Chronic Obstructive Pulmonary Diseases:
Journal of the COPD Foundation®

Brief Report

Using Electronic Health Records to Identify Participants for a Multicenter COPD Clinical Trial

Priya A. Prasad, PhD, MPH[1] Stephanie L. LaBedz, MD[2] Janet Holbrook, PhD[3] Richard A. Mularski, MD, MSHS, MCR[4] Stephen I. Rennard, MD[5] Elizabeth Sugar, PhD[3] Robert A. Wise, MD[6] Nina E. Bracken, APRN[2] Andrew Auerbach, MD, MPH[1] Jerry A. Krishnan, MD, PhD[2] for the RELIANCE study team

1. Division of Hospital Medicine, University of California, San Francisco, California, United States
2. Department of Medicine, University of Illinois Chicago, Chicago, Illinois, United States
3. Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, United States
4. Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon, United States
5. Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, United States
6. Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland, United States

Address correspondence to:

Priya A. Prasad, PhD, MPH
University of California, San Francisco
521 Parnassus Ave, Box 0131
San Francisco, CA 94143
Phone: (215) 964-0200
Email: Priya.Prasad@ucsf.edu

Keywords:

decentralized trial; comparative effectiveness; translational science; COPD

This article has an online supplement.

Abbreviations:

AI=artificial intelligence; **AMI**=acute myocardial infarction; **CI**=confidence interval; **COPD**=chronic obstructive pulmonary disease; **EHR**=electronic health record; **HF**=heart failure; **ICD-10**=International Classification of Diseases, 10th Revision; **ICD-10-CM**=International Classification of Diseases, 10th Revision, Clinical Modification; **ICS**=inhaled corticosteroid; **IRB**=institutional review board; **LABA**=long-acting beta2-agonist; **LAMA**=long-acting muscarinic antagonist; **PCORI**=Patient-Centered Outcomes Research Institute; **RELIANCE**=Roflumilast or Azithromycin to Prevent COPD Exacerbations study

Funding Support:

This work was supported by Patient-Centered Outcomes Research Institute (PCORI) Award PCS-1504-30430. The statements in this publication are solely the responsibility of the authors and do not necessarily represent the views of PCORI, its Board of Governors, or its Methodology Committee

Citation:

Prasad PA, LaBedz SL, Holbrook J, et al. Using electronic health records to identify participants for a multicenter COPD clinical trial. *Chronic Obstr Pulm Dis.* 2026;13(6):523-528. doi: https://doi.org/10.15326/jcopdf.2026.0791

Publication Dates:

Date of Acceptance: September 3, 2026
Published Online Date: September 30, 2026

Introduction

Randomized clinical trials are essential for establishing the benefits and risks of medical interventions but are slowed by inefficient methods for identification and recruitment of eligible participants. Electronic health records (EHRs) include structured clinical data stored in predefined, discrete fields with constrained values (e.g., International Classification of Diseases, 10th Revision [ICD-10] diagnosis codes) and unstructured data (free-text or nontabular content with no predefined schema, e.g., symptoms documented in a clinic note). Computer-executed queries using structured and unstructured data in EHRs (computable phenotypes)[1] represent an opportunity to accelerate the identification of potentially eligible patients as they could supplement or replace manual review of the same information. However, limited data exists about the utility of computable phenotypes across diverse research use cases.

The Roflumilast or Azithromycin to Prevent COPD Exacerbations (RELIANCE) phase 3 multicenter clinical trial (NCT04069312) was designed to compare the harms and benefits of long-term use of roflumilast versus azithromycin in patients with chronic obstructive pulmonary disease (COPD) recently hospitalized for an exacerbation.[2,3] During the start-up phase of the RELIANCE clinical trial, the study funder (Patient-Centered Outcomes Research Institute

For personal use only. Permission required for all other uses.

[PCORI]), requested an evaluation of the feasibility of an EHR-derived computable phenotype to identify potentially eligible participants. To inform the development of a candidate computable phenotype that would incorporate inclusion and exclusion criteria for the RELIANCE trial, a pragmatic clinical trial conducted within real-world clinical infrastructure, we conducted a study to estimate the number of inpatient and outpatient encounters of patients with COPD across enrolling sites. Then, in a random sample of inpatient and outpatient encounters at each enrolling site, manual review of the EHR was performed to assess the adequacy of documentation to evaluate each of the inclusion and exclusion criteria. Preliminary results were presented in the form of an abstract.[4]

Methods

We conducted a cross-sectional study across 33 U.S. clinical sites during the start-up phase of the RELIANCE trial, including 31 academic health centers, one integrated managed care organization, and one private practice. All 4 U.S. Census regions were represented by the various clinical sites (West, 7 sites; South, 8 sites; Midwest, 13 sites; Northeast, 5 sites). This study was reviewed by the institutional review board (IRB) at the University of Illinois Chicago (IRB #17102001) and was conducted with a waiver of consent. Each participating site obtained local IRB approval to conduct the data collection activities.

Each site used its local EHR to identify adults aged ≥40 years with COPD, defined by ICD-10 (Table 1)[5] codes consistent with the Centers for Medicare & Medicaid Services Hospital Readmissions Reduction Program COPD definition. EHR queries were applied to the inpatient and outpatient clinical settings:

1. Inpatient setting: adults ≥40 years hospitalized for COPD exacerbation in the prior 12 months;
2. Outpatient setting: adults ≥40 years with at least one outpatient COPD encounter in the prior 12 months.

A random sample of 15 inpatient encounters and 15 outpatient encounters identified by the EHR queries was selected at each site (30 per site, total n=990 encounters) for manual chart abstraction to assess whether patients met the following key eligibility criteria for the RELIANCE trial: (1) age ≥40 years; (2) ≥10 pack-years smoking; (3) chronic bronchitis; (4) hospitalization for COPD exacerbation within 12 months; (5) prescription for inhaled maintenance therapy (long-acting muscarinic antagonist [LAMA], long-acting beta-2-agonist [LABA], dual LABA and LAMA, inhaled corticosteroid (ICS) plus LABA, or triple therapy (LABA, LAMA, and ICS); (6) English or Spanish language; (7) not pregnant; (8) no long-term (>30 days) azithromycin or roflumilast use; and (9) no contraindications to either drug (e.g., documented allergy). A criterion was considered "indeterminate" if there was insufficient documentation in the EHR to support the criterion being met or not met (e.g., documentation of hospitalization at outside hospital, but inadequate information about dates of hospitalization). We did not assess eligibility criteria that would require a conversation with the potential study participant (i.e., willingness to provide (1) informed consent, (2) contact telephone number, and (3) social security or other health insurance claims number to study staff). The chart abstraction tool is included in the online supplement. We calculated the proportion of records meeting all criteria ("potentially eligible") and those that were indeterminate for ≥1 criteria. Binomial 95% confidence intervals (CIs) were computed for proportions.

Results

Across all 33 sites, we identified 104,097 inpatient encounters for COPD exacerbations and 160,290 outpatient encounters for patients with COPD over 12 months. In the random sample of 990 manually abstracted records, EHR documentation was insufficient to evaluate >1 eligibility criterion in 28% (95% CI, 19%–40%) and 27% (95% CI, 19%–37%) of inpatient and outpatient encounters, respectively (Figure 1). The mean proportion of potentially eligible patients was 36% (95% CI, 27%–45%) in the inpatient setting and 7% (95% CI, 5%–11%) in the outpatient setting.

The 3 criteria most frequently considered indeterminate in inpatient and outpatient encounters were: (1) chronic bronchitis (18% and 14%, respectively); (2) smoking history ≥10 pack years (13% and 10%, respectively); and (3) prior COPD hospitalization within 12 months (6% and 10%, respectively).

Discussion

Across 33 clinical sites that participated in the start-up phase of the RELIANCE clinical trial, we identified over 100,000 inpatient encounters and over 160,000 outpatient encounters in patients with COPD. Manual review of the EHR in a random sample of nearly 1000 of these encounters suggests that approximately 1 in 3 inpatients with COPD exacerbations met all RELIANCE trial criteria. Far fewer outpatients with COPD met trial eligibility criteria. We also observed variability in the completeness of documentation relevant to assessing study eligibility, with missing information most common for the number of pack years smoked and a diagnosis of chronic bronchitis.

These findings build on our previous work that indicates poor agreement between patient-reported physician diagnosis of COPD and a COPD reference standard based on smoking exposure and confirmation of postbronchodilator airflow limitation.[6] The current study, which is focused on the RELIANCE clinical trial, highlights the importance

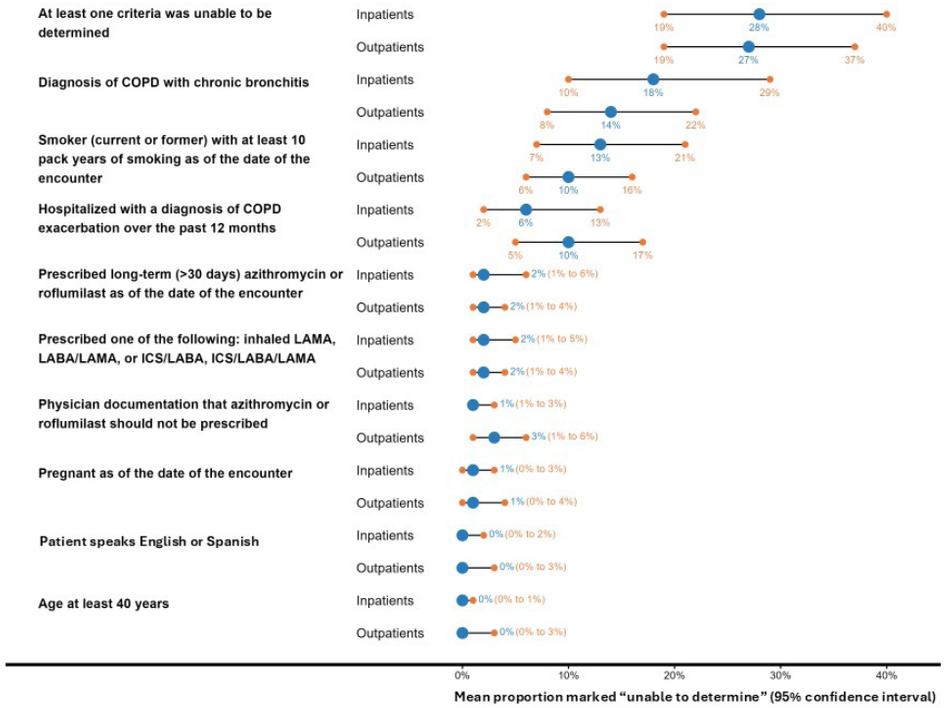
For personal use only. Permission required for all other uses.

Table 1. International Classification of Diseases-Tenth Revision Codes Used to Identify Patients With COPD

Principal Diagnosis of COPD (ICD-10-CM codes):	J41.8 Mixed simple and mucopurulent chronic bronchitis J42 Unspecified chronic bronchitis J43.0 Unilateral pulmonary emphysema (MacLeod's syndrome) J43.1 Panlobular emphysema J43.2 Centrilobular emphysema J43.8 Other emphysema J43.9 Emphysema, unspecified J44.0 Chronic obstructive pulmonary disease with acute lower respiratory infection J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation J44.9 Chronic obstructive pulmonary disease, unspecified	Source: 2017 Condition-Specific Readmission Measure Updates and Specifications Report: AMI, HF, Pneumonia, COPD, and Stroke (Appendix D.2, page 81)[5]
OR Principal Diagnosis of Respiratory Failure	J96.00 Acute respiratory failure, unspecified whether with hypoxia or hypercapnia J96.01 Acute respiratory failure with hypoxia J96.02 Acute respiratory failure with hypercapnia J96.20 Acute and chronic respiratory failure, unspecified whether with hypoxia or hypercapnia J96.21 Acute and chronic respiratory failure with hypoxia J96.22 Acute and chronic respiratory failure with hypercapnia J96.90 Respiratory failure, unspecified, unspecified whether with hypoxia or hypercapnia J96.91 Respiratory failure, unspecified with hypoxia J96.92 Respiratory failure, unspecified with hypercapnia R09.2 Respiratory Arrest	Source: 2017 Condition-Specific Readmission Measure Updates and Specifications Report: AMI, HF, Pneumonia, COPD, and Stroke (Appendix D.2, page 81) [5]
PLUS A Secondary Diagnosis of J44.0 or J44.1:	**PLUS** **2. A secondary diagnosis in any position (either):** J44.0 Chronic obstructive pulmonary disease with acute lower respiratory infection J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation	

COPD=chronic obstructive pulmonary disease; ICD-10-CM=International Classification of Diseases, 10th Revision, Clinical Modification; AMI=acute myocardial infarction; HF=heart failure

Figure 1. Proportion of Patients With Indeterminate Eligibility Using a Computable Phenotype Developed for the Roflumilast or Azithromycin to Prevent COPD Exacerbations Clinical Trial



In a random sample of 990 encounters, manual review of EHR (both structured and unstructured data) identified gaps in documentation relevant to evaluating 1 or more eligibility criteria for the RELIANCE trial in a mean of 28% of encounters in the inpatient stratum and 27% of encounters in the outpatient stratum. Gaps in documentation were most common for a diagnosis of chronic bronchitis and pack years smoked.

COPD=chronic obstructive pulmonary disease; LAMA=long-acting muscarinic antagonist; LABA=long-acting beta2-agonist; ICS=inhaled corticosteroid; EHR=electronic health record; RELIANCE=Roflumilast or Azithromycin to Prevent COPD Exacerbations

For personal use only. Permission required for all other uses.

of reviewing structured and unstructured data in the EHR as well as addressing gaps in clinical documentation when developing a computable phenotype to identify trial-eligible patients. Variation in data completeness across eligibility criteria for the RELIANCE clinical trial, especially information to identify patients with chronic bronchitis and >10 pack-years smoking history, highlights the current limitations of EHR data. The potential to further improve computable phenotype performance using natural language processing or artificial intelligence (AI)-enabled data extraction tools warrants further study.[7,8]

From a translational science perspective,[9] this study illustrates the potential and gaps in EHR-based screening to improve the efficiency of multicenter clinical trials compared to manually searching through all encounters each day. Investments by the National Institutes for Health's National Center for Advancing Translational Sciences and PCORI in health data infrastructure (e.g., PCORnet, N3C RECOVER)[10,11] provide an opportunity to refine and scale these methods in other conditions, including clinical trials in interstitial lung disease where more efficient approaches to identifying patients eligible for clinical trials are needed. We do, however, acknowledge study limitations. The study was not designed to evaluate the accuracy of COPD diagnosis at clinical trial sites. Nearly all clinical sites that contributed to the study were academic health centers; findings in this report may or may not apply to nonacademic health centers. The documentation in the EHR for a modest number of encounters per study site was examined (15 in the inpatient setting, 15 in the outpatient setting), precluding a meaningful analysis of site-level differences. In addition, the study was not designed to examine if documentation in the EHR varied according to the type of EHR vendor (e.g., Allscripts, Epic, Oracle Cerner). Also, the study was performed immediately prior to the launch of the trial in 2020. The results may or may not generalize to studies conducted in the postpandemic period. However, our approach to developing and evaluating a computable phenotype can inform future efforts to address these additional questions.

This study is innovative in that it evaluated the real-world use of a computable EHR phenotype across 33 heterogeneous U.S. clinical sites to test how it might be used for recruitment for a live, multicenter comparative effectiveness trial. Rather than describing a theoretical phenotype, we reviewed both structured and unstructured EHR data relevant to a clinical trial's eligibility criteria, thereby generating empirical estimates of enrichment yield and documentation completeness. The finding that patients hospitalized for COPD exacerbations were 5-fold more likely to be eligible for the RELIANCE clinical trial compared to patients with an outpatient encounter with a COPD diagnosis demonstrates the practical efficiency gains achievable through hospitalization-triggered (rather than outpatient visit-triggered) EHR prescreening. Our study also points out the need to develop and test advanced methods to address documentation gaps in EHRs. Based on these findings, we would advocate for developing and testing a candidate computable phenotype using queries of age and ICD-10 diagnosis codes in structured EHR data from inpatient encounters combined with natural language processing or AI-enabled data extraction of structured and unstructured data for other eligibility criteria.

In conclusion, results from our study indicate that multisite implementation of EHR-facilitated recruitment using computable phenotypes will require review of structured and unstructured data and novel strategies to overcome gaps in clinical documentation specific to each clinical trial's eligibility criteria.

Acknowledgements

Author contributions: SL, NB, and JK were responsible for the conception, and SL, JH, RM, SR, ES, RW, NB, and JK contributed to the design of the work. JH, RM, SR, ES, RW, and AA were responsible for the critical review of the manuscript. SL, NB, and JK did the analysis, and PP, SL, JH, RM, SR, ES, NB, AA, and JK interpreted the data. PP, SL, NB, and JK drafted the manuscript. All authors gave their final approval of the work and are accountable for all relevant parts of the work.

Declaration of Interests

PP reports funding from the National Center for Advancing Translational Sciences (NCATS), the National Heart, Lung, and Blood Institute (NHLBI), and the National Institute of Nursing Research paid to her institution and consulting fees paid to her by EpiExcellence, LLC. SL reports funding from the American Thoracic Society/American Lung Association (ALA)/the American College of Chest Physicians, and NCATS paid to her institution. JH reports funding from the National Institute of Allergy and Infectious Disease (NIAID), ALA, and PCORI paid to her institution. SR reports receipt of consulting fees, lecture honoraria and/or travel expenses from Roche, Sanofi, RS Biotherapeutics, GOLD, RespirAI, Regeneron, the Alpha 1 Foundation, Galvanize, Gondola, Wave, Lilly, CRISPR Therapeutics, Arrowhead, the National Institutes of Health, PCORI, and UCB. ES reports funding from the NHLBI, NIAID, and ALA paid to her institution. RM reports funding from PCORI, the Centers for Disease Control and Prevention, the NHLBI, the NIAID, Pfizer, AstraZeneca, and Sanofi paid to his institution. RW reports funding from Chiesi, AstraZeneca, Sanofi, and the ALA paid to his institution and consulting fees paid to him from the Centers for Medicare & Medicaid Services. AA reports funding from the Agency for Healthcare Research and Quality, NCATS, NHLBI, and PCORI paid to his institution and royalties for editor roles at UpToDate. JK reports funding from PCORI,

For personal use only. Permission required for all other uses.

the NHLBI, the COPD Foundation, the ALA, and BioVie paid to his institution, NHLBI and Cambridge University Hospitals National Health Service Foundation Trust for data-safety monitoring board/advisory board membership, and leadership roles at the Respiratory Health Association, the Global Initiative for Asthma, and the COPD Foundation. NB has no conflicts of interest to declare.

For personal use only. Permission required for all other uses.

References

1. NIH Pragmatic Trials Collaboratory. Electronic health records-based phenotyping. Rethinking Clinical Trials website. Accessed October 2025. https://rethinkingclinicaltrials.org/chapters/conduct/electronic-health-records-based-phenotyping/definitions/

2. Krishnan JA, Holbrook JT, Sugar EA, et al. Rationale and design of the roflumilast or azithromycin to prevent COPD exacerbations clinical trial. *Chronic Obstr Pulm Dis.* 2026;13(1):17-28. https://doi.org/10.15326/jcopdf.2025.0669

3. Krishnan JA, Albert RK, Rennard SI, on behalf of the RELIANCE study. Waiting for actionable evidence: roflumilast or azithromycin? *Chronic Obstr Pulm Dis.* 2022;9(1):1-3. https://doi.org/10.15326/jcopdf.2021.0272

4. LaBedz S, Albert RK, Avdalovic MV, et al. Feasibility of screening electronic health records to identify participants for a multi-center clinical effectiveness trial in COPD. *Am J Respir Crit Care Med.* 2019;199(Supplement_1):A3289. https://academic.oup.com/ajrccm/article/199/Supplement_1/A3289/8579484

5. Centers for Medicare & Medicaid Services: QualityNet. 2017 Condition-specific measures updates and specifications report: hospital-level 30-day risk-standardized readmission measures. QualityNet website. Published March 2017. Accessed March 2026. https://qualitynet.cms.gov/files/5d26a75fcbc787001f617348?filename=2017_Rdmsn_Updates&Specs_Repts.zip=

6. Prieto-Centurion V, Rolle AJ, Au DH, et al. Multicenter study comparing case definitions used to identify patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2014;190(9):989-995. https://doi.org/10.1164/rccm.201406-1166OC

7. Patel AK. The marriage of computable phenotypes with machine learning-a pathway to evidence-based care for critically ill children. *JAMA Netw Open.* 2025;8(2):e2457422. https://doi.org/10.1001/jamanetworkopen.2024.57422

8. Afshar M, Joyce C, Oakey A, et al. A computable phenotype for acute respiratory distress syndrome using natural language processing and machine learning. *AMIA Annu Symp Proc.* 2018;2018:157-165. https://pmc.ncbi.nlm.nih.gov/articles/PMC6371271/

9. National Institutes of Health, National Center for Advancing Translational Sciences (NIH, NCATS). About translational science. NCATS NIH website. Updated July 2025. Accessed October 2025. https://ncats.nih.gov/about/about-translational-science

10. Forrest CB, McTigue KM, Hernandez AF, et al. PCORnet® 2020: current state, accomplishments, and future directions. *J Clin Epidemiol.* 2021;129:60-67. https://doi.org/10.1016/j.jclinepi.2020.09.036

11. Hadley E, Yoo YJ, Patel S, et al. Insights from an N3C RECOVER EHR-based cohort study characterizing SARS-CoV-2 reinfections and Long COVID. *Commun Med (Lond).* 2024;4:129. https://doi.org/10.1038/s43856-024-00539-2

For personal use only. Permission required for all other uses.