Pragmatic and Observational Research

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

INSIGHTS Asthma Pragmatic Registry – A Pragmatic Approach to High-Validity Real-World Evidence for Asthma

Karynsa Kilpatrick[1], Urmila Chandran[2], Robert Urman[1], Jean-Pierre Llanos[3], Nestor A Molfino[4], Rayna K Matsuno[2], Daniel J Riskin[2,5]

[1]Center for Observational Research, Amgen, Inc, Thousand Oaks, California, USA; [2]Verantos Inc, Menlo Park, California, USA; [3]Global Medical Affairs, Amgen, Inc, Thousand Oaks, California, USA; [4]Global Development, Amgen, Inc, Thousand Oaks, California, USA; [5]Stanford University School of Medicine, Stanford, California, USA

Correspondence: Karynsa Kilpatrick, Center for Observational Research, Amgen Inc, Thousand Oaks, CA, 91320, USA, Email karynsah@amgen.com

Purpose: Asthma is a heterogenous disease, and up to 10% of asthma patients have severe disease. While critical, clinical trials in these patients are generally not representative of the broader ambulatory patient population, warranting the need for reliable and rich real-world data to enable clinical, payer, and regulatory decision-making. Currently, observational registries of consented patients rely on longitudinal, clinically rich data to support real-world evidence generation, but are resource intensive. This manuscript describes an asthma pragmatic registry that incorporates rigorous data reliability standards while offering scalability for research.

Patients and Methods: The INSIGHTS Asthma Pragmatic Registry is a real-world, retrospective cohort of adult and adolescent patients with moderate-to-severe asthma geographically distributed throughout the United States. The data set is purposely designed to meet data reliability standards of accuracy, completeness, and traceability. Efforts include continuous improvements in the capture of key clinical features (including disease severity) from electronic health records using advanced technology, linkage across electronic health records, medical and pharmacy claims, and death registry, and process to trace data elements back to a source of truth.

Results: The pragmatic registry implementation includes patients meeting moderate-to-severe asthma eligibility criteria since Jan 1, 2014, with routine data refreshes, resulting in 9,185 patients meeting eligibility criteria through August 1, 2024. The median age of patients is 53 years. At least two-thirds of patients are female, of white race, and non-Hispanic ethnicity. On average, patients were followed for more than five years.

Conclusion: The INSIGHTS Asthma Pragmatic Registry is a new paradigm in observational research, which blends the data reliability and richness of traditional, consented registries while incorporating the flexibility and scale of utilizing routinely collected data. Ongoing efforts include maintaining longitudinality of patient data and improvements to sustain high data reliability per regulatory standards to support this unique data set for asthma research.

Keywords: real-world data, asthma registry, regulatory-grade evidence, electronic health records

Introduction

Asthma is a complex respiratory disease affecting approximately 7.7% of people living in the US.[1] The heterogeneity of the disease results in varying clinical presentations, pathobiology, biomarkers, treatment management and response, and outcomes.[2–6] Patients with severe asthma, defined as asthma requiring maximal optimized high-dose inhaled corticosteroids (ICS) in combination with longer-acting medications or uncontrolled asthma despite use of these medications, make up approximately 10% of all asthma patients.[7,8] Current therapy for severe disease includes inhaled corticosteroids (ICS) plus long-acting bronchodilators with or without oral glucocorticoids. When disease remains uncontrolled, six biologics are approved for add-on treatment for severe asthma in the United States (US), including those indicated for eosinophilic asthma (benralizumab,[9] mepolizumab,[10] and reslizumab),[11] eosinophilic asthma or corticosteroid-dependent asthma (dupilumab),[12] allergic asthma (omalizumab),[13] and asthma regardless of biomarker or phenotype (tezepelumab).[14]

Received: 18 February 2025
Accepted: 12 November 2025
Published: 18 December 2025

Industry-sponsored randomized controlled trials (RCTs), while essential to establishing efficacy and safety of treatments, have restrictive eligibility criteria, and thus limited generalizability in the real-world setting. As diseases and healthcare decisions become increasingly nuanced, real-world evidence (RWE) becomes an essential means of augmenting evidence generated via RCTs. However, not every type of real-world data (RWD) source may be fit-for-purpose to answer clinical questions. For instance, relying solely on structured electronic health records (EHRs) or administrative claims to identify severe asthma patients can be problematic due to inconsistent capture of disease severity codes.[15] Obtaining a comprehensive understanding of the patient journey can also be problematic when there is limited capture of data on symptoms, pulmonary function tests, and other clinical outcomes relevant to asthma.

The inventory of severe asthma registries is growing globally,[16] with approximately 70% of the 37 identified registries being affiliates of the International Severe Asthma Registry (ISAR),[17] currently the largest with over 21,000 adult patients with severe asthma. In the US, CHRONICLE[18] and Severe Asthma Research Program (SARP)[19] are non-ISAR affiliated severe asthma registries enrolling patients across health systems. These traditional registries involve enrollment of consented patients, thus enabling primary collection of data tailored to specific research needs, such as patient-reported outcomes. However, they can have limited scalability (doubling patient count often doubles cost), flexibility (added variables can take years to populate), and efficiency (reaching enrollment targets can require several years).[20–22] There can also be sampling bias resulting from consent requirements.[23,24]

An alternative option is a real-world pragmatic registry that leverages documentation of routine care while organizing retrospective data in a standardized format for a population. A pragmatic registry offers certain advantages over a consented registry, such as enabling better representativeness, larger sample sizes, flexibility to reuse infrastructure, and more rapid adaptation to evolving research needs. Whereas data are primarily collected with strict quality controls in a consented registry, routine data are used in a pragmatic registry. Thus, ensuring data quality is paramount.

Fortunately, data quality standards have been clarified in recent years. The Food and Drug Administration (FDA) published guidance titled "Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision Making for Drug and Biological Products", calling out accuracy, completeness, and traceability as critical aspects of data reliability.[25] The life sciences industry, payers, and providers are increasingly adopting these standards. Thus, for a pragmatic registry to be considered scientifically comparable and a credible alternative to traditional registries, it should have high accuracy, completeness, and traceability.

The INSIGHTS Asthma Pragmatic Registry is implemented as a uniquely rich and high-reliability data set using routinely collected data for epidemiology and clinical research applications in asthma. Leveraging data from health systems in the US, the registry combines artificial intelligence (AI)-based extraction from EHR unstructured data and linkage of the EHR records to pharmacy claims, medical claims, and death registry data on every patient to achieve high levels of accuracy, completeness, and traceability. The purpose of this manuscript is to describe the pragmatic registry design including specific efforts to meet data quality standards and to characterize currently selected patients diagnosed with moderate-to-severe asthma.

Materials and Methods

Design and Setting

The INSIGHTS Asthma Pragmatic Registry is a real-world, retrospective, noninterventional cohort of patients with moderate-to-severe asthma. Patients are identified from a distributed collection of academic and community health systems, spanning over 65 hospitals and approximately 1,600 clinics, covering over 38 million patient visits across different geographies in the US. The registry will continue to expand to additional health systems.

Data Sources

The pragmatic registry includes data from multiple data sources, including EHR data from integrated health systems, medical and pharmacy claims data from claims aggregators, and mortality data from death registry vendors. The EHR data from unstructured notes are extracted using robust AI technology, which includes natural language processing and

machine learned inference as detailed in prior work.[15,26] EHR data for all selected asthma patients are linked to adjudicated claims and death registry data via tokenization. Details on each data source are summarized below:

- EHR structured: Lists such as problem lists, procedure lists, medication lists, unadjudicated claims, and other captured data such as labs.
- EHR unstructured: Primary care notes, specialty notes, radiology and pathology notes, and other narrative content.
- Claims: All adjudicated pharmacy (medication), provider, and facility billing claims.
- Death registry: Data on deaths from any cause sourced from government, private, and historical sources including death certificates, person-reporting, funeral homes, newspapers, and online obituaries. Death data are considered to be sufficiently tracked if at least one year has passed since the last observation date to account for the expected lag in updating the death registry.

Target Population

Patients meeting eligibility criteria for moderate-to-severe asthma (see Table 1) since January 1, 2014, to the most current data available from data sources are selected. Patients are either required to have a receipt of a medium-to-high dose ICS[27] plus an additional controller medication or required to have documentation of an explicit diagnosis of severe asthma by their physician. The index date is the earlier date of meeting either criterion. While the focus of the pragmatic registry is severe asthma, the specified criteria are broad enough to include patients who may have moderate disease at the time of pragmatic registry inclusion, thus enabling the study of disease progression. Additionally, the broad inclusion criteria provide users greater flexibility in defining specific study populations of interest. The data set is refreshed every quarter, with each refresh including additional longitudinal data for existing participants as well as new patients that meet criteria.

Data Reliability

The INSIGHTS Asthma Pragmatic Registry is the first pragmatic registry to not only measure but also require high accuracy, completeness, and traceability.

Accuracy

Accuracy is a quantitative assessment of whether extracted data properly reflects patient characteristics. The quality of data extracted from unstructured EHR using advanced technology is assessed against manual chart abstraction (reference standard) for asthma-related concepts in a subset of data.[15] Manual chart abstraction was conducted by two independent clinical annotators who were blinded to each other's annotations. A minimum inter-rater reliability of 0.70 Cohen's kappa score was required within the manual reference standard. Discrepancies between annotators or ambiguous language were escalated to a clinical informaticist for review and adjudication. The target accuracy level for high validity is an average F1-score (summary measure of recall and precision) of 0.80 or above. Recall, precision, and F1-score are considered to be important metrics to report in healthcare studies evaluating EHR data quality.[28] Recall measures to what extent a concept (such as severe asthma) noted by the physician is also captured by the AI model. Precision measures to what extent a concept identified by the AI model are also noted by the physician. An F1-score of 1 indicates maximum recall and precision, while 0 indicates zero recall and/or precision.[29] Hence, values closer to 1 are desirable. Achieving this

Table 1 Eligibility Criteria for the INSIGHTS Asthma Pragmatic Registry

Among patients with a confirmed asthma diagnosis, those who meet criteria for moderate-to-severe asthma as defined by one of the following: • Use of a medium-to-high dose ICS plus at least one additional controller medication[a] [OR] • Physician-diagnosed severe asthma[b]
Is 12 years of age or older when moderate-to-severe asthma criteria are met

Notes: [a]Additional controller medications include long-acting muscarinic antagonist (LAMA), leukotriene receptor antagonist (LTRA), and long-acting beta agonist (LABA). Patients can receive an ICS combination product (ICS/LABA combination medication) or if they received ICS monotherapy, they also need to receive an additional controller medication within 12 months of receiving ICS monotherapy.[b]Patients must have at least 2 physician-documented occurrences of severe asthma, which can occur in the structured or unstructured data.

high standard requires addition of unstructured data and optimization of AI technology specific to asthma and to each provider organization. To ensure a high level of accuracy for the INSIGHTS Asthma Pragmatic Registry, key concepts were required to have an F1-score of 0.80 or greater. Prior work on accuracy that enabled development of the INSIGHTS Asthma Pragmatic Registry as well as ongoing work has been described previously.[15,30] This methodology is also detailed in prior work related to migraine and cardiovascular disease.[26,31,32]

Completeness

Completeness is a quantitative assessment of inclusion of multiple views into the patient journey. This requires bringing together multiple data sources for each patient, typically via data linkage. The high standard is achieved by requiring linkage of data from structured EHR, unstructured EHR, medical claim, pharmacy claim, and death registry for all patients within the pragmatic registry. To ensure a high level of completeness for the INSIGHTS Asthma Pragmatic Registry, we required a minimum of 12 months of overlap of EHR structured, EHR unstructured, medical claim, and pharmacy claim data.

Traceability

Traceability is a measure of how well individual data elements are proven correct against a patient-specific source of truth. Traceability is achieved by connecting each data element within the dataset to a source of truth. For example, a reference to a medication must be traced to a pharmacy claim. Achieving a high standard of traceability requires maintenance of all data for a patient within a common system, normalization of coding, and identification of granularity changes such as asthma to severe asthma. Since the majority of clinical information about the patient resides in unstructured notes,[33,34] to ensure a high level of traceability for the INSIGHTS Asthma Pragmatic Registry, we required that unstructured data from EHR visits were available for at least half of all structured EHR visits.

Data Standardization

All data from each data source are transformed into a common data model (CDM) based on the Observational Medical Outcomes Partnership (OMOP) CDM[35] to enable standardization of the data structure. Additionally, encoding is normalized to standardized vocabularies.[36] A data quality review framework is used for verification of data received from each data source to ensure consistency in data standards.

Data Collection

The pragmatic registry includes routinely collected data for included patients. Data elements include demographics, conditions, visits, procedures, devices, measurements, observations, drug exposures, insurance coverage, site where care was provided, costs of care, and death from any cause. These concepts are defined in OMOP, with the current version being v20240830.[35] The OMOP model is extended to capture variables extracted from unstructured text via AI models. Each row of data contains several variables related to extracted concepts, including available information on attributes such as severity of a symptom and whether the symptom is experienced currently.

Privacy and Security

The Pragmatic Registry follows the most rigorous protocols available for secondary use of routinely collected healthcare data. Data are de-identified based on expert determination according to 45 C.F.R.§ 164.514: Expert Determination and Safe Harbor prior to merging data sets or adding data into the pragmatic registry.[37] For example, sensitive dates such as date of death are masked to month and year only, and variables such as age are truncated. Tokenization is used for linkage, thus ensuring a one-way matching without the possibility of viewing patient identifiers. Data storage and transmission are compliant with Health Information Trust Alliance (HITRUST) certification. Consistent with HIPAA, informed consent is not required. Additionally, the approach was reviewed by the Salus Institutional Review Board and was granted exemption from the requirements for institutional review board review under 45 CFR 46.104(d)(4(ii)).

Data Analysis

Data elements summarized include patient demographic and clinical characteristics, provider type, labs, and follow-up time. Age and data on biomarkers, including blood eosinophil and total immunoglobulin E (IgE) are summarized based on value at or date closest prior to index, as well as observations throughout study period. The baseline period includes the 12 months prior to and including index date, and follow-up period starts one day after index and ends on the earliest of dataset end (August 1, 2024), last observation, or date of death.

Continuous variables are summarized using mean and median. Categorical variables are summarized using frequency distributions. To contextualize with a US-based and a global registry, a side-by-side description of characteristics of the INSIGHTS Asthma Pragmatic Registry, CHRONICLE,[18,38] and ISAR[17,39,40] are also provided.

Results

Initial Observations

From January 1, 2014, through August 1, 2024, of 29,750 patients with an asthma diagnosis and high-reliability data, defined as data with high accuracy, completeness, and traceability,[25] 9,185 (30.9%) patients met pragmatic registry eligibility criteria. On index date, most patients (n = 8,585; 93.5%) had met eligibility criteria through receipt of asthma medication, 523 (5.7%) patients met criteria by having evidence of physician-documented severe asthma, and 77 patients (0.8%) met both criteria on the same date. Baseline characteristics and summary statistics on follow-up are shown in Table 2. Patients had an average 5.4 years and median 6 years of follow-up.

The average and median ages of patients at the time of meeting eligibility criteria are 51 and 53 years, respectively, with the majority (61.2%) of patients in the age group 40–69 years. Patients are mostly female (65.7%) of white race (65.5%), and non-Hispanic ethnicity (73.5%). Approximately one-fifth of the patients are ever smokers.

Table 2 Characterization of Patients Included in the INSIGHTS Asthma Pragmatic Registry

Characteristic	INSIGHTS Asthma Pragmatic Registry, n = 9,185
Demographics	
Age, years (at index)	
Mean (SD)	51 (17)
Median	53
Min, Max	12, 89
Age category, years (at index)	
12-17	267 (2.9%)
18-29	961 (10.5%)
30-39	1,096 (11.9%)
40-49	1,478 (16.1%)
50-59	2,327 (25.3%)
60-69	1,814 (19.5%)
70-79	883 (9.6%)
≥80	359 (3.9%)
Female, n (%)	6,033 (65.7%)

(*Continued*)

Table 2 (Continued).

Characteristic	INSIGHTS Asthma Pragmatic Registry, n = 9,185
Race, n (%)	
Black	1,096 (11.9%)
White	6,015 (65.5%)
Asian	210 (2.3%)
Other	1,058 (11.5%)
Unknown	806 (8.8%)
Ethnicity, n (%)	
Hispanic or Latino	1,073 (11.7%)
Not Hispanic or Latino	6,755 (73.5%)
Unknown	1,357 (14.8%)
Lifestyle factors	
Documentation of ever smoking, n (%)	1,715 (18.7%)
Clinical characteristics	
Highest level of asthma severity noted for patient by physician during baseline, n (%)	
Mild[a]	1,170 (12.7%)
Mild-to-moderate	55 (0.6%)
Moderate	1,156 (12.6%)
Moderate-to-severe	64 (0.7%)
Severe	667 (7.3%)
Not documented	6,073 (66.1%)
Biomarker data during baseline, closest to index	
Blood eosinophil count (cells/μL)	
Number of patients with measurement[b]	2,933 (31.9%)
Mean (SD) cells/μL	210 (301)
Median cells/μL	100
<150 cells/μL, n (%)	1,611 (54.9%)
≥150 cells/μL, n (%)	1,322 (45.1%)
Total IgE in serum level (IU/mL)	
Number of patients with measurement[b]	103 (1.1%)
Mean (SD) IU/mL	209 (355)
Median IU/mL	56

(*Continued*)

Table 2 (Continued).

Characteristic	INSIGHTS Asthma Pragmatic Registry, n = 9,185
Biomarker data availability during study period	
Blood eosinophil data	
Number of patients with measurement[b]	6,285 (68.4%)
Mean (SD) measurements per patient	12 (20)
Total IgE in serum level data	
Number of patients with measurement[b]	492 (5.4%)
Mean (SD) measurements per patient	2 (2)
Longitudinality of data	
Duration of follow-up, years	
Mean (SD)	5.4 (2.6)
Median	6
Min, Max	0, 11

Notes: [a] Eligible patients include those who qualify based on *either* the medication-based definition or explicit diagnosis of severe asthma by a physician. Inconsistencies in classification by the two methods are reflected in the distribution of highest level of asthma severity noted for patient by physician during baseline.[b]Measurements with available results.
Abbreviations: IgE, immunoglobulin E; SD, standard deviation.

Data Reliability (Accuracy, Completeness, and Traceability)

Data reliability metrics were high across all health systems and are summarized in Table 3. These are stratified by health system, as a proxy for geographic region. Accuracy, as measured by the F1-score, for the concept of severe asthma ranged from 0.80 to 0.90. The median values for traceability, measured as the percentage of structured EHR with corresponding unstructured EHR, ranged from 90% to 100%. The median values for completeness, measured as the number of months of overlap between claims, structured EHR, and unstructured EHR ranged from 37.0 months to 41.7 months. Data reliability metrics for a more comprehensive list of concepts are published in Kilpatrick et al.[15]

Summary of Asthma Registries

Design and characteristics of the INSIGHTS Asthma Pragmatic Registry, CHRONICLE, and ISAR are summarized in Table 4. Differences across the registries include variations in design, frequency of data collection, measurement of data reliability, and eligibility criteria. Registries appear to be similar with respect to select characteristics such as data elements and follow-up duration.

Table 3 Summary of Data Reliability Metrics for the INSIGHTS Asthma Pragmatic Registry

	Health System 1	Health System 2	Health System 3
Accuracy for severe asthma, F1-score[a]	90.0	85.7	80.0
Traceability (%), median (IQR)[b]	100% (85%, 100%)	100% (86%, 100%)	90% (71%, 100%)
Completeness (months), median (IQR)[c]	37.0 (22.7 60.8)	41.7 (25.2, 63.7)	38.8 (23.5, 56.1)

Notes: [a] Accuracy is measured by the F1-score. Requirement is 0.80.[b] Traceability is measured as the percentage of structured electronic health record data that has corresponding unstructured data (source of truth). Requirement is 50%.[c] Overlap in months between claims, structured electronic health record data, and unstructured electronic health record data. Requirement is 12 or more months.

Table 4 Characteristics of the INSIGHTS Asthma Pragmatic Registry, CHRONICLE,[17,30] and ISAR[16,31,32]

Characteristic	INSIGHTS Asthma Pragmatic Registry	CHRONICLE	ISAR
Age range	≥12 years	≥18 years	≥18 years
Type of asthma	Moderate-to-severe	Severe	Severe
Data collection period	2014 – no end date	2018 – no end date	2014 – no end date
Patients, n	9,185 (as of August 2024)	2,793 (as of Feb 2022)	20,113 (as of 2023)
Country	US	US	Global
Data reliability	Reported accuracy, completeness, and traceability metrics	Not reported	Not reported
Data elements collected	Demographics, clinical characteristics, treatments, medical history, measurements, procedures, healthcare resource utilization, clinical outcomes	Demographics, clinical characteristics, treatments, medical history, measurements, procedures, healthcare resource utilization, clinical outcomes, patient reported outcomes	Demographics, clinical characteristics, treatments, medical history, measurements, procedures, healthcare resource utilization, clinical outcomes, patient-reported outcomes
Physician specialty for patient enrollment	No required specialty	Allergists/immunologists, pulmonologists	Allergists/immunologists, pulmonologists
Design	Non-interventional, retrospective	Non-interventional, prospective and retrospective	Non-interventional, prospective and retrospective
Duration of follow-up	Indefinite; Average 5.4 years	Indefinite; target average of 5 years	Indefinite
Timing of data collection	Routinely collected data with longitudinal refresh every quarter	Baseline and at intervals depending on variable (monthly, every 3 months or 6 months)	Baseline, with no mandatory follow-up visit. Data are collected at subsequent visits

Abbreviation: ISAR, International Severe Asthma Registry.

Discussion

This manuscript describes the ongoing US-based INSIGHTS Asthma Pragmatic Registry as a resource-efficient, yet comparably robust alternative to a traditional consented disease registry of moderate-to-severe asthma patients. Approximately 9,200 patients are currently included with patient numbers continuing to grow each quarter. The pragmatic registry is differentiated from other uses of RWD through substantial efforts to achieve high data reliability[25] (including accuracy, completeness, and traceability) and richness, while relying solely on data collected in routine clinical practice. Pragmatic registries may be used independently or combined with primary data collection as required to ensure relevant and reliable data are applied to research questions.

Data Reliability (Accuracy, Completeness, and Traceability)

Accuracy of key clinical variables is increased by measuring and optimizing technology-based extraction from unstructured data. A robust reference standard with minimum inter-rater reliability of 0.70 and high expectations of curated variable accuracy of minimum 0.80 F1-score are among the most rigorous requirements in the industry. Measurement and optimization for each key concept at each local health system represents a critical step to ensure proper application of data reliability.

Completeness is increased by requiring all patients included in the pragmatic registry to have linked EHR (structured and unstructured), medical and pharmacy claims, and death data (as appropriate). The linkage requirement is enhanced to avoid issues with linkage scenarios where a patient's EHR and claims data are years apart. In the pragmatic registry, while every individual had at least 12 months of overlap in EHR structured, EHR unstructured, medical claim, and pharmacy claim data, 82% of patient-years had two or more data sources and 49% of patient-years had three or more data sources.

Traceability is tested for all variables within the pragmatic registry. For example, a medical claim for asthma is traced back to a clinician's note documenting asthma or a subtype. Through tracing, errors in linkage, limitations in care coverage for a given patient, and reimbursement-related upcoding for a given condition become immediately obvious. Hence, each data element included in the pragmatic registry is indicated with the source data from which that record originated. This enables calculation of percent of data elements of interest that can be traced back to the source document, thus demonstrating the suitability of the dataset to meet Health and Human Services audit requirements.

Data Richness

Curation, linkage, and enrichment support rich and reliable data, enabling the extraction of variables that are incompletely captured in claims data such as pulmonary function tests (eg, forced expiratory volume in 1 second [FEV1], forced vital capacity [FVC]), as well as variables that are rarely populated such as disease severity and symptoms. Symptoms such as chest tightness are poorly captured in structured data due to lack of reimbursement incentive to track symptoms.[15] Disease subtypes such as severe asthma are poorly captured in structured data due to lack of granularity required in the reimbursement pathway but may be captured either from curation of stated concepts in unstructured data or enrichment based on multiple linked data sources (eg, asthma diagnosis code, high-dose ICS, and lack of control). The pragmatic registry includes findings, symptoms, granular disease subtypes, and other rich clinical content that is typically absent or rarely captured.[15] These can be critical for research questions which may influence clinical care.

Contextualizing with Severe Asthma Registries

Consistent with demographic distributions observed in severe asthma registries,[18,39] patients are predominantly female, white, and of non-Hispanic ethnicity. A differentiating aspect between the INSIGHTS Asthma Pragmatic Registry and traditional severe asthma registries (such as CHRONICLE and ISAR) is the broader eligibility criteria with respect to diagnosing asthma severity. To generate future RWE for heterogeneous asthma clinical subtypes while ensuring sufficient sample sizes for subgroups, eligibility criteria are broad to enrich with patients meeting the definition of severe asthma according to current guidelines,[41,42] as well as patients with less severe disease. These criteria allow researchers the flexibility to define cohorts based on level of clinical confirmation for severe asthma and research question at-hand. Consequently, some patient characteristics may not be directly comparable with other severe asthma registries with stricter criteria (such as requiring diagnosis by a sub-specialist).

Another differentiating factor is the inclusion of adolescent patients aged 12 years and older. In the US, five of the six approved biologics include adolescents in their label, underscoring the need for high-reliability RWE and long-term safety data in both adult and non-adult populations.[43] Inclusion of adolescent patients and not requiring that patients are being seen by a subspecialist enable a more comprehensive understanding of prescription patterns and disease management across care settings, which could potentially facilitate systems-wide improvements. Finally, unlike other severe asthma registries, the pragmatic registry is purely observational in nature with no intervention to collect non-routine data.

Strengths and Limitations

The main strength of the INSIGHTS Asthma Pragmatic Registry lies in the combination of high-reliability data with the flexible, scalable, and efficient use of data collected in routine clinical practice. Additionally, the high standards for data reliability are maintained as the registry is expanded to include increased data capture. Patients are selected from a network of academic and community inpatient and outpatient facilities across various geographies in the US and spanning specialist and non-specialist clinics, increasing the generalizability of the asthma patient population in the dataset. The observed longitudinality of the pragmatic registry is comparable to other registries[18] and sufficient to enable long-term treatment safety and effectiveness assessments. The dataset reflects real world care without the biases inherent in selective trial enrollment or registry consent.

Limitations of the pragmatic registry largely reflect the reliance on data collected through routine clinical care. Patients may not visit a physician if they are feeling well and information during that period may not be recorded. Similarly, a physician will not perform every test or document every finding on every encounter. Patients may move into and out of a given health system, resulting in both left and right censoring of the patient journey. Certain clinical

outcomes of interest, such as asthma exacerbations, are not explicitly captured or extracted. However, in the absence of such data, algorithms can be applied to claims and/or structured EHR to analytically define exacerbations. Finally, the greatest limitation of the pragmatic registry is a lack of data elements that are not part of routine care and require patient consent. This includes patient-reported outcomes and structured assessments that are not captured at a frequent cadence in practice, such as quality of life measures and Asthma Control Test scores. This limitation, however, makes the pragmatic registry a truly real-world data set.

Conclusion

The INSIGHTS Asthma Pragmatic Registry can meet the demand for clinically rich, highly valid, and timely RWE as the data set is specifically developed to meet data reliability standards while continuing to offer the advantages of leveraging data collected in routine clinical practice. By incorporating the strengths of traditional registries in obtaining clinically nuanced data but doing so in a more scalable and efficient manner, the pragmatic registry presents a new paradigm in generating high validity RWE for clinical and epidemiological research.

Acknowledgments

The authors would like to thank Ms. Erin Murray and Dr. Robert Kalfus from Verantos Inc for providing project management and clinical informatics support, respectively, for the pragmatic registry.

The work in this manuscript was supported by funding from Amgen Inc.

The INSIGHTS Asthma Pragmatic Registry is not publicly available and follows a subscription model.

Disclosure

KK, RU, JPL, and NM are employees and stockholders of Amgen Inc. UC reports that this work was funded by support from Amgen Inc. DR and UC are employees and stockholders of Verantos Inc. DR reports Commercial engagement with Amgen Inc. The authors report no other conflicts of interest in this work.

References

1. "Most recent national asthma data | CDC". Available from: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm. Accessed 2024 February 2024.
2. Chen W, Reddel HK, FitzGerald JM, Beasley R, Janson C, Sadatsafavi M. Can we predict who will benefit most from biologics in severe asthma? A post-hoc analysis of two Phase 3 trials. *Respir Res*. 2023;24(1):120. doi:10.1186/s12931-023-02409-2
3. Papi A, Blasi F, Canonica GW, Morandi L, Richeldi L, Rossi A. Treatment strategies for asthma: reshaping the concept of asthma management. *Allergy Asthma Clin Immunol*. 2020;16:75. doi:10.1186/s13223-020-00472-8
4. Bostantzoglou C, Delimpoura V, Samitas K, Zervas E, Kanniess F, Gaga M. Clinical asthma phenotypes in the real world: opportunities and challenges. *Breathe*. 2015;11(3):186–193. doi:10.1183/20734735.008115
5. Trevor J, Lugogo N, Carr W, et al. Severe asthma exacerbations in the United States:: incidence, characteristics, predictors, and effects of biologic treatments. *Ann Allergy Asthma Immunol*. 2021;127(5):579–587.e1. doi:10.1016/j.anai.2021.07.010
6. FitzGerald JM, Barnes PJ, Chipps BE, et al. The burden of exacerbations in mild asthma: a systematic review. *ERJ Open Res*. 2020;6 (3):00359–2019. doi:10.1183/23120541.00359-2019
7. Rönnebjerg L, Axelsson M, Kankaanranta H, et al. Severe asthma in a general population study: prevalence and clinical characteristics. *J Asthma Allergy*. 2021;16(14):1105–1115. doi:10.2147/JAA.S327659
8. Chung KF, Wenzel SE, Brozek JL, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J*. 2014;43(2):343–373. doi:10.1183/09031936.00202013
9. "AstraZeneca pharmaceuticals. prescribing information for benralizumab" April 2024. Available from: https://den8dhaj6zs0e.cloudfront.net/50fd68b9-106b-4550-b5d0-12b045f8b184/3647bed4-ce91-4fe7-9bc5-32dbee73f80a/3647bed4-ce91-4fe7-9bc5-32dbee73f80a_viewable_rendition__v.pdf. [Accessed June 25, 2024.].
10. GlaxoSmithKline LLC. Prescribing information for mepolizumab." March 2023. Available from: https://gskpro.com/content/dam/global/hcpportal/en_US/Prescribing_Information/Nucala/pdf/NUCALA-PI-PIL-IFU-COMBINED.PDF. Accessed June 25, 2024.
11. Teva Respiratory, LLC. Prescribing information for reslizumab. Available from: https://www.cinqair.com/globalassets/cinqair/prescribinginformation.pdf. [Accessed June 25, 2024.].
12. "Regeneron pharmaceuticals, inc. prescribing information for dupilumab" January 2024. Available from: https://www.regeneron.com/downloads/dupixent_fpi.pdf. Accessed June 25, 2024.
13. Genentech USA. Inc. prescribing information for omalizumab." 2024. Available from: https://www.gene.com/download/pdf/xolair_prescribing.pdf. Accessed June 25, 2024.
14. "Amgen inc and astrazeneca. prescribing information for tezepelumab" May 2023. Available from: https://den8dhaj6zs0e.cloudfront.net/50fd68b9-106b-4550-b5d0-12b045f8b184/e306dc06-d580-4457-b15f-9f28545ad63a/e306dc06-d580-4457-b15f-9f28545ad63a_viewable_rendition__v.pdf. [Accessed June 25, 2024.].

15. Kilpatrick K, Cahill K, Chandran U, Riskin D. Advanced approaches to generating high-validity real-world evidence in asthma. *Epidemiology*. 2025;36(1):20–27. doi:10.1097/EDE.0000000000001803
16. Cushen B, Koh MS, Tran TN, et al. Adult severe asthma registries: a global and growing inventory. *Pragmat Obs Res*. 2023;14:127–147. doi:10.2147/POR.S399879
17. "International Severe Asthma Registry. ISAR | Collaborative Global Asthma Registry. ISAR." 2023. Available from: https://www.isar.opcglobal.org. [Accessed August 30, 2024.].
18. Ambrose CS, Chipps BE, Moore WC, et al. The CHRONICLE study of US adults with subspecialist-treated severe asthma: objectives, design, and initial results. *Pragmat Obs Res*. 2020;11:77–90. doi:10.2147/POR.S251120
19. "National Heart, Lung & Blood Institute. Severe Asthma Research Program." 2019. Available from: http://www.severeasthma.org/home.html. [Accessed September 5, 2024.].
20. Tangka FK, Subramanian S, Beebe MC, et al. Cost of operating central cancer registries and factors that affect cost: findings from an economic evaluation of centers for disease control and prevention national program of cancer registries. *J Public Health Manag Pract*. 2016;22(5):452–460. doi:10.1097/PHH.0000000000000349
21. Delaunay C. Registries in orthopaedics. *Orthop Traumatol Surg Res*. 2015;101, suppl. 1:S69–S75. doi:10.1016/j.otsr.2014.06.029
22. Gliklich RE. N.A. Registries for evaluating patient outcomes: a user's guide. Rockville (MD): Agency for Healthcare Research and Quality (US). Planning a Registry. Available from: https://www.ncbi.nlm.nih.gov/books/NBK208631/. [Accessed October 22, 2025.].
23. "Food and drug administration. real-world data: assessing registries to support regulatory decision-making for drug and biological products guidance for industry" December 2023. Available from: https://www.fda.gov/media/154449/download. [Accessed October 20, 2025.].
24. Wang K, Ftang CN, Jacobsen RB, Årøen A. Review of response rates over time in registry-- based studies using patient-- reported outcome measures. *BMJ Open*. 2020;10:e030808. doi:10.1136/bmjopen-2019-030808
25. "Food and drug administration. real-world data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products" July 2024. Available from: https://www.fda.gov/media/152503/download. [Accessed May 20, 2025.].
26. Hernandez-Boussard T, Monda KL, Crespo BC, Riskin D. Real world evidence in cardiovascular medicine: ensuring data validity in electronic health record-based studies. *J Am Med Inform Assoc*. 2019;26(11):1189–1194. doi:10.1093/jamia/ocz119
27. "Global Initiative for Asthma. Global strategy for asthma management and prevention (2022 Update)," 2022. Available from: https://ginasthma.org/wp-content/uploads/2022/07/GINA-Main-Report-2022-FINAL-22-07-01-WMS.pdf. [Accessed May 30, 2025.].
28. Lee S, Doktorchik C, Martin EA, et al. Electronic medical record-based case phenotyping for the Charlson conditions: scoping review. *JMIR Med Inform*. 2021;9(2):e23934. doi:10.2196/23934
29. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. 2022;12(1):5979. doi:10.1038/s41598-022-09954-8
30. Riskin DJ, Monda KL, Gagne JJ, et al. Implementing accuracy, completeness, and traceability for data reliability. *JAMA Netw Open*. 2025;8(3):e250128. doi:10.1001/jamanetworkopen.2025.0128
31. Garan AR, Monda KL, Dent-Acosta RE, Riskin DJ, Gluckman TJ. Retrospective comparison of traditional and artificial intelligence-based heart failure phenotyping in a US health system to enable real-world evidence. *BMJ Open*. 2023;13(8):e073178. doi:10.1136/bmjopen-2023-073178
32. Riskin DJ, Cady R, Shroff A, Hindiyeh NA, Smith T, Kymes S. Using artificial intelligence to identify patients with migraine and associated symptoms and conditions within electronic health records. *BMC Med Inform Decis Mak*. 2023;23(1):121. doi:10.1186/s12911-023-02190-8
33. Negro-Calduch E, Azzopardi-Muscat N, Krishnamurthy RS, Novillo-Ortiz D. Technological progress in electronic health record system optimization: systematic review of systematic literature reviews. *Int J Med Inform*. 2021;152:104507. doi:10.1016/j.ijmedinf.2021.104507
34. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. *Yearb Med Inform*. 2014;9(1):14–20. doi:10.15265/IY-2014-0020
35. "OMOP common data model" [Available from: https://ohdsi.github.io/CommonDataModel/. [Accessed March 20, 2024.].
36. Athena. Available from: https://athena.ohdsi.org/search-terms/start. [Accessed March 20, 2024.].
37. "US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.". Available from: https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html#safeharborguidance. [Accessed May 30, 2025.].
38. Ledford DK, Soong W, Carr W, et al. Real-world severe asthma biologic administration and adherence differs by biologic: CHRONICLE study results. *Ann Allergy Asthma Immunol*. 2023;131(5):598–605.e3. doi:10.1016/j.anai.2023.07.017
39. Wang E, Wechsler ME, Tran TN, et al. Characterization of severe asthma worldwide: data from the international severe asthma registry. *Chest*. 2020;157(4):790–804. doi:10.1016/j.chest.2019.10.053
40. FitzGerald JM, Tran TN, Alacqua M, et al. International severe asthma registry (ISAR): protocol for a global registry. *BMC Med Res Methodol*. 2020;20(1):212. doi:10.1186/s12874-020-01065-0
41. "Global Initiative for Asthma. Global strategy for asthma management and prevention (2024 update)," May 2024. Available from: https://ginasthma.org/wp-content/uploads/2024/05/GINA-2024-Strategy-Report-24_05_22_WMS.pdf. [Accessed July2, 2024.].
42. "National Institutes of Health. Expert panel report 3: guidelines for the diagnosis and management of asthma. NIH Publication. Available from: https://www.nhlbi.nih.gov/sites/default/files/media/docs/EPR-3_Asthma_Full_Report_2007.pdf. [Accessed July 2, 2024.].
43. Bacharier LB, Jackson DJ. Biologics in the treatment of asthma in children and adolescents. *J Allergy Clin Immunol*. 2023;151(3):581–589. doi:10.1016/j.jaci.2023.01.002