Generating real-world evidence in asthma: Advanced approaches to achieve high validity

Karynsa Kilpatrick,[1] Keri L Monda,[1] Daniel Riskin[2]

[1]Center for Observational Research, Amgen Inc., Thousand Oaks, CA; [2]Verantos Inc., Menlo Park, CA

Background

- Regulatory, market access, and prescribing decisions increasingly rely on insights derived from analyses of health data collected in routine care, i.e., real-world evidence (RWE)
- As RWE can potentially affect change in clinical practice, underlying data quality is necessarily scrutinized
- To quantitatively understand data quality, rigorous approaches are required, particularly in research questions intended to influence the standard of care and in diseases that are phenotypically complex and difficult to define in real-world data

Purpose

- In this study, asthma was used as a testbed to compare data quality between Traditional and Advanced RWE approaches (***defined below***)

Methods

Study Design

- A retrospective analysis leveraging electronic health record (EHR) primary encounter data from a tertiary care academic medical center between 2018 and 2020

Evaluation of Traditional vs Advanced RWE Approaches

- A set of predefined asthma-related clinical concepts (or features) was extracted using:
  - ***Traditional RWE*** approach: Structured EHR using traditional query techniques
  - ***Advanced RWE*** approach: Unstructured EHR data (full clinical narrative) using natural language processing (NLP) and artificial intelligence (AI)-based inference (machine learning)
- Performance of each approach was evaluated against a reference standard manually created via chart abstraction (annotated by two independent clinical annotators)
  - A minimum inter-rater reliability of 0.8 Cohen's kappa score was required to consider the manual reference sufficient
- Accuracy of each approach was measured as recall, precision, and F1-score (**Table 1**)

Table 1. Accuracy measures to compare Traditional and Advanced RWE approaches with a manual reference standard

Condition based on proposed RWE approach (Advanced or Traditional)	Condition based on manual reference standard	
	Yes	No
Yes	True positive (A)	False positive (B)
No	False negative (C)	True negative (D)
Recall (sensitivity) = A/(A+C) Proportion correctly identified (true positives) among those that should have been identified (true positives + false negatives)		
Precision (positive predictive value) = A/(A+B) Proportion correctly identified (true positives) among all those that were identified (true positives + false positives)		
F1-score = 2 × ([precision × recall] / [precision + recall]) Summary score (harmonic mean) of precision and recall		

Methods (contd.)

Statistical Analysis

- A two-sided p-value of 0.05 and a Chi-squared test were used to compare the statistical differences between the Advanced and Traditional RWE approaches
- Success of the technology was defined as either average recall and precision >80%, or accuracy of Advanced RWE ≥25% higher than Traditional RWE as measured by F1-score

Asthma-related Features Evaluated

- A set of features were selected *a priori*, encompassing conditions, comorbidities, symptoms, findings, and procedures considered to be important in asthma and with varying levels of phenotypic complexity (**Table 2**); list not considered comprehensive but selected to provide sufficient visibility, ensuring proper choice of approach and highlight areas requiring better technology

Table 2. Asthma-related features evaluated

Category	Feature	
Conditions	Asthma Moderate asthma Eosinophilic asthma	Mild asthma Severe asthma
Comorbidities	Chronic obstructive pulmonary disease (COPD) Gastroesophageal reflux disease (GERD) Interstitial lung disease	Bronchitis Smoking
Symptoms	Chest tightness Dyspnea	Cough Wheezing
Findings	Forced expiratory volume (FEV) Forced vital capacity (FVC)	Functional residual capacity (FRC)
Procedures	Pulmonary function test (PFT)	

Results

- A total of 18 asthma-related features were abstracted from 6,037 encounters among 3,481 patients
- Across these 18 features, there was a 43% absolute increase and 82% relative increase in the average F1-score between the Traditional and Advanced RWE approaches (**Table 3**), with a *p*-value <0.001 for all features (**Table 4**)
- Cohen's kappa score of 0.8 indicated inter-rater reliability, reflecting a credible reference standard

Table 3. Average recall, precision, and F1-score for Traditional and Advanced RWE approaches across the 18 asthma-related features evaluated

Traditional RWE			Advanced RWE			*P*-value for difference in F1-score	% change in F1-score from Traditional to Advanced approach	
Average recall (%)	Average precision (%)	Average F1-score (%)	Average recall (%)	Average precision (%)	Average F1-score (%)		Absolute	Relative
40.7	72.4	52.1	95.6	93.8	94.7	< 0.001	43%	82%

Results (contd.)

Table 4. Recall, precision, and F1-score for Traditional and Advanced RWE approaches for each asthma-related feature evaluated

	Traditional RWE			Advanced RWE			*P*-value for difference in F1-score	% change in F1-score from Traditional to Advanced approach	
	Recall (%)	Precision (%)	F1-score (%)	Recall (%)	Precision (%)	F1-score (%)		Absolute	Relative
Condition									
Asthma	78.2	91.8	84.5	97.2	96.2	96.7	<0.001	12%	14%
Mild asthma	15.1	22.9	18.2	86.5	87.9	87.2	<0.001	69%	379%
Moderate asthma	40.4	25.8	31.5	94.7	96.4	95.6	<0.001	64%	203%
Severe asthma	32.8	35.5	34.1	89.6	90.9	90.2	<0.001	56%	165%
Eosinophilic asthma	0.0	N/A	0.0	75.0	100.0	85.7	<0.001	86%	N/A
Comorbidities									
COPD	69.3	87.1	77.2	96.4	92.3	94.3	<0.001	17%	22%
Bronchitis	29.4	80.0	43.0	96.8	89.0	92.8	<0.001	50%	116%
Smoking	29.8	90.8	44.9	91.1	91.1	91.1	<0.001	46%	103%
Interstitial lung disease	42.9	84.2	56.8	85.7	80.7	83.1	<0.001	26%	46%
GERD	69.0	89.3	77.9	97.3	92.8	95.0	<0.001	17%	22%
Symptoms									
Chest tightness	0.0	N/A	0.0	87.6	95.8	91.5	<0.001	92%	N/A
Cough	22.1	79.8	34.6	96.2	96.3	96.3	<0.001	62%	178%
Dyspnea	28.9	72.7	41.4	95.5	95.5	95.5	<0.001	54%	131%
Wheezing	2.2	50.0	4.2	95.4	94.6	95.0	<0.001	91%	2162%
Findings									
FEV	0.0	N/A	0.0	99.1	97.9	98.5	<0.001	99%	N/A
FRC	0.0	N/A	0.0	96.2	100.0	98.0	<0.001	98%	N/A
FVC	0.0	N/A	0.0	98.4	96.6	97.5	<0.001	98%	N/A
Procedures									
PFT	24.5	67.3	35.9	96.1	87.8	91.8	<0.001	56%	156%

- There was considerable heterogeneity in accuracy by concept in the Traditional RWE approach (**Table 4**), whereas the Advanced RWE approach consistently met success criteria and demonstrated higher measures of accuracy, overall and across features
- As an example for conditions, there was a 56% absolute increase and 165% relative increase in the F1-score between the Traditional and Advanced approaches in identifying severe asthma; the Advanced approach was also able to measure accuracy of phenotypic concepts unavailable in the structured fields of EHR (e.g., eosinophilic asthma)
- For the comorbidities evaluated, while the Traditional approach had ≥80% precision across comorbidities and recall ranged from 29.4% (interstitial lung disease) to 69.3% (COPD), the Advanced approach led to a 27%–67% absolute increase and 39%–229% relative increase in precision across these comorbidities, with all accuracy measures being ≥80%
- Differences in accuracy between the two approaches were more pronounced for symptoms, findings, and procedures, with a 54%–99% absolute increase and 131%–2162% relative increase in the F1-score between the two approaches

Key Takeaway

High-validity RWE in asthma from routinely collected real-world data is possible and could enable the generation of increasingly reliable evidence to support healthcare decision making and improve patient care.

Conclusions

- Results demonstrated that using narrative data and optimized AI technologies, advanced approaches resulted in higher measures of recall, precision, and F-1 score compared to traditional approaches in an academic medical center.
- Advanced approaches are particularly beneficial when there is a need to define clinical concepts, such as disease subtypes and symptoms, that are poorly or inaccurately represented within structured fields of the EHR.
- A potential limitation of this study is that EHR data from a tertiary care medical center (i.e., a specialized care setting) were used and may not be generalizable to other healthcare settings.
- Data type and technology approach led to highly variable data quality, highlighting the importance of measuring accuracy for key variables once a specific approach is selected.
- Heterogeneity in data quality across clinical concepts also underscores the importance of ensuring a fit-for-purpose design, where required data quality and relevance depend on the research question of interest and to what extent the resulting evidence may inform or influence the standard of care via clinical, regulatory, or reimbursement decision-making.

Disclosures & Funding

This study was sponsored by Amgen Inc. KK and KLM are employees and stockholders of Amgen Inc. DR is the Founder and Chief Executive Officer of Verantos Inc. Medical writing support was funded by Amgen Inc. and provided by Tim Harrision, PharmD (Amgen Inc.), Kate Smigiel, PhD (Amgen Inc.), and Meenakshi Mukherjee, PhD (Cactus Life Sciences).

Presented at the CHEST Annual Meeting 2023, October 8–11, Honolulu, Hawai'i, USA