FORIAN

Introduction to CHRONOS

Forian's Hybrid Real-World Data Ecosystem

Presented By: Seth Kuranz, PhD
Email: seth.kuranz@forian.com

Table of Contents

Background

Real-world data (RWD) is an important and growing data source for the generation of real-world evidence (RWE) in public health, clinical, healthcare, and life sciences sectors. RWD refers to the data collected through routine patient care and clinical practice and is usually transactional or administrative in nature; however, RWD may also refer to data reported by patients, wearable device data, and registry data. RWE refers to the findings, insights, and solutions generated using RWD (Franklin et al., 2017; Jarow et al., 2017; Yuan et al., 2018).

Although efforts to improve the reliability and validity of RWD and RWE are ongoing (Berger et al., 2017; Concato & Corrigan-Curay, 2022; Franklin et al., 2017; Jarow et al., 2017; Prada-Ramallal et al., 2019; Rassen et al., 2019; Wang, Pottegård, et al., 2022; Wang, Sreedhara, et al., 2022), to date, RWD has contributed to hypothesis generation, assessment of safety and patient outcomes, measurement of healthcare resource use, characterization of the quality of care, and other aspects of health and wellbeing in real-world populations (Jarow et al., 2017; Yuan et al., 2018). The passage of the 21 Century Cures Act in 2016 highlighted the role of RWD in the acceleration of regulatory decision-making around expanding indications of approved prescription drugs, and there have since been further developments and dissemination of regulatory guidance. The use of RWD and RWE in regulatory decision-making will continue to expand as stakeholders fully understand when and how RWD may serve as a valid substitute for the expensive, resource-intensive collection of trial data (Franklin et al., 2021; Wang, Sreedhara, et al., 2022).

A substantial barrier to using RWD successfully is aligning research aims with appropriate data sources. Given the pragmatic nature of RWD collection, assessing fit-for-purpose use cases requires an understanding of available data elements, the provenance of the data sources, and the representativeness of the patient populations available to researchers.

Introduction to CHRONOS

Forian, Inc. is a leader in RWD and RWE generation and has developed one of the largest, integrated, repositories of healthcare data in the US. CHRONOS, Forian's linked RWD ecosystem, is a novel offering that brings together open claims, closed claims, and consumer data to meet research aims across the clinical and life sciences industry.

CHRONOS includes patient data linked across the following RWD sources:

- Closed medical claims data,
- Closed pharmacy claims data,
- Open submit and remit medical claims data,
- Open pharmacy claims data,
- Consumer data [i.e., social determinants of health (SDoH) data].

Healthcare claims flow into CHRONOS through three different paths. Open claims flow into CHRONOS directly from the routing services that manage the submit (1) and remit (3) stages of the billing process. Providers generate claims at service locations and submit these claims to routing services. Routing services standardize and direct the claims to the appropriate payers. Reimbursement decisions are made by payers and claim remittances are sent back to the provider through routing services, which may or may not be the same routing service used in the submission stage of the revenue cycle. Closed claims flow into CHRONOS directly from payers after an adjudication decision is made and the claim is resolved (2). In CHRONOS, the claims data from each source are normalized and standardized to a proprietary data model that allows for streamlined analysis across sources (Figure 1).

Claims data is supplemented by consumer data in the CHRONOS RWD ecosystem. Consumer databases that aggregate sociodemographic information and purchasing behavior of individuals are

linked to patients' open and closed claims data. Linked consumer data elements expand and broaden the demographic, socioeconomic, and healthcare profile created based on the claims data alone.

Strengths and Limitations of RWD

Open Claims Data

The open claims data in CHRONOS includes both submit and remit medical claims and pharmacy claims. Open claims data provides visibility into patient care independent of a health insurance plan. Sourcing open claims from clearinghouses and switches results in longer periods of follow-up and a more frequent refresh cadence (~weekly) than what is generally available in closed claims data. Open claims include data elements related to the medical billing cycle including the reason for the rejection of a claim, rebate or coupon use for prescriptions, patient responsibility for payments, and provider details. A limitation of open claims data is the incomplete capture of care (Table 1), which can, for example, lead to missing data or unmeasured confounding in outcomes studies (Pauly et al., 2016; Sanchez et al., 2021; Wade et al., 2017).

Closed Claims Data

The closed claims data in CHRONOS includes institutional, professional, and pharmacy claims that have been adjudicated by healthcare payers. A crucial strength of the closed claims data is the near complete capture of care while the patient is enrolled in a health insurance plan. Follow-up studies may be limited as patients enroll and disenroll in health insurance plans, which prevents the measurement of longer-term health outcomes. Due to the adjudication process, closed claims are lagged by 3-6 months, on average, limiting potential research questions related to emerging health characteristics and therapeutics (Table 1).

Consumer Data

The consumer data in CHRONOS provides data on demographics, socioeconomic factors, and purchasing behaviors of patients in the closed and open claim data. The consumer data expands the list of data elements traditionally available in administrative RWD and includes patients' and/or households' race/ethnicity, occupation, education, marital status, family and household structure, income, and purchasing behavior. Purchasing behavior is associated with medication adherence, independently of other SDoH (Krumme et al., 2016), and can provide details on access to care and health behaviors not documented in a clinical setting. Given the data collection methods, the consumer data is cross-sectional in nature limiting the ability to assess changes in social determinants over time.

Hybrid RWD

A hybrid RWD model combines the strengths of each data source while mitigating the limitations found in the complementary data sources. Open claims data provides insight into uptake and treatment patterns of newly approved therapies; measures of incidence and prevalence of disease stratified by patient, payer, and provider characteristics; and commercial targeting, market sizing, and patient profiling. Combining the open data and closed claims data supports the stabilization of longitudinal patient cohorts and the evaluation of potential biases. Closed claims data supports outcomes and comparative effectiveness studies that examine the burden of illness and cost of care; unmet needs; adherence, persistence, and treatment patterns; and the natural history of disease. Combining the closed data with the open data can provide insight into out-of-pocket or out-of-network care and can expand visibility into the patient journey beyond periods of enrollment in a commercial health insurance plan. In all instances, the consumer data provides insights into SDoH that cannot be sourced from the administrative claims data.

Research Opportunities with CHRONOS

Forian's CHRONOS ecosystem includes data from over 350 million linked patients in the US with healthcare claims captured between January 1, 2017, and 2023. Across the closed, open, and consumer data, 43% of patients have a link to at least two of the three sources and over 21 million patients are present in all three sources. Of the patients present in all three sources, 98% have both medical and pharmacy claims data providing rich detail on the healthcare received by this subset of CRHONOS patients (Figure 2).

To demonstrate the value of a combined open and closed claims data system, Figure 3 presents a hypothetical journey for a patient diagnosed with breast cancer. Episodes of care 2 through 4 occur while the patient is enrolled in a health plan associated with the closed claims data and these episodes provide detailed information about the patient's care in a physician's office and inpatient setting. Although a breast cancer diagnosis occurs in Episode 2, without the history provided by the open data, the diagnostic mammogram associated with the diagnosis would not be visible to a researcher. Similarly, without the open data providing visibility into Episode 6, a change in health plan would be missing from the patient's journey. In Episode 2, the capture of a sertraline prescription in the open data demonstrates how out-of-pocket costs paid for by the patient may not be present in the closed data (Figure 3).

To further characterize opportunities available using the CHRONOS data ecosystem, we provide an example analysis and results using a chronic kidney disease (CKD) cohort. The objectives of the analyses in the context of this white paper are to:

- Describe the value of linking SDoH data to claims data.
- Demonstrate how patient volumes change when combining closed and open claims data.
- Assess changes in study results when using hybrid RWD compared to using closed or open claims data alone.

Example: Progression to ESRD in Patients With and Without Diabetes

Background

The prevalence of diagnosed chronic kidney disease (CKD) among US adults is approximately 20% with 2,269 patients per million persons treated for end stage renal disease (ESRD) in 2018 (CDC surveillance). Diabetes is one of the most common causes of progression from CKD to ESRD, with males more likely than females to progress (Koye et al., 2018; Ricardo et al., 2019). Published evidence on the differences in the risk of progression between racial/ethnic groups is inconsistent (Hounkpatin et al., 2020).

Objectives

1. Describe the associations between SDoH and progression to stage 5 (S5) CKD or ESRD among newly diagnosed CKD patients in CHRONOS.
2. Assess the impact of race/ethnicity on the rate of progression to S5 CKD or ESRD among newly diagnosed CKD patients with and without diabetes in CHRONOS.
3. Evaluate changes to the study cohorts and effect estimates when using hybrid claims data as compared to closed claims data alone in CHRONOS.

Methods

This was a retrospective observational analysis using CHRONOS. For all objectives, a primary cohort was defined using closed claims data elements. For objective 3, a secondary cohort was defined using hybrid RWD (i.e., closed claims combined with open claims data elements) with open claims data elements supplementing the inclusion/exclusion criteria.

Primary Cohort Definition

The index event was defined as the date of the first claim for stage 3 (S3) or stage 4 (S4) CKD (ICD-10 codes N18.3, N18.30, N18.31, N18.32, N18.4) between January 1, 2018, and March 31, 2022, in the closed claims data. Patients were required to be at least 18 years old and have at least 12 months of continuous enrollment prior to the index event (i.e., the baseline period). Progression was defined as the date of the first claim for S5 CKD (N18.5) or ESRD (N18.6) after the index event and before March 31, 2022. Variable follow-up ended at the date of progression or end of enrollment in a health insurance plan in the primary cohort (Figure 4).

Patients were excluded from the analysis if they were first diagnosed with S3 or S4 CKD before January 1, 2018, if a claim for S5 CKD or ESRD was submitted prior to the first claim for S3 or S4 CKD, if patients did not have continuous enrollment in a health insurance plan for 12 months before the index event, or if patients did not link to the consumer data available in CHRONOS. A link to the consumer data was required in this analysis due to the focus on SDOH and race/ethnicity in objective 2 (Table 2). All patients with a link to the consumer data but missing a specific consumer data element were placed into an 'unknown' category for that element.

Patient characteristics defined in the claims data included age at index, sex, diagnoses in the Charlson Comorbidity Index (CCI), CCI score, CKD stage at index, and baseline diabetes diagnosis. CCI comorbidities were used to calculate a CCI score based on published guidance (Glasheen et al., 2019). Given that a diagnosis of S3 or S4 CKD was a required criterion for inclusion in this analysis and stratification by diabetes was required to meet the study objectives, renal- and diabetes-related CCI categories were excluded from the calculation of the CCI score.

Patient characteristics from the consumer data included race/ethnicity (White, Black, Hispanic, other), smoking status (yes or no), presence (children: yes or no) and number (0, 1, 2, 3) of children, marital status (married: yes or no), household size (1, 2, 3, 4, or 5 household members), educational attainment (completed high school, completed college, completed graduate school, completed vocational or technical training), occupation (professional/technical, student, homemaker, retired), dwelling type (multi family or single family unit), home ownership (owner, renter), household income (USD), and net worth (USD).

Study Analyses

The analysis for objective 1 was conducted in the overall CKD patient cohort. Descriptive and objective 2 and 3 analyses stratified CKD patients into cohorts with and without a diagnosis of diabetes (E08* – E13*) in the 12 months before the index event. Categorical variables are reported as frequencies and percentages and continuous variables as means and standard deviations (SD). Standardized mean differences (SMD) are described for all baseline and SDoH characteristics.

Objective 1 Analysis

Rates of progression per 100,000 person-years, incidence rate ratios (IRRs), and 95% confidence intervals (CIs) were calculated for baseline study characteristics and SDoH in the primary cohort using a Poisson regression model. The model adjusted for all baseline characteristics and SDoH.

Objective 2 Analysis

A Poisson regression model was used to examine the interaction between diabetes and race/ethnicity, adjusting for baseline characteristics and SDoH. An interaction term was included in the model using patients without diabetes and patients who are white as the combined reference category. IRRs and 95% CIs are reported.

Objective 3 Analysis

Propensity score matching was used to adjust for baseline differences between CKD patients with and without diabetes. Propensity scores were generated using a logistic regression model that included all baseline and SDOH characteristics as predictors. A model was created separately for the primary and secondary cohorts. Patients with and without diabetes were matched 1-to-1 using greedy-nearest-neighbor matching algorithm and a caliper of 0.25 times the pooled standard deviation. Adjusted rates per 100,000 PYs and IRRs were calculated using a Poisson model, accounting for the 1-to-1 matching. For the secondary cohort, variable follow-up ended at the date of progression, end of enrollment in a health insurance plan in the primary cohort, or the date of the last claim in the open claims data if open claims are present for a patient following disenrollment in the closed claims data.

All analyses were conducted using SAS Analytics Pro.

Results

In the primary cohort analysis, 32,639 and 80,219 CKD patients with and without diabetes, respectively, met the study inclusion/exclusion criteria (Table 2). Most patients were newly diagnosed with S3 CKD at index (94.1% and 94.9%). Mean follow-up among patients with and without diabetes was 538 (SD: 396) and 575 (SD: 425) days with 5.7% and 2.8% progressing to S5 CKD or ESRD within the follow-up time, respectively (Table 3).

Patients with diabetes were older [mean 59.4 (SD: 8.3) vs 56.9 (SD: 10.2)] and more likely to be Black (15.4% vs 12.4%) or Hispanic (8.8% vs 6.0%). Patients without diabetes had a higher income (≥ $125,000) and higher net worth (≥ $1,000,000). Patients with diabetes had a higher mean CCI score [mean 1.2 (SD: 1.7) vs 0.9 (SD: 1.6)] than patients without diabetes. Comorbid cerebrovascular disease, myocardial

infarction, peripheral vascular disease, and congestive heart failure were more prevalent in patients with than without diabetes (Table 3).

Objective 1 Results

In the primary cohort, patients diagnosed with S4 CKD at index were more likely to progress than patients with S3 [7.9 (7.4-8.4)]. Female [0.8 (0.8-0.9)] and Black [0.7 (0.7-0.8)] patients were less likely to progress than male and White patients, respectively, and Hispanic patients [1.1 (1.0-1.2) more likely to progress than White patients (Table 4).

Objective 2 Results

After adjusting for baseline characteristics and SDoH, there was a significant interaction between baseline diabetes and race/ethnicity regarding the risk of progression. Compared to White patients without diabetes, Black [1.5 (1.4-1.7)], Hispanic [2.5 (2.2-2.9)], and White [2.1 (1.8-2.4)] patients with diabetes were more likely to progress. Black patients without diabetes [0.7 (0.6-0.8)] were less likely to progress than White patients without diabetes (Figure 5).

Objective 3 Results

The secondary cohort identified 105,214 newly diagnosed CKD patients from the primary cohort after excluding prevalent S3 or S4 CKD patients identified in the open claims data (i.e., patients with a CKD diagnosis in the open claims before the first CKD diagnosis in the closed claims data was recorded). After excluding the 7,644 patients with prevalent CKD that went unidentified in the closed claims data, an additional 1,518 patients with a diabetes diagnosis and an additional 957 patients with progression to S5 CKD or ESRD were identified in the open claims. After supplementing the closed claims data with the open claims, fewer patients had a CCI score of zero with a corresponding increase in scores of one or more during the baseline period (Table 5).

In the primary cohort, 32,639 patients with diabetes were matched to 32,639 patients without diabetes. In the secondary cohort, 31,846 patients with diabetes and 31,846 patients without diabetes were matched with two patients with diabetes excluded because no match was available. In both the primary (Table 6) and secondary cohort analyses, propensity score matching resulted in balanced baseline characteristics (i.e., all SMD <10%).

Patients with diabetes were 2.2 (2.0-2.4) and 2.3 (2.2-2.5) times as likely to progress in the primary and secondary cohorts, respectively. Differences were observed in the absolute rates of progression with rates attenuated in the primary compared to secondary cohort. In the primary cohort, patients with diabetes had a rate of progression of 10.8/100,000 PYs (10.3/100,000 PYs -11.3/100,000 PYs). In the secondary cohort, patients with diabetes had a rate of progression of 13.8/100,000 PYs (13.2/100,000 PYs -14.3/100,000 PYs) (Figure 6).

Discussion

This analysis examined CKD progression among patients with and without diabetes. Expectedly, patients diagnosed with S4 CKD at index were more likely to progress over follow-up than patients diagnosed with S3 CKD. As reported in the literature, female patients were less likely to progress than males (Ricardo et al., 2019).

Reported differences in risk of progression between racial/ethnic groups are inconsistent in the published literature. Studies have suggested differences between Black and White patients can be completely or at least partially explained by comorbidities, such as diabetes (Hounkpatin et al., 2020). This analysis found Black patients less likely and Hispanic patients more likely to progress as compared to White patients. When examining race/ethnicity in combination with diabetes, Black, Hispanic, and

White patients with diabetes were more likely to progress than White patients without diabetes. Among patients without diabetes, Black patients were less likely to progress compared to White patients. Although this analysis is suggestive of an interaction between race and diabetes, further work is required to explore treatments patterns and more fully characterized outcomes (i.e., dialysis and transplants) among racial/ethnic groups.

Comparing the primary cohort to the secondary cohort, the inclusion of open claims data identified the following: additional patients with a history of CKD and diabetes; additional patients that progressed to S5 CKD or ESRD; and additional patients with a history of comorbidities that were not present in the closed claims data. Excluding patients with prevalent CKD, defined as patients with a diagnosis of CKD in the open data before the index event defined in the closed claims data alone, reduced the overall cohort size but increased the certainty that newly diagnosed CKD patients were identified for this analysis. Rates of progression among diabetic and non-diabetic patients increased in the secondary, compared to the primary, cohort, likely due to the additional cases of progression that occurred following disenrollment from the closed claims data.

For the analysis presented in this white paper, the definitions of CKD and ESRD were restricted to ICD-10 diagnosis codes recorded in CHRONOS. Rates or progression may differ when accounting for renal replacement therapy, dialysis, and kidney transplants. In regards to differences between racial/ethnic groups, further analysis should explore the role of CKD treatments, which have been shown to be used in a higher proportion of people in ethnic minority groups (Hounkpatin et al., 2020). This analysis does not include clinical characteristics, such as estimated glomerular filtration rates (eGFR), which may provide a more detailed assessment of progression than diagnosis codes alone but are not generally available in medical billing data. Additional linkages to electronic medical records could better characterize these details.

Incidence rates and rate ratios were used to assess CKD progression, accounting for person-time accrued over follow-up in each cohort. Time to progression was not assessed directly in the reported findings; however, Cox Proportional Hazards models were conducted in a sensitivity analysis (not presented) and did not result in meaningful differences between hazards ratios and IRRs. This was likely due to similar distributions of time to progression and over all follow-up in cohorts of patients with and without diabetes.

Summary of CHRONOS

CHRONOS, Forian's hybrid data ecosystem, presents an opportunity to conduct RWE studies taking full advantage of the strengths of open, closed, and consumer data. Open claims data provides a broad perspective on the health care provided to the US population. Closed claims data provides a near complete picture of patient care covered by a payer in the US. Combining closed and open claims data with consumer data, expands the instances in which each data source is fit for purpose. Linking CHRONOS data to additional data sets, both novel and traditional, will continue to increase its utility for researchers in public health, clinical, healthcare, and life sciences sectors.

Ethics Statement

CHRONOS is a retrospective observational data product licensed from Forian Inc. that includes patient-level data deidentified in compliance with the federal privacy and security rules of the Health Insurance Portability and Accountability Act of 1996, as amended ("HIPAA"), including by the Health Information Technology for Economic and Clinical Health Act and any implementing regulations issued by the U.S. Department of Health and Human Services pursuant thereto, and the regulations published thereunder at 45 CFR Parts 160, 162 and 164, as may be amended from time to time, known as the Standards for Privacy of Individually Identifiable Health Information ("Privacy Rule") and the Security Standards (the "Security Rule", and together with the Privacy Rule, the "HIPAA Regulations"), as certified by the determination of a qualified expert in accordance with Section §164.514(b)(1) of the Privacy Rule. At the time of the analysis yielding the study, Forian's CHRONOS data product included healthcare claims data from over 300 million deidentified patients receiving care in the United States between 2017 and 2022. The analyses presented here utilize only deidentified patient-level data and was therefore exempt from Institutional Review Board approval.

References

Berger, M. L., Sox, H., Willke, R. J., Brixner, D. L., Eichler, H.-G., Goettsch, W., Madigan, D., Makady, A., Schneeweiss, S., Tarricone, R., Wang, S. V., Watkins, J., & Daniel Mullins, C. (2017). Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiology and Drug Safety*, *26*(9), 1033–1039. https://doi.org/10.1002/pds.4297

Concato, J., & Corrigan-Curay, J. (2022). Real-World Evidence—Where Are We Now? *The New England Journal of Medicine*, *386*(18), 1680–1682. https://doi.org/10.1056/NEJMp2200089

Franklin, J. M., Patorno, E., Desai, R. J., Glynn, R. J., Martin, D., Quinto, K., Pawar, A., Bessette, L. G., Lee, H., Garry, E. M., Gautam, N., & Schneeweiss, S. (2021). Emulating randomized clinical trials with nonrandomized real-world evidence studies: First results From the RCT DUPLICATE initiative. *Circulation*, *143*(10), 1002–1013. https://doi.org/10.1161/CIRCULATIONAHA.120.051718

Franklin, J. M., Schneeweiss, S., & Solomon, D. H. (2017). Assessment of confounders in comparative effectiveness studies from secondary databases. *American Journal of Epidemiology*, *185*(6), 474–478. https://doi.org/10.1093/aje/kww136

Glasheen, W. P., Cordier, T., Gumpina, R., Haugh, G., Davis, J., & Renda, A. (2019). Charlson Comorbidity Index: ICD-9 update and ICD-10 translation. *American Health & Drug Benefits*, *12*(4), 188–197.

Hounkpatin, H. O., Fraser, S. D. S., Honney, R., Dreyer, G., Brettle, A., & Roderick, P. J. (2020). Ethnic minority disparities in progression and mortality of pre-dialysis chronic kidney disease: A systematic scoping review. *BMC Nephrology*, *21*(1), 217. https://doi.org/10.1186/s12882-020-01852-3

Jarow, J. P., LaVange, L., & Woodcock, J. (2017). Multidimensional evidence generation and FDA regulatory decision making: Defining and using “Real-World” data. *JAMA*, *318*(8), 703–704. https://doi.org/10.1001/jama.2017.9991

Koye, D. N., Magliano, D. J., Reid, C. M., Jepson, C., Feldman, H. I., Herman, W. H., & Shaw, J. E. (2018). Risk of progression of nonalbuminuric CKD to end-stage kidney disease in people with diabetes: The CRIC (Chronic Renal Insufficiency Cohort) study. *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation*, *72*(5), 653–661. https://doi.org/10.1053/j.ajkd.2018.02.364

Pauly, N. J., Talbert, J. C., & Brown, J. (2016). Low-cost generic program use by medicare beneficiaries: Implications for medication exposure misclassification in administrative claims data. *Journal of Managed Care & Specialty Pharmacy*, *22*(6), 741–751. https://doi.org/10.18553/jmcp.2016.22.6.741

Prada-Ramallal, G., Takkouche, B., & Figueiras, A. (2019). Bias in pharmacoepidemiologic studies using secondary health care databases: A scoping review. *BMC Medical Research Methodology*, *19*(1), 53. https://doi.org/10.1186/s12874-019-0695-y

Rassen, J. A., Bartels, D. B., Schneeweiss, S., Patrick, A. R., & Murk, W. (2019). Measuring prevalence and incidence of chronic conditions in claims and electronic health record databases. *Clinical Epidemiology*, *11*, 1–15. https://doi.org/10.2147/CLEP.S181242

Ricardo, A. C., Yang, W., Sha, D., Appel, L. J., Chen, J., Krousel-Wood, M., Manoharan, A., Steigerwalt, S., Wright, J., Rahman, M., Rosas, S. E., Saunders, M., Sharma, K., Daviglus, M. L., Lash, J. P., & CRIC Investigators. (2019). Sex-related disparities in CKD progression. *Journal of the American Society of Nephrology: JASN*, *30*(1), 137–146. https://doi.org/10.1681/ASN.2018030296

Sanchez, R. J., Ge, W., Wei, W., Ponda, M. P., & Rosenson, R. S. (2021). The association of triglyceride levels with the incidence of initial and recurrent acute pancreatitis. *Lipids in Health and Disease*, *20*(1), 72. https://doi.org/10.1186/s12944-021-01488-8

Wade, R. L., Patel, J. G., Hill, J. W., De, A. P., & Harrison, D. J. (2017). Estimation of Missed Statin Prescription Use in an Administrative Claims Dataset. *Journal of Managed Care & Specialty Pharmacy*, *23*(9), 936–942. https://doi.org/10.18553/jmcp.2017.23.9.936

Wang, S. V., Pottegård, A., Crown, W., Arlett, P., Ashcroft, D. M., Benchimol, E. I., Berger, M. L., Crane, G., Goettsch, W., Hua, W., Kabadi, S., Kern, D. M., Kurz, X., Langan, S., Nonaka, T., Orsini, L.,

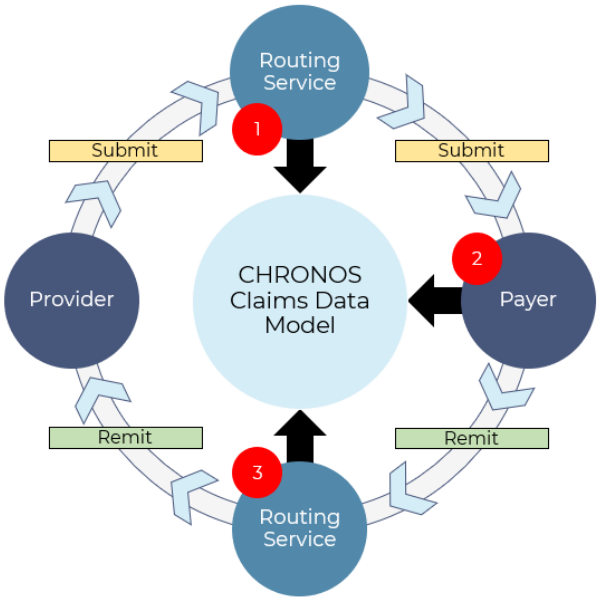
Perez-Gutthann, S., Pinheiro, S., Pratt, N., … Williams, R. J. (2022). HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force. *Pharmacoepidemiology and Drug Safety*. https://doi.org/10.1002/pds.5507

Wang, S. V., Sreedhara, S. K., Schneeweiss, S., & REPEAT Initiative. (2022). Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. *Nature Communications*, *13*(1), 5126. https://doi.org/10.1038/s41467-022-32310-3

Yuan, H., Ali, M. S., Brouwer, E. S., Girman, C. J., Guo, J. J., Lund, J. L., Patorno, E., Slaughter, J. L., Wen, X., Bennett, D., & ISPE Comparative Effectiveness Research Special Interest Group. (2018). Real-world evidence: What it is and what it can tell us according to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG). *Clinical Pharmacology and Therapeutics*, *104*(2), 239–241. https://doi.org/10.1002/cpt.1086

Figures

Figure 1. Medical Billing and the Flow of Claims Data Into CHRONOS

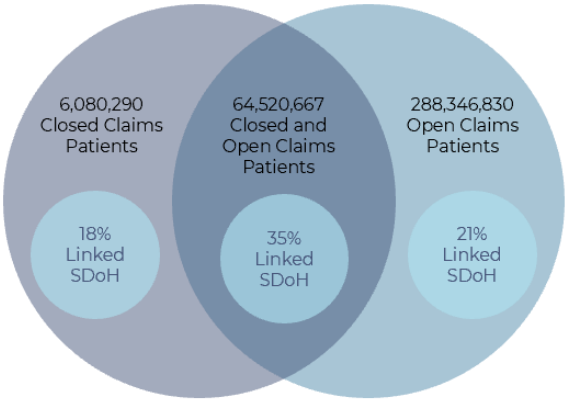


Open and closed claims data flow into CHRONOS at different revenue cycle stages for medical billing.

1. Open claims data is sourced from clearinghouses and switches that manage submitted claims from providers to payers.
2. Payers receive submitted claims and return remitted claims to the provider following adjudication. Closed claims data is sourced following the adjudication process, which summarizes the submitted and remitted data and may include multiple revenue cycles.
3. Open claims data is sourced from clearinghouses and switches that manage remitted claims from payers to providers.

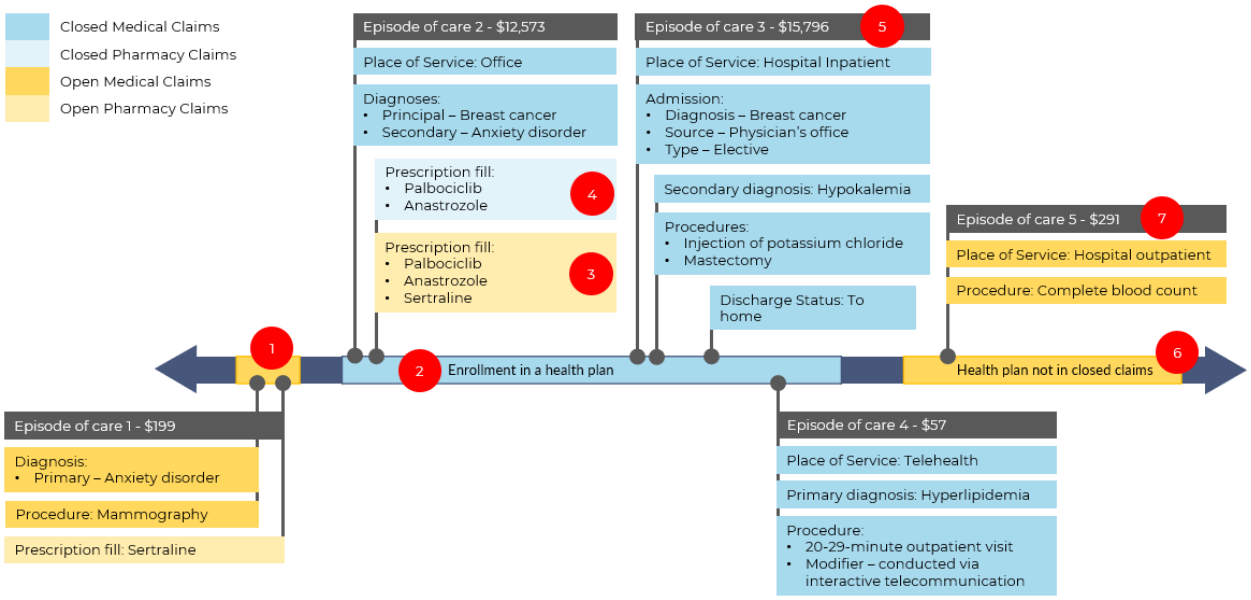
Figure 2. CHRONOS Patient Counts Between 2017 and 2022 by Data Source

Total CHRONOS Patient Count	**N = 358,947,787**
Closed Claims Patients	**N = 70,600,957** 98% Medical claims 61% Pharmacy claims
Open Claims Patients	**N = 352,867,497** 90% Medical claims 57% Pharmacy claims
SDoH Patients	**N = 96,175,806**



Note: Patient counts current as of March 2023

Figure 3. A Patient Journey Combining Open and Closed Claims in CHRONOS

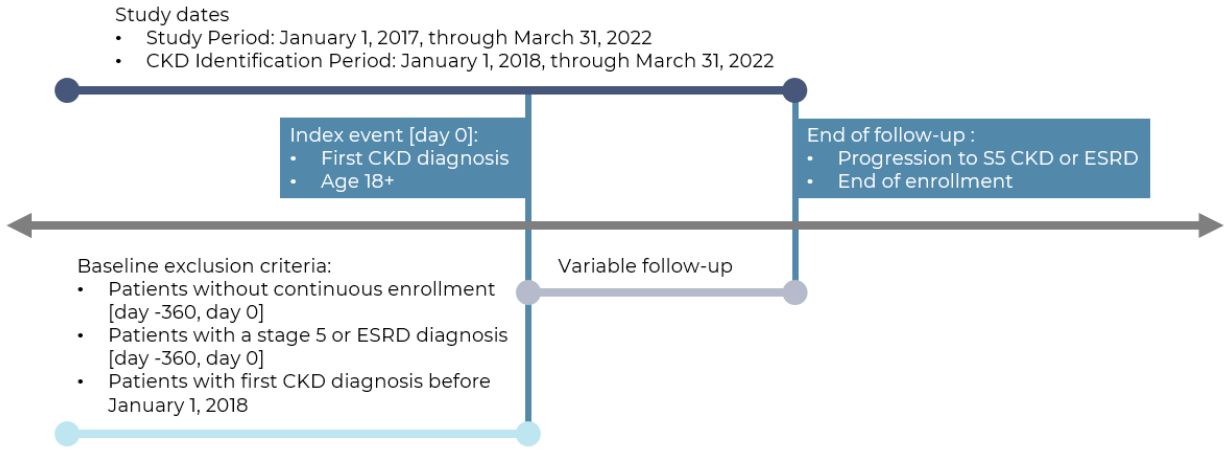


Characterizing the patient journey across open and closed claims data:

1. A history of care from open claims data is visible before enrollment in the health insurance plan associated with closed claims data.
2. Enrollment in a health insurance plan in the closed claims data provides near complete capture of care during covered time periods.
3. Of the prescription fills captured in the open claims data, only two are also found in the closed claims data. The sertraline fill is only represented in the open claims data.
4. Of the prescription fills captured in the closed claims data only sertraline is missing. The sertraline fill is likely related to the patient's anxiety diagnosis and paid for out-of-pocket.
5. The depth of clinical detail available in the closed claims data covers the entirety of the inpatient visit including the source of admission and discharge status.
6. After disenrollment from the insurance plan associated with the closed claims data, the patient can be followed in a subsequent health insurance plan.
7. Cost of care reflects the total cost associated with each episode, including costs from institutional, professional, and pharmacy claims data.

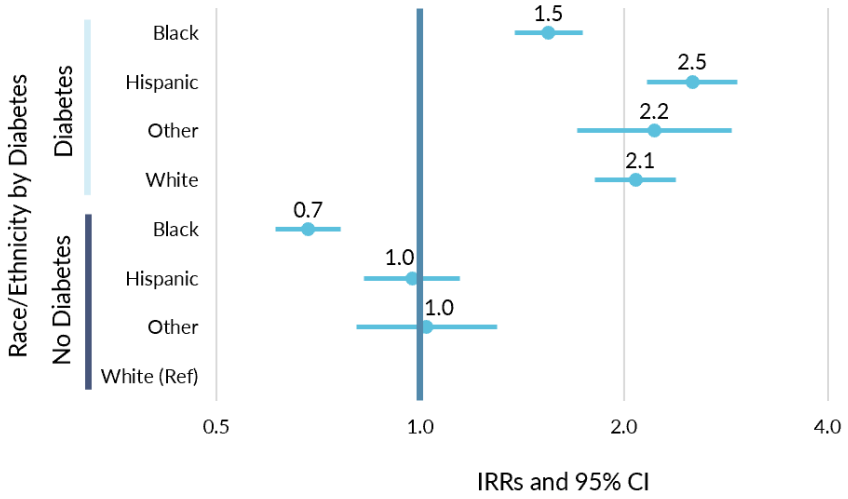
Note: data elements and patterns displayed in the figure represent results seen in multiple CHRONOS patient records. These results are collapsed into the single patient journey displayed with select details excluded for obfuscation.

Figure 4. Progression to ESRD Study Diagram



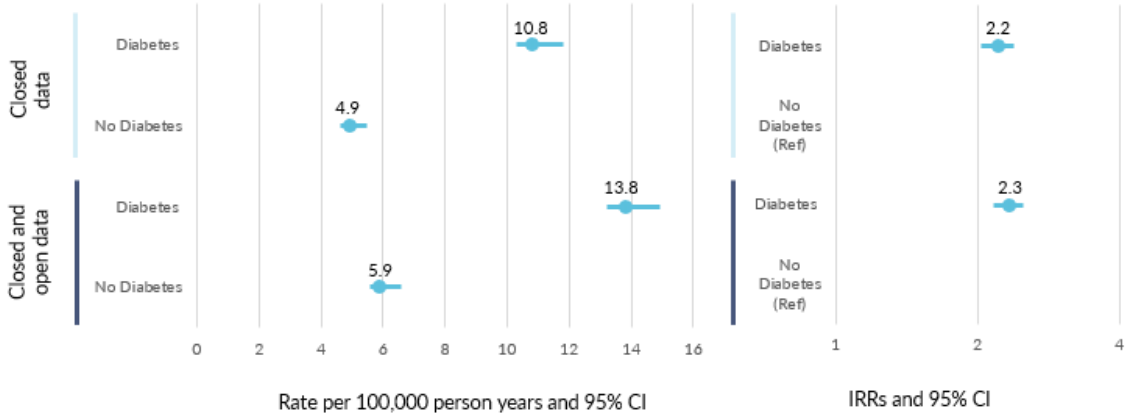
Abbreviations: CKD, chronic kidney disease; ESRD, end stage renal disease; S5, stage 5

Figure 5. Associations Between Race/Ethnicity, Diabetes, and Progression to ESRD



Abbreviations: CI, confidence interval; IRRs, incidence rate ratios; Ref, reference

Figure 6. Rates and IRRs for Progression to ESRD by Data Source



Abbreviations: CI, confidence interval; IRRs, incidence rate ratios; Ref, reference

Tables

Table 1. Strengths and Limitations of Open and Closed Claims Data in CHRONOS

	Open Claims Data	Closed Claims Data
Source	• The data is sourced from routing systems such as claims clearinghouses and pharmacy switches.	• The data is sourced from healthcare insurers/payers.
Strengths	• The length of a patient's record is not restricted by enrollment in an insurance plan. • The geographic representation is not limited by the location of the payer. • The lag between the current date and date of the last claim is less than two weeks.	• The capture of patient care is nearly complete while the patient is enrolled in an insurance plan. • The inclusion of institutional, professional, and pharmacy claims provides a near total cost of care.
Limitations	• The capture of care for a patient may be incomplete because clearinghouses and pharmacy switches vary by provider. • The medical billing process may be unresolved (i.e., submitted claims without remitted claims).	• The average length of enrollment in an insurance plan is 1-2 years. • The lag between the current date and the date of the last claim in the date is 3-6 months, on average.

Table 2. Patient Attrition in the Primary CKD cohort

	n	% of row above
Patients with a claim for S3 or S4 CKD between 2017 and 2022	647,694	--
Exclusions:		
Patients with S5 CKD or ESRD before the first diagnosis for S3 or S4 CKD	630,686	97
Patients with a first diagnosis for S3 or S4 CKD before January 1, 2018	445,163	71
Patients without 12 months of continuous enrollment before the index event	311,953	70
Patients without a link to SDoH data elements	112,858	36

Abbreviations: CKD, chronic kidney disease; ESRD, end stage renal disease; SDoH, social determinants of health; S3, stage 3; S4, stage 4

Table 3. Baseline and SDoH Characteristics Among CKD Patients With and Without Diabetes in the Primary CKD Cohort

	CKD With Diabetes		CKD Without Diabetes		SMD
	n or mean	% or SD	n or mean	% or SD	%
Total	32,639		80,219		
Study Cohort Characteristics					
Index CKD Diagnosis (n, %)					
S3 CKD	30,728	94.1	76,156	94.9	3
S4 CKD	1,911	5.9	4,063	5.1	3
Days of Follow-Up (mean, SD)	528.4	396.3	575.0	425.0	11
Progression (n, %)	1,863	5.7	2,260	2.8	14
Demographics					
Age (mean, SD)	59.4	8.3	56.9	10.2	27

Sex (n, %)					
Female	15,496	47.5	39,486	49.2	3
Male	17,143	52.5	40,733	50.8	3
SDoH Characteristics					
Race / Ethnicity (n, %)					
Black	5,029	15.4	9925	12.4	9
Hispanic	2,869	8.8	4,834	6.0	11
White	23,671	72.5	63,100	78.7	14
Other	1,070	3.3	2,360	2.9	2
Smoker (n, %)	4,369	13.4	9,145	11.4	6
Presence of Children (n, %)	13,751	42.1	35,186	43.9	3
Number of Children (n, %)					
0 Children	19,021	58.3	45,432	56.6	3
1 Child	7,503	23.0	18,614	23.2	1
2 Children	3,212	9.8	8,377	10.4	2
3 Children	2,903	8.9	7,796	9.7	3
Married (n, %)	20,698	63.4	50,853	63.4	0
Single Parent (n, %)	4,070	12.5	10,256	12.8	1
Household Size (n, %)					
1 Member	4,871	14.9	12,204	15.2	1
2 Members	11,447	35.1	27,432	34.2	2
3 Members	7,617	23.3	18,431	23.0	1
4 Members	4,163	12.8	10,299	12.8	0
5 Members	4,541	13.9	11,853	14.8	2
Education (n, %)					
Completed High School	14,882	45.6	36,535	45.5	0
Completed College	9,600	29.4	23,389	29.2	1
Completed Graduate School	4,307	13.2	11,758	14.7	4
Vocational/Technical Training	246	0.8	524	0.7	1
Unknown	3,604	11.0	8,013	10.0	3
Occupation (n, %)					
Professional/Technical	21,446	65.7	52,284	65.2	1
Student	233	0.7	675	0.8	1
Homemaker	2,133	6.5	4,808	6.0	2
Retired	756	2.3	1,546	1.9	3
Unknown	8,071	24.7	20,906	26.1	3
Dwelling type (n, %)					
Multi Family Unit	3,959	12.1	9,473	11.8	1
Single Family Unit	28,436	87.1	70,132	87.4	1
Unknown	244	0.7	614	0.8	0
Homeownership (n, %)					
Owner	27,220	83.4	67,448	84.1	2

Renter	5,383	16.5	12,641	15.8	2
Unknown	36	0.1	130	0.2	1
Household Income (n, %)					
< $15,000	1,599	4.9	3,345	4.2	4
$15,000 - $19,999	1,592	4.9	3,361	4.2	3
$20,000 - $29,999	2,772	8.5	5,743	7.2	5
$30,000 - $39,999	3,256	10.0	6,825	8.5	5
$40,000 - $49,999	3,523	10.8	7,749	9.7	4
$50,000 - $74,999	7,725	23.7	17,676	22.0	4
$75,000 - $99,999	4,955	15.2	12,626	15.7	2
$100,000 - $124,999	2,661	8.2	7,516	9.4	4
≥ $125,000	4,556	14.0	15,378	19.2	14
Net worth (n, %)					
$0	2,097	6.4	3,931	4.9	7
$1 - $4,999	1,457	4.5	2,705	3.4	6
$5,000 - $9,999	1,210	3.7	2,393	3.0	4
$10,000 - $24,999	2,194	6.7	4,123	5.1	7
$25,000 - $49,999	2,221	6.8	4,349	5.4	6
$50,000 - $99,999	3,554	10.9	7,347	9.2	6
$100,000 - $249,999	6,421	19.7	14,712	18.3	3
$250,000 - $499,999	5,264	16.1	13,294	16.6	1
$500,000 - $999,999	3,620	11.1	10,832	13.5	7
$1,000,000 - $1,999,999	2,464	7.5	8,418	10.5	10
≥ $2,000,000	1,707	5.2	6,991	8.7	14
Unknown Net Worth	430	1.3	1,124	1.4	1
Charlson Comorbidity Index (CCI)					
Comorbidities (n, %)					
AIDS	35	0.1	103	0.1	1
Cerebrovascular Disease	2,888	8.8	4,578	5.7	12
Chronic Pulmonary Disease	5,615	17.2	11,353	14.2	8
Dementia	446	1.4	868	1.1	3
Hemiplegia or Paraplegia	412	1.3	613	0.8	5
HIV	216	0.7	797	1.0	4
Mild Liver Disease	3,465	10.6	6,423	8.0	9
Moderate-to-Sever Liver Disease	523	1.6	824	1.0	5
Any Malignancy	2,479	7.6	6,573	8.2	2
Myocardial Infarction	6,313	19.3	7,487	9.3	29
Peptic Ulcer Disease	495	1.5	985	1.2	2
Peripheral Vascular Disease	4,406	13.5	6,025	7.5	20
Rheumatic disease	1,124	3.4	3,551	4.4	5
Metastatic Solid Tumors	596	1.8	1,924	2.4	4
Congestive Heart Failure	5,229	16.0	6,650	8.3	24

CCI Score - continuous (mean, SD)	1.2	1.7	0.9	1.6	18
CCI Score – categorical (n, %)					
0	15,092	46.2	46,569	58.1	24
1	7,543	23.1	16,024	20.0	8
2	4,502	13.8	8,134	10.1	11
3	2,647	8.1	4,462	5.6	10
4+	2,855	8.7	5,030	6.3	9

Abbreviations: CCI, Charlson Comorbidity Index; CKD, chronic kidney disease; SD, standard deviation; SDoH, social determinants of health; SMD, standardized mean difference; S3, stage 3; S4, stage 4

Table 4. Predictors of Progression to ESRD in the Primary CKD Cohort

	Rate per 100,000 PYs	(95% CI)	IRR	(95% CI)
CKD at Index				
S4 CKD	44	(30 - 64)	7.9	(7.4 - 8.4)
S3 CKD	6	(4 - 8)	Ref	--
Demographics				
Sex				
Female	14	(10 - 21)	0.8	(0.8 - 0.9)
Male	17	(12 - 25)	Ref	--
Race / Ethnicity				
Black	12	(8 - 17)	0.7	(0.7 - 0.8)
Hispanic	18	(12 - 26)	1.1	(1.0 - 1.2)
Other	17	(12 - 26)	1	(0.9 - 1.3)
White	17	(11 - 24)	Ref	--
Smoker	16	(11 - 24)	1.1	(1.0 - 1.2)
Non-Smoker	15	(10 - 22)	Ref	--
SDoH Characteristics				
Number of Children				
1 Child	16	(11 - 23)	0.9	(0.8 – 1.0)
2 Children	15	(10 - 22)	0.9	(0.7 – 1.0)
3 Children	14	(9 - 21)	0.8	(0.7 – 1.0)
0 Children	18	(12 - 26)	Ref	--
Married	16	(11 - 23)	1	(0.9 - 1.1)
Not Married	16	(11 - 23)	Ref	--
Single Parent	17	(11 - 25)	1.1	(1.0 - 1.3)
Not a Single Parent	15	(10 - 21)	Ref	--
Household Size				
2 Members	15	(10 - 22)	1	(0.9 - 1.1)
3 Members	16	(11 - 23)	1	(0.9 - 1.2)
4 Members	16	(11 - 24)	1.1	(0.9 - 1.3)
5 Members	17	(12 - 25)	1.2	(1.0 - 1.4)
1 Member	15	(10 - 22)	Ref	--
Education				
Completed High School	16	(11 - 23)	Ref	--

Completed College	15	(10 - 21)	0.9	(0.9 – 1.0)
Completed Graduate School	15	(10 - 21)	0.9	(0.8 – 1.0)
Vocational/Technical Training	17	(10 - 29)	1.1	(0.8 - 1.5)
Unknown	16	(11 - 23)	1	(0.9 - 1.1)
Occupation				
Professional/Technical	15	(11 - 22)	Ref	--
Student	13	(8 - 22)	0.9	(0.6 - 1.2)
Homemaker	14	(10 - 21)	0.9	(0.8 - 1.1)
Retired	18	(12 - 28)	1.2	(1.0 - 1.5)
Unknown	18	(12 - 26)	1.2	(1.1 - 1.3)
Dwelling type				
Multi Family Unit	16	(11 - 24)	1	(0.9 - 1.1)
Single Family Unit	16	(11 - 23)	Ref	--
Unknown	15	(9 - 24)	1	(0.7 - 1.3)
Homeownership				
Owner	17	(14 - 20)	0.9	(0.9 – 1.0)
Renter	18	(15 - 21)	Ref	--
Unknown	13	(5 - 36)	0.7	(0.3 – 2.0)
Household Income				
< $15,000	15	(10 - 23)	0.9	(0.8 - 1.1)
$15,000 - $19,999	15	(10 - 23)	0.9	(0.8 - 1.1)
$20,000 - $29,999	15	(10 - 23)	0.9	(0.8 - 1.1)
$30,000 - $39,999	14	(9 - 20)	0.8	(0.7 – 1.0)
$40,000 - $49,999	16	(11 - 23)	0.9	(0.8 - 1.1)
$50,000 - $74,999	16	(11 - 24)	1	(0.9 - 1.1)
$75,000 - $99,999	16	(11 - 23)	0.9	(0.8 - 1.1)
$100,000 - $124,999	17	(11 - 25)	Ref	--
≥ $125,000	17	(12 - 25)	1	(0.9 - 1.2)
Net worth				
$0	16	(11 - 24)	1.1	(0.9 - 1.3)
$1 - $4,999	17	(11 - 25)	1.1	(0.9 - 1.3)
$5,000 - $9,999	19	(12 - 28)	1.2	(1.0 - 1.5)
$10,000 - $24,999	16	(11 - 24)	1.1	(0.9 - 1.2)
$25,000 - $49,999	15	(10 - 23)	Ref	--
$50,000 - $99,999	16	(11 - 24)	1.1	(0.9 - 1.2)
$100,000 - $249,999	16	(11 - 23)	1	(0.9 - 1.2)
$250,000 - $499,999	15	(10 - 22)	1	(0.8 - 1.2)
$500,000 - $999,999	16	(11 - 24)	1.1	(0.9 - 1.3)
$1,000,000 - $1,999,999	15	(10 - 23)	1	(0.8 - 1.2)
≥ $2,000,000	13	(8 - 19)	0.8	(0.7 – 1.0)
Unknown Net Worth	14	(9 - 23)	0.9	(0.7 - 1.3)

Note: Rates and IRRs adjusted for baseline characteristics and SDoH, including age and CCI score (not shown in table)
Abbreviations: CCI, Charlson Comorbidity Index; CKD, chronic kidney disease; IRRs, incidence rate ratio; PYs, person-years; SDoH, social determinants of health; S3, stage 3; S4, stage 4

Table 5. Changes in Patient Counts in the Primary and Secondary CKD Cohorts

	Primary Cohort		Secondary Cohort	
	n or mean	% or SD	n or mean	% or SD
Number of patients in final cohort	112,858		105,214	
A history of diabetes defined in the closed claims data (n, %)	32,639	28.9	30,330	28.8
A history of diabetes defined in the closed or open claims data (n, %)	34,756	30.8	31,848	30.3
Progression defined in the closed claims data (n, %)	4,123	3.7	3,696	3.5
Progression defined in the closed or open claims data (n, %)	5,241	4.6	4,653	4.4
CCI score defined by closed claims data (mean, SD)	1.0	1.6	1.0	1.6
CCI score defined by closed or open claims data (mean, SD)	1.1	1.6	1.0	1.6
CCI score categories in closed claims data (n, %)				
0	61,661	54.6	57,541	54.7
1	23,567	20.9	21,934	20.8
2	12,636	11.2	11,764	11.2
3	7,109	6.3	6,588	6.3
4+	7,885	7.0	7,387	7.0
CCI score categories in closed or open claims data (n, %)				
0	60,159	53.3	56,429	53.6
1	23,823	21.1	22,130	21.0
2	12,964	11.5	12,007	11.4
3	7,434	6.6	6,815	6.5
4+	8,478	7.5	7,833	7.4

Abbreviations: CCI, Charlson Comorbidity Index; SD, standard deviation

Table 6. Baseline Characteristics Among CKD Patients With and Without Diabetes After Matching in the Primary CKD Cohort

	CKD With Diabetes		CKD Without Diabetes		SMD
	n or mean	% or SD	n or mean	% or SD	%
Total	32,639		32,639		
Demographics					
Age (mean, SD)	59.4	8.3	59.5	8.7	1
Sex (n, %)					
Female	15,496	47.5	15,525	47.6	0
Male	17,143	52.5	17,114	52.4	0
Race / Ethnicity (n, %)					
Black	5,029	15.4	4,988	15.3	0
Hispanic	2,869	8.8	2,655	8.1	2
White	23,671	72.5	23,920	73.3	2
Other	1,070	3.3	1,076	3.3	0

Smoker (n, %)	4,369	13.4	4,366	13.4	0
SDoH Characteristics					
Number of Children (n, %)					
0 Children	19,021	58.3	19,142	58.6	1
1 Child	7,503	23.0	7,467	22.9	0
2 Children	3,212	9.8	3,196	9.8	0
3 Children	2,903	8.9	2,834	8.7	1
Married (n, %)	20,698	63.4	50,853	63.4	0
Single Parent (n, %)	4,070	12.5	10,256	12.8	1
Household Size (n, %)					
1 Member	4,871	14.9	4,910	15.0	0
2 Members	11,447	35.1	11,498	35.2	0
3 Members	7,617	23.3	7,643	23.4	0
4 Members	4,163	12.8	4,135	12.7	0
5 Members	4,541	13.9	4,453	13.6	1
Education (n, %)					
Completed High School	14,882	45.6	14,930	45.7	0
Completed College	9,600	29.4	9554	29.3	0
Completed Graduate School	4,307	13.2	4,280	13.1	0
Vocational/Technical Training	246	0.8	276	0.8	1
Unknown	3,604	11.0	3,599	11.0	0
Occupation (n, %)					
Professional/Technical	21,446	65.7	21,523	65.9	0
Student	233	0.7	212	0.6	1
Homemaker	2,133	6.5	2,170	6.6	0
Retired	756	2.3	725	2.2	1
Unknown	8,071	24.7	8,009	24.5	0
Dwelling type (n, %)					
Multi Family Unit	3,959	12.1	3,918	12.0	0
Single Family Unit	28,436	87.1	28,477	87.2	0
Unknown	244	0.7	244	0.7	0
Homeownership (n, %)					
Owner	27,220	83.4	27,223	83.4	0
Renter	5,383	16.5	5,380	16.5	0
Unknown	36	0.1	36	0.1	0
Household Income (n, %)					
< $15,000	1,599	4.9	1,614	4.9	0
$15,000 - $19,999	1,592	4.9	1,607	4.9	0
$20,000 - $29,999	2,772	8.5	2,801	8.6	0
$30,000 - $39,999	3,256	10.0	3,229	9.9	0
$40,000 - $49,999	3,523	10.8	3,539	10.8	0
$50,000 - $74,999	7,725	23.7	7,769	23.8	0

$75,000 - $99,999	4,955	15.2	4,996	15.3	0
$100,000 - $124,999	2,661	8.2	2,626	8.0	0
≥ $125,000	4,556	14.0	4,458	13.7	1
Net worth (n, %)					
$0	2,097	6.4	2,057	6.3	1
$1 - $4,999	1,457	4.5	1,447	4.4	0
$5,000 - $9,999	1,210	3.7	1,165	3.6	1
$10,000 - $24,999	2,194	6.7	2,179	6.7	0
$25,000 - $49,999	2,221	6.8	2,198	6.7	0
$50,000 - $99,999	3,554	10.9	3,537	10.8	0
$100,000 - $249,999	6,421	19.7	6,585	20.2	1
$250,000 - $499,999	5,264	16.1	5,429	16.6	1
$500,000 - $999,999	3,620	11.1	3,673	11.3	1
$1,000,000 - $1,999,999	2,464	7.5	2,418	7.4	1
≥ $2,000,000	1,707	5.2	1,487	4.6	3
Unknown Net Worth	430	1.3	464	1.4	1
CCI					
CCI Score (mean, SD)	1.2	1.7	1.2	1.8	2

Abbreviations: CCI, Charlson Comorbidity Index; CKD, chronic kidney disease; SD, standard deviation; SDoH, social determinants of health.