

Original Paper

Agentic AI for Clinical Outcomes Research, Population Health Management Analyses With Large Administrative Databases, and Generating Epidemiological Estimates of Diseases: Feasibility and Validation Study

Daniel Frisch¹, MD, FHRS, FACC; Michael Fronstin², MBA; Carla Zema³, PhD

¹Bruce & Robbi Toll Heart and Vascular Institute, Division of Cardiology, Thomas Jefferson University, Philadelphia, PA, United States

²Vireo Strategies, Philadelphia, PA, United States

³Zema Consulting, Athens, AL, United States

Corresponding Author:

Carla Zema, PhD

Zema Consulting

22549 Mary Grace Dr

Athens, AL, 35613

United States

Phone: 1 4704226860

Email: carla@zema.consulting

Abstract

Background: There is tremendous enthusiasm for the use of AI in health care because of the ability to analyze existing data for preventative, diagnostic, and treatment support. Agentic AI can feasibly provide access to large real-world datasets for the generation of real-world evidence for health care and clinical applications to health care providers, researchers, and administrators without access to large analytic programming resources.

Objective: The objective of this study was to understand the feasibility of using agentic AI for clinical outcome research and population health management. Specifically, this study used an agentic AI evidence-generation platform to obtain epidemiological estimates of diverse medical conditions, with the results evaluated against existing AI frameworks.

Methods: Prevalence estimates of 6 conditions (amyotrophic lateral sclerosis, acute myeloid leukemia, bladder cancer, Huntington disease, elevated lipoprotein (a), and Parkinson disease) were estimated using an agentic AI evidence-generation platform applied to an administrative claims database with representation from every US state. Gender-specific rates were calculated within the following age categories: 0 to 17, 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 to 74 years and 75 years and older. Period prevalence was estimated from January 1, 2020, to June 30, 2025, and annual prevalence rates for each year were estimated from 2020 to 2024. Continuous enrollment for 12 months was required during the study period for inclusion. Source code generated by the platform as part of the analysis was reviewed by an independent programmer for validation of methods and programming, and analyses were replicated using traditional programming methods. Results obtained throughout the process were evaluated against several existing AI application frameworks.

Results: Regarding accuracy, epidemiological estimates obtained using the agentic AI platform were consistent with published estimates for all 6 conditions, as well as with estimates obtained from traditional programming methods. Regarding rigor, the agentic AI platform conducted the analysis with rigor by confirming acceptable methods in published literature for the type of data source used. Code lists used for the analysis were confirmed against existing algorithms when available. Appropriate statistical methods were used to compare differences in prevalence rates by age and gender. Regarding trust (explainability, transparency, replicability, traceability, and validation), the agentic AI platform generated all source code used for the analyses, which was reviewed and validated for accuracy and appropriateness. The analysis included a “human in the loop” to validate the research question, data extraction method, statistical analysis plan, and output plan prior to proceeding with each step.

Conclusions: With specific design considerations to ensure responsible use, agentic AI can be invaluable to increasing the accessibility of large datasets for applied clinical outcome research and population health management analyses.

(*J Med Internet Res* 2026;28:e100284) doi: [10.2196/100284](https://doi.org/10.2196/100284)

KEYWORDS

agentic artificial intelligence; agentic AI; clinical outcome research; population health; administrative data; prevalence; real-world evidence; artificial intelligence; AI

Introduction

There is tremendous enthusiasm for the use of AI in health care because of the ability to analyze existing data for preventative, diagnostic, and treatment support [1-4]. AI is a broad term and encompasses a diverse array of tools and functions. As AI continues to evolve, it is important to differentiate between AI agents, or generative AI, and agentic AI to understand the value and limitations of various applications in health care. AI agents have the ability to cull information from massive datasets using large language models and can generate content and answer questions, whereas agentic AI refers to systems capable of autonomous decision-making and self-learning to accomplish complex scientific tasks [5-8]. An example of differentiation is a literature review: generative AI can identify appropriate literature, whereas agentic AI can identify literature and apply relevant findings to adapt study methodology to conform to accepted standards.

Over the past decade, the recognition of the value of real-world evidence (RWE), defined as evidence generated during real-world patient care, has exploded [9,10]. RWE is generated from real-world data (RWD) sources, meaning data that are generated through the process of patient care, such as electronic health records, administrative claims and billing data, laboratory data, registries, patient monitoring devices, etc. The use of RWD to generate RWE has typically been limited to researchers with sophisticated programming skills capable of analyzing such large datasets, especially as the data were collected for other purposes, thereby requiring a significant amount of preparation and data cleaning for analytic purposes. For example, administrative claims and billing data represent a rich source of information on health care services provided as a claim is generated for practically every health care interaction. However, understanding how to use the data requires knowledge of medical coding, application of continuous enrollment criteria, and understanding of the claims adjudication process. This complexity can make the use of RWD difficult for health care and clinical applications such as clinical outcome monitoring and population health management.

AI, specifically agentic AI, can make access to large real-world datasets for the generation of RWE for health care and clinical applications feasible for health care providers, researchers, and administrators without access to large analytic programming resources. A recent survey of physician attitudes toward the use of AI in medicine found that acceptance and enthusiasm were shaped more by experience with and engagement in the use of AI than by demographic or physician characteristics [11]. AI is only as good as the data used to create the results [12]. Accountability and trust are also important factors [2,13,14]. Trust is multifactorial and encompasses fairness, transparency, robustness, and reliability. Moreover, while agentic AI has the capability to operate autonomously, having a “human in the loop” to confirm decisions at certain critical times during the process is important to ensuring accountability and trust [2,15].

The objective of this study was to understand the feasibility of using agentic AI for clinical outcome research and population health management evaluated against applicable frameworks of the responsible use of AI in health care. Specifically, this study used HealthVerity eXOs, an agentic AI evidence-generation platform powered by Medeloop, to obtain epidemiological estimates of several diverse medical conditions as subgroup identification is the foundation of clinical outcome research and population health management.

Methods

Study Design

Prevalence estimates of 6 diverse conditions were obtained using an agentic AI evidence-generation platform applied to an administrative claims database with representation from every US state. Using claims databases to estimate prevalence has been well documented in many disease areas [16-19]. Prevalence estimates were obtained for amyotrophic lateral sclerosis (ALS), acute myeloid leukemia (AML), bladder cancer, Huntington disease (HD), elevated lipoprotein (a), and Parkinson disease (PD). The *International Classification of Diseases, 10th Revision (ICD-10)*, diagnosis codes used to define each of the conditions are available in [Multimedia Appendix 1](#). The conditions were selected to represent diverse conditions across varying specialty areas and to include both common and rare conditions. Gender-specific rates were calculated within the following age categories: 0 to 17, 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 to 74 years and 75 years and older. Period prevalence was estimated from January 1, 2020, to June 30, 2025. A minimum of 12 months of continuous enrollment was required for inclusion as is typical in analyses using closed claims data. Clopper-Pearson, Wald, or Wilson 95% CIs, as appropriate, were computed for each stratum-specific prevalence estimate, and Pearson chi-square tests were conducted to test for differences within the strata. Statistical significance was assessed at an α value of .05.

AI Validation

The following standardized prompt was used across analyses for consistency: “What is the period prevalence of <condition> (expressed as a rate per 100,000), and how do these rate differ according to these age categories (0–17, 18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75 and older) and gender within these age categories? Require continuous enrollment for a minimum of 12 months at any time during the study period.” The agentic AI platform allows for human-in-the-loop acceptance at the following steps of the analysis: query validation, data extraction plan, statistical plan, and output plan. The platform also conducts targeted literature reviews to identify accepted methodologies, including disease identification algorithms. All source code used for the analysis is also generated with the results. Source code was reviewed by an independent programmer for validation of methods and programming. The independent programmer was blinded to the creator of the source code and was not aware

that it was generated by agentic AI. The programmer had direct experience with the dataset used for the analysis and was provided with the study objective of estimating the period prevalence of 6 diverse conditions, including the age and gender breakdown. The programmer was asked to review the code to determine whether the source code could meet the study objectives and used appropriate methods and to replicate the analysis using traditional programming methods.

The underlying architecture comprises a researcher agent (orchestrator), a semantic engine, an AI-powered notebook, and resources. The researcher agent interacts with the user to both receive and respond to questions. The semantic engine comprises an agent that has access to a semantic search querier and semantic database for concept extraction, code extraction, code lookup, and code retrieval. The researcher agent sends tasks to the AI-powered notebook, which comprises an agent that writes code and uses the notebook, which is connected to the cloud-based dataset, to complete tasks. The resources include custom logic and other reference files. The results from the semantic engine and AI-powered notebook are sent back to the researcher agent to coordinate all research tasks per the user-approved plan. The researcher agent conducts the final synthesis of information to present it back to the user. At the time of this study, fine-tuned versions of GPT-4.1 (OpenAI) and text-embedding-3-large were used for search and embeddings, respectively, to power the semantic engine.

Ethical Considerations

Because this study used anonymized, deidentified data collected in the ordinary course of care, institutional review board approval and Declaration of Helsinki considerations were not required in accordance with Title 45 of the Code of Federal Regulations part 46.104(d)(4)(ii). All study data were handled through protocols compliant with HIPAA (Health Insurance Portability and Accountability Act).

Results

The administrative claims data included 59 million individuals from January 1, 2020, to June 30, 2025, with every state in the United States represented. The demographics of the study population, as well as geographic region, are shown in [Table 1](#).

Of the 59 million individuals in the study population, 38.4 million (65.1%) met the continuous enrollment requirement of 12 months during the study period. The demographics of the study population are similar to those of the US population. Although not used in the analysis, every geographic region of the United States was represented in the data, and the distribution of those with known race was similar to that of the population of the United States.

The overall period prevalence for each of the 6 conditions is shown in [Table 2](#), and the age-gender breakdowns for each condition are illustrated in [Figure 1](#).

As expected, period prevalence was higher than the annual prevalence estimates. There was a greater difference between annual prevalence and period prevalence estimates for conditions where there was likely to be less overlap in the patients represented in each year. Interestingly, there were differences in how the annual prevalences were calculated across conditions within the study period. For ALS, HD, elevated lipoprotein (a), and PD, an individual who was confirmed in the numerator in one year was carried forward to the numerators of subsequent years as all are lifelong diseases; however, annual prevalences were calculated independently without carrying forward for the other conditions. The differences in prevalence rates when calculated using traditional programming methods ranged from -1.0 to +0.1 in all instances and are listed in [Multimedia Appendix 2](#).

Table 1. Retrospective US claims data study population characteristics for individuals in the dataset from January 1, 2020, to June 30, 2025 (N=38.4 million)^a.

Characteristic	Patients (millions), n (%)
Gender	
Female	20 (52.2)
Male	17.5 (45.5)
Other or unknown	0.9 (2.2)
Age (y)	
0-17	8.9 (23.3)
18-24	3.9 (10.2)
25-34	5.2 (13.5)
35-44	4.9 (12.8)
45-54	4.3 (11.2)
55-64	4.6 (12.0)
65-74	3.4 (8.9)
≥75	2.3 (6.0)
Missing or unknown	0.9 (2.3)
Race and ethnicity	
Asian	0.7 (1.8)
Black	2.7 (7.0)
White	11.1 (28.9)
Other	0.9 (2.3)
Missing or unknown	23.0 (59.9)
Region	
Midwest	10.4 (27.1)
Northeast	6.8 (17.7)
South	11.8 (30.7)
West	8.3 (21.6)
Missing or unknown	1.1 (2.9)

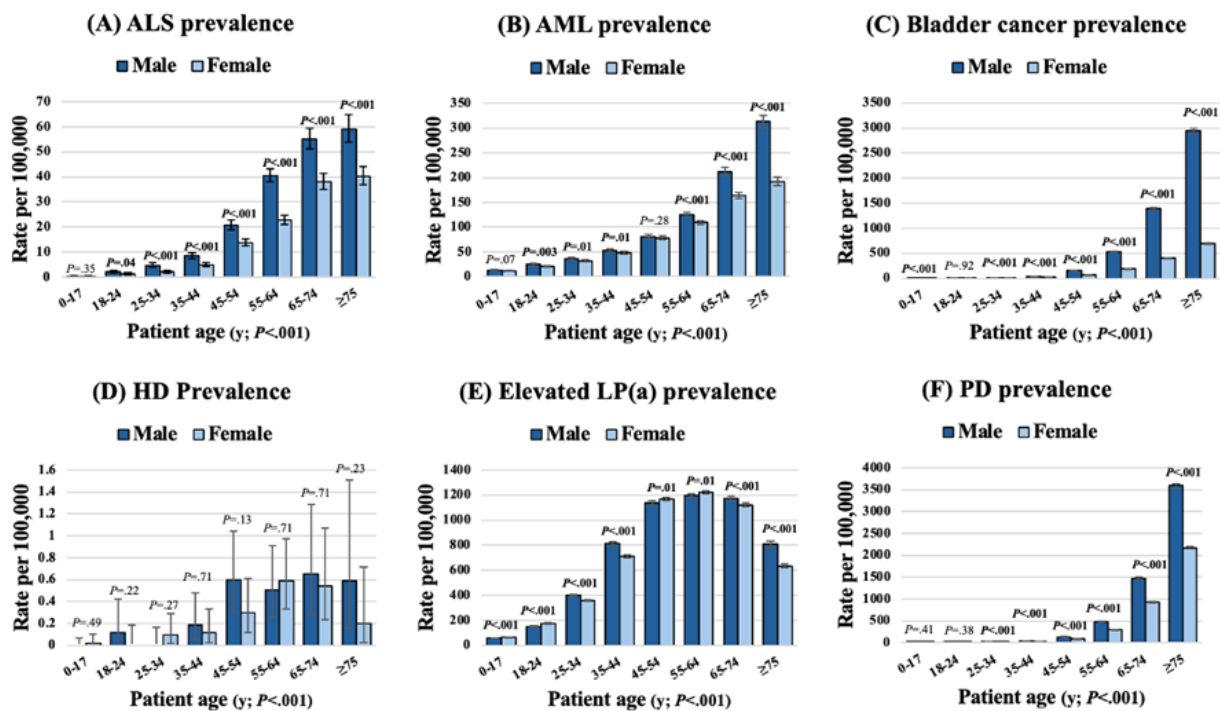
^aPercentages do not add up to 100 due to rounding.

Table 2. Overall period prevalence of 6 diverse conditions in a retrospective administrative US claims data analysis from January 1, 2020 to June 30, 2025.

Condition	Prevalence rate per 100,000 (95% CI)					
	Overall	2020	2021	2022	2023	2024
ALS ^a	16.1 (15.6-16.6)	7.1 (6.7-7.5)	9.0 (8.6-9.4)	10.0 (9.5-10.4)	10.5 (10.1-10.9)	11.0 (10.5-11.4)
AML ^b	51.6 (50.7-52.4)	41.1 (36.6-46.1)	49.9 (47.1-52.9)	46.1 (43.6-48.7)	31.8 (30.3-33.4)	28.9 (27.6-30.4)
Bladder cancer	228.0 (226.3-229.8)	144.7 (142.8-146.6)	132.1 (130.6-133.8)	136.0 (134.4-137.6)	131.9 (130.3-133.4)	132.3 (130.7-134.0)
HD ^c	10.2 (9.8-10.6)	6.7 (6.3-7.1)	7.4 (7.0-7.8)	7.9 (7.5-8.3)	8.4 (8.0-8.8)	9.4 (8.9-9.8)
Elevated lipoprotein (a)	173.9 (172.4-175.5)	47.7 (46.4-49.0)	76.0 (74.5-77.5)	100.9 (99.2-102.6)	130.6 (128.5-132.7)	167.0 (164.2-170.0)
PD ^d	333.8 (331.7-335.9)	206.1 (203.9-208.4)	232.0 (229.9-234.1)	263.0 (260.9-265.2)	273.8 (271.6-276.0)	288.9 (286.4-291.4)

^aALS: amyotrophic lateral sclerosis.
^bAML: acute myeloid leukemia.
^cHD: Huntington disease.
^dPD: Parkinson disease.

Figure 1. Period prevalence of 6 diverse conditions in a retrospective administrative US claims data analysis from January 1, 2020, to June 30, 2025, by gender and age category. *Statistically significant difference for age and gender comparisons; statistical significance assessed at $\alpha=.05$. ALS: amyotrophic lateral sclerosis; AML: acute myeloid leukemia; HD: Huntington disease; LP(a): lipoprotein (a); PD: Parkinson disease.



For the gender- and age-specific estimates, there were statistically significant differences among the age groups for all conditions. The period prevalence of ALS, AML, bladder cancer, and PD was significantly higher in male than in female individuals in adults except for the ages of 18 to 24 years in bladder cancer and PD as well as the ages of 45 to 54 years in AML. The rarity of HD made gender differences within age groups difficult to detect. The prevalence of elevated lipoprotein (a) was similar and highest among the age groups of 45 to 74 years and then decreased at the ages of 75 years and older. The largest difference in gender was observed in the 35 to 44 years

and 75 years and older age groups, where prevalence was significantly higher in male individuals. While statistically significant, female individuals had a slightly higher prevalence of elevated lipoprotein (a) than male individuals in the age groups of less than 25 years and 45 to 64 years.

With regard to human in the loop, there were 8 instances in which modifications to the specific analyses being suggested by the agentic AI were made to ensure consistency across the analyses. Table 3 shows the 8 instances of the analysis that were proposed by the agentic AI platform and the changes made

during the human-in-the-loop confirmations of the proposed analyses.

Table 3. Human-in-the-loop changes to the analyses proposed by the agentic AI for the retrospective administrative US claims data prevalence analyses from January 1, 2020, to June 30, 2025.

Condition	Proposed by AI	Modification made	Stage in the process
Bladder cancer and Huntington disease	Continuous enrollment definition of a >45-d gap	Continuous enrollment definition of a >30-d gap	Acceptance of the statistical plan
Elevated lipoprotein (a) and acute myeloid leukemia	Calculation of prevalence rates for gender groups and age groups separately	Calculation of gender groups within each age subgroup	Validation of analysis question
Bladder cancer	C65.x, C66.x, and C67.x codes to align with the work by Gitlin et al [20]	Removal of C65.x and C66.x codes to align with clinical expert input	Acceptance of data extraction plan
Parkinson disease	G20, G20.Ax, G20.Bx, and G20.C codes	G20, G20.Ax, and G20.Bx codes	Acceptance of data extraction plan
Elevated lipoprotein (a) and Parkinson disease	No carryforward in estimating annual prevalence rates from 2020 to 2024	Specified carryforward to be used in estimating annual prevalence rates from 2020 to 2024	Acceptance of the statistical plan

Two instances were related to the definition of an acceptable gap allowed for the implementation of continuous enrollment criteria and were edited prior to acceptance of the statistical analysis plan, 2 instances were related to inconsistent interpretation of producing age-gender breakdowns of prevalence rates and were edited prior to the acceptance of the query validation, 2 instances involved the modification of the codes used to identify the target population and were edited prior to the acceptance of the data extraction plan, and 2 instances required the addition of the specification to use a carryforward approach to calculate the annual prevalence estimates (elevated lipoprotein (a) and PD). The selection of continuous enrollment criteria to be defined by less than a 45-day gap in enrollment as opposed to 30 days was not wrong per se but was modified for consistency across the 6 conditions. Both 30- and 45-day definitions are commonly used for acceptable gaps in enrollment to still be considered continuously enrolled. Although prompts for each condition used the exact same wording, there were 2 instances in which the subgroups were not interpreted correctly to be gender groups *within* the age groups. These instances were corrected with further clarification at the validation of the analysis question. The 2 instances in which the selection of the *ICD-10* coding was updated were based on clinical expert input. For bladder cancer, the code list initially selected was based on a published bladder cancer study using administrative claims data. There were several studies published that all had different code lists for the selection of the study population. The clinical expert advised that the lists were not wrong but would result in a narrower definition with a higher sensitivity for identifying those with the condition. For PD, the code changed in late 2023 from G20 with no modifiers to having 5 different modifiers. G20.C represented parkinsonisms. The clinical expert felt that a patient could have parkinsonisms without having a diagnosis of PD, thus advising to exclude that code. The agentic AI platform opted to run a sensitivity analysis with and without the code given the uncertainty of the inclusion in defining PD. Finally, when initially running the inquiry for elevated lipoprotein (a) and PD, annual estimates were not calculated using a carryforward approach. Because PD is a chronic condition with no current “cure” and elevated lipoprotein (a) is a genetic

marker, use of the carryforward approach was appropriate for these conditions.

Due to the size of some of the analyses based on the number in the target population, some analyses were broken down and conducted in several iterations on the platform. In these instances, the overall period prevalence estimates were calculated separately from the gender and age category breakdown.

Discussion

Principal Findings

Epidemiological estimates across 6 diverse conditions were calculated using both traditional programming methods and an agentic AI platform that included a human in the loop to validate the analysis at several key times throughout. The estimates that were generated were practically the same and were completed in much less time (ie, approximately 80%-90% less time on average) using the agentic AI platform.

Application of Relevant Frameworks Addressing the Use of AI in Health Care

Several frameworks have emerged to provide guidance for the responsible use of AI in health care [2,4,12,21]. These frameworks can be used to understand the feasibility of using agentic AI for clinical outcome research and population health management.

Accuracy

Prevalence estimates can vary widely for a variety of reasons: representativeness and accuracy of the data source used to calculate, methodology used to estimate, period, geography, etc [22-24]. This study produced prevalence estimates that were broadly consistent with previously published estimates for all the conditions, including gender and age differences [25-40]. Given the difficulty of replicating published prevalence estimates in differing data sources, this study also estimated prevalence through traditional programming methods using the same administrative claims dataset. The results were consistent with the estimates obtained using the agentic AI platform.

While there were 8 instances in which human intervention was necessary to achieve the desired analysis, this study was unique in that consistency across individual analyses was important. The agentic AI platform for continuous enrollment criteria and diagnosis code selection was based on published literature. The choices were scientifically acceptable options, although they were not what was preferred for *this* study. Two of the instances dealt with correcting the interpretation of the prompt. The desired analysis was to calculate prevalence of gender within age subgroups but was interpreted as gender and age subgroups despite having the exact same prompt wording across conditions. Additionally, for the 4 conditions in which carryforward methodology was appropriate when calculating serial annual prevalences, 2 instances had to be corrected to specify using carryforward. That is why having a human in the loop is necessary to validate the analysis. In many instances, there are several options that are equally valid, such as continuous enrollment criteria; in other instances, nuances of methodology specific to the subject area need to be confirmed. The user may have specific reasons for choosing one method over another. Thus, the human in the loop ensures that the analysis approach meets the specific needs and preferences of the user.

Rigor

Identifying individuals with specific conditions and estimating disease prevalence are the foundation for quantifying burden, managing population health, and evaluating clinical outcomes [4]. While some may consider this type of analysis descriptive, there is a need for rigor as there are factors that must be accounted for in the analysis that make the use of agentic AI valuable. Depending on the data source, there are accepted methods for estimating prevalence [23,24,41]. For example, when using closed claims data, continuous enrollment requirements are standard practice as this ensures that an individual is represented in the data for sufficient time to be able to be identified for the numerator of the prevalence rate. Prevalence rates calculated using closed claims data that do not use continuous enrollment criteria will most likely underestimate the true prevalence. Before conducting the analysis, the agentic AI platform used in this study performed a targeted literature review to verify that the analytical methods selected were established and appropriate for addressing the specific research question.

Another example of potential bias in estimating prevalence using administrative claims data stems from the use of coding to identify the condition. Identifying some conditions is straightforward when there are clear diagnosis codes that are used. Prior to late 2023, PD was a condition with G20 *ICD-10* diagnosis codes representing all its various presentations. Diagnosis codes are created to represent conditions and relevant variations of them for reimbursement. An example where coding is not as straightforward to capture a condition is bladder cancer. *ICD-10* C67.* codes represent malignant neoplasm of the bladder. However, a recent study defined bladder cancer to also include C65.* (malignant neoplasm of the renal pelvis) and C66.* (malignant neoplasm of the ureter) codes with C67.* codes [20]. Other relevant codes include C79.1* (secondary malignant neoplasm of the bladder and other and unspecified urinary organs), D41.4 (neoplasm of uncertain behavior of the

bladder), and D09.0 (carcinoma in situ of the bladder). Differences in the inclusion or exclusion of codes will influence the prevalence estimate obtained using administrative claims data. The agentic AI platform used for this study generates a code list that will be used, conducts a targeted literature search to identify algorithms for identification that have been used in previous studies, and allows the user to edit the list prior to implementation in the analysis.

Trust

Trust is a broad concept and is often represented by factors such as explainability, transparency, reliability, traceability, and validation. These factors can be challenging with agentic AI given the autonomous nature of the technology. Having a human in the loop for collaboration is essential to building trust. The specific agentic AI platform used for the analysis pauses at various points in the process to give the user the ability to edit. Key steps in which having a human in the loop are critical are (1) the validation of the overall analysis prompt (ie, was the prompt interpreted as intended?), (2) the data extraction process and creation of the analytic data file (ie, what criteria will be used to create the analytic data file?), (3) generation of the statistical analysis plan, and (4) creation of the output plan (ie, how will results be presented?). Providing input into these critical steps in the process can ensure that the results are clinically relevant, explainable, and scientifically appropriate.

Even with the human in the loop at critical steps in the process, there are a myriad of decisions that are made during processing, including the order in which steps are taken, that can influence the specific results obtained. Being able to view the specific source code that is used throughout the process is the only way to have the ability to replicate the analysis fully. This level of transparency is essential for validation and reliability. For this study, an independent programmer reviewed all source code generated from the agentic AI platform to validate the methods and process used and was able to produce the results through traditional programming methods.

There are some limitations to using agentic AI to conduct RWD analyses. Understanding how to develop appropriate prompts that result in the desired analysis takes some iteration if the user is not experienced in using agentic AI. Most people think of generative AI when they hear “AI.” Previous research has shown that comfort with using AI increases with experience [11]. Using agentic AI to conduct analyses do not provide instantaneous results.

Comparison to Traditional Programming Methods

While not as fast as generative AI, agentic AI in RWD analyses is significantly faster than traditional programming methods. An analysis from the start of data extraction, creation of the analytic file, and conduct of the analysis would take an experienced programmer a minimum of 20 programming hours for each condition and includes targeted literature reviews, code list validation, and protocol generation. The agentic AI platform was able to complete analyses of conditions in between 40 and 90 minutes depending on the complexity of the identification of the condition and the size of the patient population with that condition.

When considering the time and resources required to conduct a comprehensive epidemiological estimate that includes targeted literature review, validation of coding, creation of the analytic file, and programming, it is easy to understand the potential in efficiency that using an agentic AI platform for such analyses can bring.

Regardless of increases in efficiency, stakeholders who are interested in conducting such analyses may not have the expertise or resources necessary to conduct a comprehensive analysis. Consequently, analyses of large datasets may not be feasible for clinical outcome research and population health management. The ability to use an agentic AI platform makes such analyses feasible for many stakeholders that would benefit from generating evidence that can be used for clinical decision-making and population health management.

Limitations

This study has several limitations. First, in understanding the application of agentic AI for clinical outcome research and population health management, generating epidemiological estimates was selected to represent these broad types of research and real-world use as most begin by identifying and estimating the population of interest. However, there are many types of research studies and analyses for which this technology is applicable. In addition to epidemiological estimates, other analyses in which it may be valuable to use an agentic AI platform include clinical practice population health management, use and treatment patterns, clinical comparative effectiveness, quality improvement, assessment of guideline-directed care, adherence and persistency with treatment, clinical trial design and feasibility, and understanding the patient health care

journey. Second, it is difficult to evaluate the accuracy of the results as so many factors can influence epidemiological estimates, such as data source, time frame, continuous enrollment criteria, criteria for numerator identification. Results compared with published estimates had broad face validity; however, statistical comparison was not made as differences may be based on other factors and not on the actual accuracy of the agentic AI. Third, this study was limited to analyses using a large administrative claims database. The value in the use of agentic AI for other types of data, such as electronic health records, registries, clinical trials, etc, may not be generalizable to these datasets and should be explored in future research. Another limitation that is common to most administrative datasets is the lack of completeness of race and ethnicity data. While the distribution of race and ethnicity in our study population seemed to be similar to that of the general population, over half of the patients in the dataset were missing race and ethnicity. The study design of only requiring a period of 12 months of continuous enrollment at any time during the study period of over 5 years was intended to maximize the number of patients that could be identified. On the other hand, this study design could have underestimated prevalence if the denominator was inflated. Finally, the agentic AI platform was specifically designed for health outcome research. Other agentic AI platforms may not produce similar results in these types of studies and analyses.

Conclusions

With specific design considerations to ensure responsible use, agentic AI can be invaluable to making large datasets accessible for applied clinical outcome research and population health management analyses.

Acknowledgments

The authors would like to acknowledge both HealthVerity and Medeloop for their support and technical assistance with the analysis. All analyses were conducted using the HealthVerity eXOs powered by Medeloop, an agentic AI evidence studio that translates plain English prompts into transparent analyses with step-by-step logic and auditable programming code. Generative AI was not used in the writing of the manuscript.

Data Availability

The dataset analyzed for this study is not publicly available and is accessible via subscription only. Data and source code for the analyses for verification purposes can be made available from the corresponding author on reasonable request.

Funding

This study was funded by HealthVerity. HealthVerity contracted author CZ to conduct the study and provided support for the article processing fee payment; however, no direct financial support was provided to any authors specifically for the preparation, drafting, or review of this manuscript.

Authors' Contributions

CZ was involved in all aspects of the study, including conceptualization, formal analysis, methodology, project administration, validation, visualization, and writing. MF was involved in conceptualization, funding acquisition, methodology, project administration, supervision, and writing. DF was involved in methodology, validation, visualization, supervision, and writing.

Conflicts of Interest

CZ was contracted to conduct the study. At the time of the study, MF was an executive contractor for HealthVerity. The other author declares no other conflicts of interest.

Multimedia Appendix 1

International Classification of Diseases, 10th Revision, diagnosis codes used to identify patients in the retrospective administrative US claims database analysis of prevalence from January 1, 2020, to June 30, 2025.

[\[DOCX File, 15 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Comparison of overall prevalence in administrative US claims data from January 1, 2020, to June 30, 2025, using traditional programming methods and an agentic AI platform.

[\[DOCX File, 16 KB-Multimedia Appendix 2\]](#)

References

1. Arbelaez Ossa L, Milford SR, Rost M, Leist AK, Shaw DM, Elger BS. AI through ethical lenses: a discourse analysis of guidelines for AI in healthcare. *Sci Eng Ethics*. Jun 04, 2024;30(3):24. [\[FREE Full text\]](#) [doi: [10.1007/s11948-024-00486-0](https://doi.org/10.1007/s11948-024-00486-0)] [Medline: [38833207](#)]
2. Alelyani T. A validated framework for responsible AI in healthcare autonomous systems. *Sci Rep*. Dec 19, 2025;15(1):44432. [\[FREE Full text\]](#) [doi: [10.1038/s41598-025-25266-z](https://doi.org/10.1038/s41598-025-25266-z)] [Medline: [41419784](#)]
3. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. *Healthc Manage Forum*. Jan 2020;33(1):10-18. [doi: [10.1177/0840470419873123](https://doi.org/10.1177/0840470419873123)] [Medline: [31550922](#)]
4. Leist AK, Klee M, Kim JH, Rehkopf DH, Bordas SP, Muniz-Terrera G, et al. Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. *Sci Adv*. Oct 21, 2022;8(42):eabk1942. [\[FREE Full text\]](#) [doi: [10.1126/sciadv.abk1942](https://doi.org/10.1126/sciadv.abk1942)] [Medline: [36260666](#)]
5. Liawrungrueang W. Artificial intelligence (AI) agents versus agentic AI: what's the effect in spine surgery? *Neurospine*. Jun 2025;22(2):473-477. [\[FREE Full text\]](#) [doi: [10.14245/ns.2550308.154](https://doi.org/10.14245/ns.2550308.154)] [Medline: [40625011](#)]
6. Soetikno BT, Nielsen CS, Pollreis A, Ting DS. Toward autonomous discovery: agentic AI and the future of ophthalmic research. *Curr Opin Ophthalmol*. Jan 01, 2026;37(1):60-65. [doi: [10.1097/ICU.0000000000001179](https://doi.org/10.1097/ICU.0000000000001179)] [Medline: [41175385](#)]
7. Tripathi S, Cook TS, Kim W. Agentic AI in radiology. *Radiology*. Feb 2026;318(2):e252730. [doi: [10.1148/radiol.252730](https://doi.org/10.1148/radiol.252730)] [Medline: [41631988](#)]
8. Zou J, Topol EJ. The rise of agentic AI teammates in medicine. *Lancet*. Feb 08, 2025;405(10477):457. [doi: [10.1016/S0140-6736\(25\)00202-8](https://doi.org/10.1016/S0140-6736(25)00202-8)] [Medline: [39922663](#)]
9. Dang A. Real-world evidence: a primer. *Pharmaceut Med*. Jan 2023;37(1):25-36. [\[FREE Full text\]](#) [doi: [10.1007/s40290-022-00456-6](https://doi.org/10.1007/s40290-022-00456-6)] [Medline: [36604368](#)]
10. Schad F, Thronicke A. Real-world evidence-current developments and perspectives. *Int J Environ Res Public Health*. Aug 16, 2022;19(16):10159. [\[FREE Full text\]](#) [doi: [10.3390/ijerph191610159](https://doi.org/10.3390/ijerph191610159)] [Medline: [36011793](#)]
11. Heinrichs H, Kies A, Nagel SK, Kiessling F. Physicians' attitudes toward artificial intelligence in medicine: mixed methods survey and interview study. *J Med Internet Res*. Aug 26, 2025;27:e74187. [\[FREE Full text\]](#) [doi: [10.2196/74187](https://doi.org/10.2196/74187)] [Medline: [40857713](#)]
12. Truong T, Gilbank P, Johnson-Cover K, Ieraci A. A framework for applied AI in healthcare. *Stud Health Technol Inform*. Aug 21, 2019;264:1993-1994. [doi: [10.3233/SHTI190751](https://doi.org/10.3233/SHTI190751)] [Medline: [31438445](#)]
13. Olawade DB, Wada OJ, David-Olawade AC, Kunonga E, Abaire O, Ling J. Using artificial intelligence to improve public health: a narrative review. *Front Public Health*. Oct 26, 2023;11:1196397. [\[FREE Full text\]](#) [doi: [10.3389/fpubh.2023.1196397](https://doi.org/10.3389/fpubh.2023.1196397)] [Medline: [37954052](#)]
14. Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: a systematic review. *Soc Sci Med*. Mar 2022;296:114782. [\[FREE Full text\]](#) [doi: [10.1016/j.socscimed.2022.114782](https://doi.org/10.1016/j.socscimed.2022.114782)] [Medline: [35152047](#)]
15. Griffen Z, Owens K. From "human in the loop" to a participatory system of governance for AI in healthcare. *Am J Bioeth*. Sep 2024;24(9):81-83. [\[FREE Full text\]](#) [doi: [10.1080/15265161.2024.2377114](https://doi.org/10.1080/15265161.2024.2377114)] [Medline: [39226015](#)]
16. Wallin MT, Culpepper WJ, Campbell JD, Nelson LM, Langer-Gould A, Marrie RA, et al. The prevalence of MS in the United States: a population-based estimate using health claims data. *Neurology*. Mar 05, 2019;92(10):e1029-e1040. [\[FREE Full text\]](#) [doi: [10.1212/WNL.0000000000007035](https://doi.org/10.1212/WNL.0000000000007035)] [Medline: [30770430](#)]
17. Mortimer K, Hartmann N, Chan C, Norman H, Wallace L, Enger C. Characterizing idiopathic pulmonary fibrosis patients using US Medicare-advantage health plan claims data. *BMC Pulm Med*. Jan 10, 2019;19(1):11. [\[FREE Full text\]](#) [doi: [10.1186/s12890-018-0759-5](https://doi.org/10.1186/s12890-018-0759-5)] [Medline: [30630460](#)]
18. Xue AZ, Anderson C, Cotton CC, Gaber CE, Feltner C, Dellon ES. Prevalence and costs of esophageal strictures in the United States. *Clin Gastroenterol Hepatol*. Sep 2024;22(9):1821-9.e4. [doi: [10.1016/j.cgh.2024.03.026](https://doi.org/10.1016/j.cgh.2024.03.026)] [Medline: [38588766](#)]
19. Dawwas GK, Weiss A, Constant BD, Parlett LE, Haynes K, Yang JY, et al. Development and validation of claims-based definitions to identify incident and prevalent inflammatory bowel disease in administrative healthcare databases. *Inflamm Bowel Dis*. Dec 05, 2023;29(12):1993-1996. [\[FREE Full text\]](#) [doi: [10.1093/ibd/izad053](https://doi.org/10.1093/ibd/izad053)] [Medline: [37043675](#)]

20. Gitlin M, McGarvey N, Shivaprakash N, Cong Z. Time duration and health care resource use during cancer diagnoses in the United States: a large claims database analysis. *J Manag Care Spec Pharm*. Jun 2023;29(6):659-670. [[FREE Full text](#)] [doi: [10.18553/jmcp.2023.29.6.659](https://doi.org/10.18553/jmcp.2023.29.6.659)] [Medline: [37276034](#)]
21. Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, et al. Consolidated health economic evaluation reporting standards for interventions that use artificial intelligence (CHEERS-AI). *Value Health*. Sep 2024;27(9):1196-1205. [[FREE Full text](#)] [doi: [10.1016/j.jval.2024.05.006](https://doi.org/10.1016/j.jval.2024.05.006)] [Medline: [38795956](#)]
22. Riedel O, Braitmaier M, Langner I. Dementia in health claims data: the influence of different case definitions on incidence and prevalence estimates. *Int J Methods Psychiatr Res*. Jun 2023;32(2):e1947. [[FREE Full text](#)] [doi: [10.1002/mpr.1947](https://doi.org/10.1002/mpr.1947)] [Medline: [36168670](#)]
23. Jensen ET, Cook SF, Allen JK, Logie J, Brookhart MA, Kappelman MD, et al. Enrollment factors and bias of disease prevalence estimates in administrative claims data. *Ann Epidemiol*. Jul 2015;25(7):519-25.e2. [[FREE Full text](#)] [doi: [10.1016/j.annepidem.2015.03.008](https://doi.org/10.1016/j.annepidem.2015.03.008)] [Medline: [25890796](#)]
24. Kopec JA. Estimating disease prevalence in administrative data. *Clin Invest Med*. Jun 26, 2022;45(2):E21-E27. [doi: [10.25011/cim.v45i2.38100](https://doi.org/10.25011/cim.v45i2.38100)] [Medline: [35752980](#)]
25. Wolfson C, Gauvin DE, Ishola F, Oskoui M. Global prevalence and incidence of amyotrophic lateral sclerosis: a systematic review. *Neurology*. Aug 08, 2023;101(6):e613-e623. [doi: [10.1212/WNL.0000000000207474](https://doi.org/10.1212/WNL.0000000000207474)] [Medline: [37308302](#)]
26. Xu L, Liu T, Liu L, Yao X, Chen L, Fan D, et al. Global variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol*. Apr 2020;267(4):944-953. [doi: [10.1007/s00415-019-09652-y](https://doi.org/10.1007/s00415-019-09652-y)] [Medline: [31797084](#)]
27. Berry JD, Blanchard M, Bonar K, Drane E, Murton M, Ploug U, et al. Epidemiology and economic burden of amyotrophic lateral sclerosis in the United States: a literature review. *Amyotroph Lateral Scler Frontotemporal Degener*. Aug 2023;24(5-6):436-448. [[FREE Full text](#)] [doi: [10.1080/21678421.2023.2165947](https://doi.org/10.1080/21678421.2023.2165947)] [Medline: [36748473](#)]
28. Mehta P, Raymond J, Punjani R, Larson T, Bove F, Kaye W, et al. Prevalence of amyotrophic lateral sclerosis (ALS), United States, 2016. *Amyotroph Lateral Scler Frontotemporal Degener*. May 2022;23(3-4):220-225. [[FREE Full text](#)] [doi: [10.1080/21678421.2021.1949021](https://doi.org/10.1080/21678421.2021.1949021)] [Medline: [34423697](#)]
29. Shallis RM, Wang R, Davidoff A, Ma X, Zeidan AM. Epidemiology of acute myeloid leukemia: recent progress and enduring challenges. *Blood Rev*. Jul 2019;36:70-87. [doi: [10.1016/j.blre.2019.04.005](https://doi.org/10.1016/j.blre.2019.04.005)] [Medline: [31101526](#)]
30. Zhou Y, Huang G, Cai X, Liu Y, Qian B, Li D. Global, regional, and national burden of acute myeloid leukemia, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Biomark Res*. Sep 11, 2024;12(1):101. [[FREE Full text](#)] [doi: [10.1186/s40364-024-00649-y](https://doi.org/10.1186/s40364-024-00649-y)] [Medline: [39256810](#)]
31. Zi H, Liu MY, Luo LS, Huang Q, Luo PC, Luan HH, et al. Global burden of benign prostatic hyperplasia, urinary tract infections, urolithiasis, bladder cancer, kidney cancer, and prostate cancer from 1990 to 2021. *Mil Med Res*. Sep 18, 2024;11(1):64. [[FREE Full text](#)] [doi: [10.1186/s40779-024-00569-w](https://doi.org/10.1186/s40779-024-00569-w)] [Medline: [39294748](#)]
32. Lobo N, Afferi L, Moschini M, Mostafid H, Porten S, Psutka SP, et al. Epidemiology, screening, and prevention of bladder cancer. *Eur Urol Oncol*. Dec 2022;5(6):628-639. [doi: [10.1016/j.euo.2022.10.003](https://doi.org/10.1016/j.euo.2022.10.003)] [Medline: [36333236](#)]
33. Medina A, Mahjoub Y, Shaver L, Pringsheim T. Prevalence and incidence of Huntington's disease: an updated systematic review and meta-analysis. *Mov Disord*. Dec 2022;37(12):2327-2335. [[FREE Full text](#)] [doi: [10.1002/mds.29228](https://doi.org/10.1002/mds.29228)] [Medline: [36161673](#)]
34. Rawlins MD, Wexler NS, Wexler AR, Tabrizi SJ, Douglas I, Evans SJ, et al. The prevalence of Huntington's disease. *Neuroepidemiology*. 2016;46(2):144-153. [doi: [10.1159/000443738](https://doi.org/10.1159/000443738)] [Medline: [26824438](#)]
35. Elbaz A, Carcaillon L, Kab S, Moisan F. Epidemiology of Parkinson's disease. *Rev Neurol (Paris)*. Jan 2016;172(1):14-26. [doi: [10.1016/j.neurol.2015.09.012](https://doi.org/10.1016/j.neurol.2015.09.012)] [Medline: [26718594](#)]
36. Ascherio A, Schwarzschild MA. The epidemiology of Parkinson's disease: risk factors and prevention. *Lancet Neurol*. Nov 2016;15(12):1257-1272. [doi: [10.1016/S1474-4422\(16\)30230-7](https://doi.org/10.1016/S1474-4422(16)30230-7)] [Medline: [27751556](#)]
37. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. *J Neural Transm (Vienna)*. Aug 2017;124(8):901-905. [doi: [10.1007/s00702-017-1686-y](https://doi.org/10.1007/s00702-017-1686-y)] [Medline: [28150045](#)]
38. Abraham DS, Gruber-Baldini AL, Magder LS, McArdle PF, Tom SE, Barr E, et al. Sex differences in Parkinson's disease presentation and progression. *Parkinsonism Relat Disord*. Dec 2019;69:48-54. [[FREE Full text](#)] [doi: [10.1016/j.parkreldis.2019.10.019](https://doi.org/10.1016/j.parkreldis.2019.10.019)] [Medline: [31677455](#)]
39. Reyes-Soffer G, Ginsberg HN, Berglund L, Duell PB, Heffron SP, Kamstrup PR, et al. Lipoprotein(a): a genetically determined, causal, and prevalent risk factor for atherosclerotic cardiovascular disease: a scientific statement from the American Heart Association. *Arterioscler Thromb Vasc Biol*. Jan 2022;42(1):e48-e60. [[FREE Full text](#)] [doi: [10.1161/ATV.000000000000147](https://doi.org/10.1161/ATV.000000000000147)] [Medline: [34647487](#)]
40. Duarte Lau F, Giugliano RP. Lipoprotein(a) and its significance in cardiovascular disease: a review. *JAMA Cardiol*. Jul 01, 2022;7(7):760-769. [doi: [10.1001/jamacardio.2022.0987](https://doi.org/10.1001/jamacardio.2022.0987)] [Medline: [35583875](#)]
41. Edwards JK, Cole SR, Shook-Sa BE, Zivich PN, Zhang N, Lesko CR. When does differential outcome misclassification matter for estimating prevalence? *Epidemiology*. Mar 01, 2023;34(2):192-200. [[FREE Full text](#)] [doi: [10.1097/EDE.0000000000001572](https://doi.org/10.1097/EDE.0000000000001572)] [Medline: [36722801](#)]

Abbreviations

ALS: amyotrophic lateral sclerosis
AML: acute myeloid leukemia
HD: Huntington disease
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Classification of Diseases, 10th Revision
PD: Parkinson disease
RWD: real-world data
RWE: real-world evidence

Edited by A Coristine; submitted 04.May.2026; peer-reviewed by A Bhatti, J Santos; comments to author 19.May.2026; revised version received 07.Sep.2026; accepted 10.Sep.2026; published 30.Sep.2026

Please cite as:

Frisch D, Fronstin M, Zema C

Agentic AI for Clinical Outcomes Research, Population Health Management Analyses With Large Administrative Databases, and Generating Epidemiological Estimates of Diseases: Feasibility and Validation Study

J Med Internet Res 2026;28:e100284

URL: <https://www.jmir.org/2026/1/e100284>

doi: [10.2196/100284](https://doi.org/10.2196/100284)

PMID:

©Daniel Frisch, Michael Fronstin, Carla Zema. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.