pagepress

Italian Journal of Medicine

Italian Journal of Medicine 2026; 20:2657
doi:10.4081/itjm.2026.2657

©Copyright: the Author(s), 2026
Licensee PAGEPress, Italy

Agreement between general practitioners and central reviewers on spirometry interpretation in primary care

Giuseppe Misiano,[1] Daniele Angioni,[2] Andrea Pizzini,[1] Daniela Ferrante,[3] Mirko Parasiliti-Caprino,[4] Vittorio Picottino,[1] Marzio Uberti[1]

[1]RIMEG Research Group, General Practice, Turin; [2]RIMEG Research Group, General Practice and Clinical Ultrasonography, Novara; [3]Unit of Medical Statistics and Cancer Epidemiology, Department of Translational Medicine, University of Eastern Piedmont and CPO Piemonte, Novara; [4]Endocrinology, Diabetes and Metabolism, Department of Medical Sciences, University of Turin, Italy

Abstract

We assessed agreement between general practitioners (GPs) and central reviewers within a supported primary-care spirometry program in Piedmont, Italy. GP and central-review labels were recorded separately; reviewer blinding was not documented. Among 658 uploads dated 18 May 2022-18 January 2023, 481 from 25 GPs had paired obstruction labels. Confidence intervals (CIs) used 50,000 GP-cluster bootstrap resamples. The obstruction agreement was 92.9% (95% CI 90.3-96.0), and Cohen's κ was 0.855 (0.797-0.918); discordance comprised 28 GP-positive/reviewer-negative and 6 GP-negative/reviewer-positive records. Relative sensitivity was 96.8% and specificity 90.4%. Agreement was 95.0% for possible restriction (κ 0.818), 97.5% for recorded reversibility (κ 0.737), and 86.2% for severity (weighted κ 0.789). Four Global Initiative for Chronic Obstructive Lung Disease categories were available, but no very severe classifications were recorded. One reviewer account contributed 397/481 assessments. Incomplete pairs, the concentration of reviews, possible repeat examinations, and the limited number of GP clusters constrain interpretation. Technical-quality data and the bronchodilator-tested denominator were unavailable. High agreement describes completed pairs within this supported pathway and does not establish diagnostic accuracy or clinical benefit. Hospital-community applications require prospective evaluation.

Key words: chronic obstructive pulmonary disease, spirometry; primary care, integrated care, inter-rater agreement.

Correspondence to: Giuseppe Misiano, RIMEG Research Group, General Practice, Via Monginevro 59, int. 7, 10141 Turin, Italy.
E-mail: dottormisiano@gmail.com

Introduction

Chronic obstructive pulmonary disease (COPD) is managed across multiple settings. General practitioners (GPs) identify symptoms and risk exposure, hospital clinicians manage exacerbations and multimorbidity, and respiratory services provide physiological and diagnostic support. Spirometry is essential, but an obstructive pattern is not equivalent to a clinical diagnosis. Current Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidance requires an appropriate clinical context and a post-bronchodilator ratio of forced expiratory volume in one second (FEV1) to forced vital capacity (FVC) below 0.70 to confirm COPD.[1] Contemporary European Respiratory Society/American Thoracic Society (ERS/ATS) standards also require interpretation against suitable reference equations and lower limits of normal.[2,3]

Decentralized spirometry has two independent quality domains: technically valid acquisition and clinically appropriate interpretation. ATS/ERS standards specify equipment verification, patient preparation, maneuver acceptability, repeatability and test-session grading.[4] Primary-care standards additionally emphasize operator competency, continuing quality assurance and clinical governance.[5] Published studies have shown marked variation in maneuver quality, device accuracy, agreement with pulmonary-function laboratories and clinical interpretation.[6-12] A high level of agreement between two readers can therefore coexist with a technically suboptimal test and cannot, by itself, establish diagnostic accuracy.

Implementation evidence suggests that a training course alone is insufficient. Feedback, repeated performance review, digital transmission and specialist access appear more useful when combined within a structured pathway.[13-20] Broader integrated disease-management programs can reduce respiratory admissions and emergency-department use, while specialist integration into general practice can improve guideline-concordant COPD care.[21,22] These findings support a hospital-community model, but they do not demonstrate that spirometry interpretation alone produces those outcomes.

Medication continuity is a second major gap. In the 2024 OsMed supplementary analyses, among new users aged at least 45 years with at least two dispensings in the combined asthma/COPD drug category, 52.2% had low adherence, only 19.5% had high adherence, and 8.4% remained persistent at 12 months; high adherence was defined as coverage of at least 80% of the observation period and treatment interruption as absence of a dispensing within

OPEN ACCESS

60 days.[23] These data combine asthma and COPD and cannot be transferred directly to the present cohort, but they define the national implementation problem. A 2024 meta-analysis found that poor adherence to inhaled COPD treatment was associated with a higher exacerbation risk [odds ratio 1.40, 95% confidence interval (CI) 1.21-1.62].[24]

The primary objective was to quantify agreement between trained GPs and central reviewers for recorded spirometric obstruction. Secondary objectives were agreement on possible restriction, recorded bronchodilator reversibility and obstruction severity, and description of missing paired reviews and practitioner-level variation. Possible applications in integrated hospital-community care are discussed as service hypotheses; hospital utilization, treatment, adherence and clinical outcomes were not measured.

Materials and Methods

Study design setting and governance

This was a retrospective agreement study based on an operational extract generated during a clinical spirometry implementation program in Piedmont, Italy. The database contained 658 uploads from 18 May 2022 to 18 January 2023, 11 health-area identifiers, 26 GP identifiers, and 5 central-review account identifiers. The primary paired analysis included 25 GPs.

Ethics approval and specific research consent were not required for the present secondary analysis. The analysis used an operational extract generated in the course of a clinical spirometry implementation program. No patient was recruited, contacted, randomized, or exposed to a procedure for this analysis. The study team received no patient names, dates of birth, contact details or medical-record identifiers and no re-identification key. This statement concerns only the secondary analysis and is distinct from ordinary clinical consent for spirometry as a healthcare procedure.

Participants and data source

Adult patients were selected by participating GPs because they were considered at risk for COPD or had a previous COPD diagnosis. GPs could assess their own patients or patients referred by colleagues. The extract did not contain patient age, sex, symptoms, smoking exposure, comorbidity, medication, exacerbation, emergency-department attendance, or hospitalization data. Exact case-finding criteria and the number of unique patients could not be reconstructed because the extract was organized by upload rather than by an identifiable patient key. GPs were instructed to perform no more than one spirometry examination per patient within the program. No formal procedure was used to identify duplicate patients or to select one record from potentially repeated examinations. Adherence to the instruction could not be verified from the extract, and the unit of analysis was the uploaded record.

Training and supported implementation

The program was developed by the *Scuola Piemontese di Medicina Generale "Massimo Ferrua"* in collaboration with the University of Eastern Piedmont. GPs attended a 20-hour, 2-day theoretical and practical course on 2-3 April 2022, followed by a planned 6-month period of practice-based training. Content covered indications and contraindications, equipment calibration and maintenance, infection control, patient preparation, slow and forced maneuvers, acceptability and repeatability, bronchodilator testing, reference values and interpretation. Practical sessions included performing spirometry and interpreting prepared obstructive, restrictive, mixed and different bronchodilator-response patterns. Program documentation indicated that a continuing medical education assessment questionnaire was scheduled at the end of the residential course.

The 20-hour residential course was accredited for 17.5 Italian Continuing Medical Education credits. Completion of the course and of 20 spirometry examinations was required for a course-completion certificate jointly issued by the *Scuola Piemontese di Medicina Generale*, constituted by the *Federazione Italiana Medici di Medicina Generale* and the *Società Italiana di Medicina Generale e delle Cure Primarie*, and by the Pulmonology Unit of the University of Eastern Piedmont. The certificate documented completion of the training pathway and was not a specialist qualification. The 20-examination requirement applied to completion of the training pathway, not a minimum number of paired records required for inclusion of a GP in the agreement analysis. Participating GPs used a Minispir Light spirometer (Medical International Research, Rome, Italy). Access to ECWMed was available free of charge to all GPs in Piedmont; traces and classifications were uploaded through the platform for central review.

The training and workflow description was based on program materials and information supplied by the organizers.

Spirometry procedures and technical-quality limitations

Spirometry was performed in GP practices. The source materials indicated that bronchodilator reversibility testing with salbutamol was offered when obstruction was identified. The database did not identify who actually received bronchodilator, distinguish an unperformed test from a negative result, or provide dose and timing. It also lacked raw FEV1 and FVC values, FEV1/FVC ratios, pre- or post-bronchodilator status, reference equations, device-verification logs, maneuver counts, ATS/ERS acceptability and repeatability grades, and test-session quality grades. Technical quality and diagnostic thresholds could therefore not be independently reassessed.

Central review

Program documentation described a tutor pool comprising pulmonologists affiliated with the Department of Translational Medicine at the University of Eastern Piedmont and the Pulmonology Unit of ASL Vercelli, GPs holding a specialist qualification in respiratory medicine, and GPs with a special interest in respiratory care. These tutors are referred to as central reviewers throughout the manuscript. Review was conducted through ECWMed using the categorical classification framework described below. The five reviewer account identifiers could not be linked in the analytic extract to named individuals, their specific qualification category or years of experience; the extract also did not establish whether each account corresponded to a single reviewer. The 481 primary paired records were distributed across five reviewer account identifiers, and 397 assessments (82.5%) were recorded under one account.

GP and central-review classifications were recorded separately. The available documentation did not establish whether central reviewers were blinded to the GP label or identify a written standard operating procedure beyond the categorical classification framework. No standardized duplicate-reading or consensus-adjudication process was documented, and inter-reviewer agreement was not

OPEN ACCESS

assessed. Central-review labels therefore served as the operational comparator for agreement and not as an independently validated clinical reference standard.

Outcomes

The primary outcome was agreement on the recorded presence or absence of an obstructive spirometric pattern. Secondary outcomes were agreement on a possible restrictive pattern, recorded bronchodilator reversibility, and obstruction severity. The program framework included four recorded severity categories corresponding nominally to the GOLD spirometric grades: GOLD 1 (mild), FEV1≥80% predicted; GOLD 2 (moderate), 50%≤FEV1<80%; GOLD 3 (severe), 30%≤FEV1<50%; and GOLD 4 (very severe), FEV1<30%. GOLD defines these grades using post-bronchodilator FEV1 in patients with FEV1/FVC<0.70.[1] Neither GPs nor central reviewers recorded a very severe classification. The three categories represented in the analysis were therefore the observed subset of the four-category framework; severe and very severe grades were not combined.

Outcomes were analyzed as stored categorical labels. Because raw physiological values and pre/post-bronchodilator status were absent from the extract, correspondence of individual recorded grades with GOLD thresholds could not be independently verified. Relative sensitivity, specificity, positive predictive value and negative predictive value describe classification against the central reviewer and are not diagnostic accuracy estimates for COPD.

Statistical analysis

All available uploads were considered; no formal sample-size calculation was performed. Each outcome was analyzed using records with both GP and central-review labels, without imputation. For obstruction, 177 records lacked a complete pair: 149 had a GP label only, 2 a reviewer label only, and 26 neither label. These exclusions concerned missing classifications and did not establish that the spirometry procedures themselves were incomplete. Individual reasons for missing labels were not recorded. Binary agreement was summarized with observed agreement and Cohen κ. Severity was assessed using unweighted and linearly weighted κ across the three observed categories. Observed agreement was reported alongside κ because κ is affected by prevalence and marginal distributions.[25]

To account for clustering of records within GPs, 95% CIs for agreement, κ and classification measures were estimated using 50,000 cluster-bootstrap samples, resampling GPs with replacement. Practitioner-level record counts and agreement ranges were described. A binomial generalized estimating equation (GEE) with exchangeable working correlation evaluated change in discordance by cumulative case order within practitioner; the effect was reported per 10 additional cases. Missing paired labels were classified by which label was absent and were described by calendar month. Analyses followed the principles of the Guidelines for Reporting Reliability and Agreement Studies and Strengthening the Reporting of Observational Studies in Epidemiology.[26,27]

Results

Record flow and completeness

The database contained 658 uploads. Paired obstruction labels were available for 481 records (73.1%), while 177 records were excluded from the primary analysis: 149 had a GP label, but no reviewer label, 2 had a reviewer label, but no GP label, and 26 had neither label. Paired possible-restriction labels were available for 479 records and paired recorded-reversibility labels for 480 records. Completion of paired obstruction review was 96-100% from May through August 2022, 56%-84% from September through December, and 52% in January 2023. The primary paired records were contributed by 25 GPs, with a median of 17 records per GP (interquartile range 10-28; range 1-57). Individual reasons for missing classifications were not recorded. Program characteristics and record flow are shown in Table 1 and Figure 1.

Agreement on obstruction

GPs and reviewers agreed on 447 of 481 obstruction classifications (92.9%; cluster-bootstrap 95% CI 90.3-96.0%). Cohen κ was 0.855 (95% CI 0.797-0.918). The cross-classification comprised 263 GP-negative/reviewer-negative, 184 GP-positive/reviewer-positive, 28 GP-positive/reviewer-negative and 6 GP-negative/reviewer-positive records. Relative to the operational reviewer label, sensitivity was 96.8%, specificity 90.4%, positive predictive value 86.8% and negative predictive value 97.8%. Cross-classifications and agreement measures are presented in Tables 2 and 3.

Table 1. Operational characteristics of the Piedmont primary-care spirometry program from 18 May 2022 to 18 January 2023.

Characteristic	Value
All uploads	658
Database timestamp range	18 May 2022 to 18 January 2023
Health-area identifiers	11
GP identifiers in extraction	26
GP identifiers in primary paired analysis	25
Central-review account identifiers	5
Paired obstruction labels	481/658 (73.1%)
Paired possible-restriction labels	479/658 (72.8%)
Paired recorded-reversibility labels	480/658 (72.9%)
Paired severity labels (three observed of four recorded categories)	181
Primary paired records per GP	Median 17; IQR 10-28; range 1-57

GP, general practitioner; IQR, interquartile range. Uploads could not be linked to demonstrably unique patients because the extract contained no patient key.

Secondary agreement outcomes

For possible restriction, 455 of 479 complete pairs agreed (95.0%), and κ was 0.818 (95% CI 0.737-0.887). For recorded reversibility, 468 of 480 complete pairs agreed (97.5%), and κ was 0.737 (95% CI 0.589-0.872). The reversibility estimates cannot distinguish a negative test from a test that was not performed.

Paired obstruction-severity labels were available for 181 records and occupied three of the four recorded categories. Exact agreement was 86.2% (156/181), unweighted κ was 0.766 (95% CI 0.664-0.853), and linearly weighted κ was 0.789 (95% CI 0.699-0.866). Most disagreements were between adjacent severity categories. Of these 181 records, 179 also had positive binary obstruction labels from both evaluators; in two records, paired severity labels were

Table 2. Cross-classification of general practitioner and central-reviewer labels in the Piedmont spirometry program, 2022-2023.

Outcome	GP label	Reviewer: no	Reviewer: yes	Total
Obstruction	No	263	6	269
	Yes	28	184	212
	Total	291	190	481
Possible restrictive pattern	No	388	12	400
	Yes	12	67	79
	Total	400	79	479
Recorded reversibility	No	450	5	455
	Yes	7	18	25
	Total	457	23	480

GP, general practitioner. The reviewer label is an operational reference, not an independent diagnostic gold standard. Recorded reversibility does not identify which patients actually received bronchodilator testing.

Table 3. Agreement and classification measures relative to central review in the Piedmont spirometry program, 2022-2023.

Outcome	n	Agreement, %	κ	Relative sensitivity, %	Relative specificity, %	PPV, %	NPV, %
Obstruction	481	92.9 (90.3-96.0)	0.86 (0.80-0.92)	96.8 (94.0-99.1)	90.4 (86.5-94.8)	86.8 (80.2-93.3)	97.8 (95.9-99.3)
Possible restrictive pattern	479	95.0 (93.2-96.7)	0.82 (0.74-0.89)	84.8 (74.1-93.6)	97.0 (95.0-98.6)	84.8 (76.1-92.7)	97.0 (95.3-98.6)
Recorded reversibility	480	97.5 (95.7-99.1)	0.74 (0.59-0.87)	78.3 (68.4-92.3)	98.5 (97.2-99.6)	72.0 (50.0-90.9)	98.9 (97.9-99.8)

Values are estimates (95% cluster-bootstrap confidence intervals), based on 50,000 resamples of general practitioners. PPV, positive predictive value; NPV, negative predictive value. Measures describe classification relative to the operational reviewer label and are not estimates of COPD diagnostic accuracy.

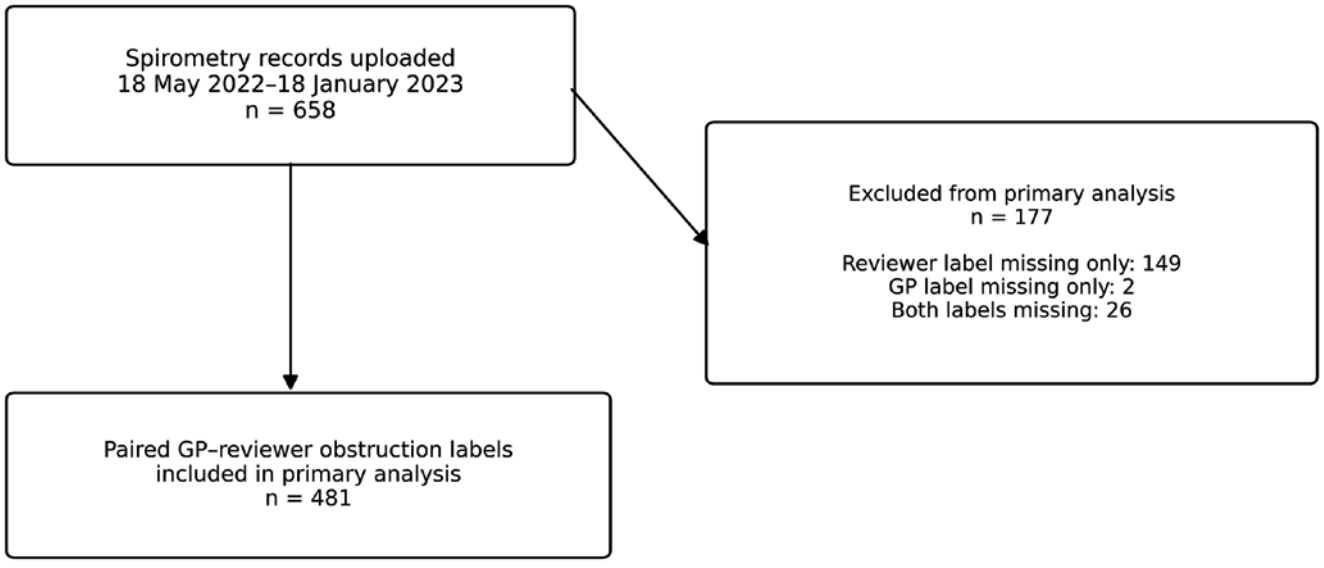


Figure 1. Flow of spirometry records through the primary agreement analysis, Piedmont, 18 May 2022-18 January 2023. Flow diagram showing 658 uploaded records, 177 exclusions because one or both obstruction labels were missing, and 481 records with paired general practitioner-reviewer obstruction labels; secondary complete-case denominators were 479 for possible restriction, 480 for recorded reversibility, and 181 for severity, of which 179 also had positive binary obstruction labels from both evaluators.

OPEN ACCESS

available although the two binary obstruction labels were not both positive. Neither GP nor central-review classifications included the very severe category, shown with zero counts in Table 4.

Practitioner and temporal variation

Obstruction agreement by GP ranged from 83.3% to 100%; among the 19 GPs contributing at least 10 paired records, the range remained 83.3%-100%. The GEE did not show a clear change in discordance with cumulative experience (odds ratio 0.76 per 10 additional cases, 95% CI 0.48-1.21; p=0.24). Because several practitioners contributed few records, these practitioner-level estimates are descriptive.

Discussion

Principal findings

Within a program combining training, equipment, digital submission and central review, agreement on recorded obstruction was 92.9%, and Cohen κ was 0.855, with CI accounting for clustering within GPs. GP-positive/reviewer-negative classifications were more frequent than the reverse. Agreement was also high for possible restriction and the three observed severity categories. These results describe categorical interpretation among completed pairs within this pathway; they do not establish technical quality, clinical diagnosis or outcome benefit.

Possible applications in integrated hospital-community care

The following service applications are hypotheses, not outcomes demonstrated by this study. Hospital internists encounter COPD during exacerbations, multimorbidity assessment and discharge planning. A shared pathway could support selective escalation of technically doubtful, discordant, severe or clinically atypical community tests. Conversely, an emergency-department visit or admission could trigger community reassessment with explicit diagnostic status, treatment plan and follow-up arrangements. Whether such a pathway reduces duplicate testing, non-selective referrals or fragmentation requires prospective evaluation.

The hypothesis is consistent with evidence for multicomponent integrated disease management rather than with spirometry alone. A 2021 Cochrane review reported fewer respiratory-related admissions (odds ratio 0.64, 95% CI 0.50-0.81), emergency-department visits (0.69, 0.50-0.93) and all-cause admissions (0.75, 0.57-0.98) with integrated programs.[21] Specialist-led integration into general practices has also improved delivery of guideline-concordant COPD care.[22] Regional integrated pathways provide another model for outcome evaluation.[28] Interpretation support may be one pathway component; it is not itself evidence of reduced hospital use.

Spirometry as a pharmacological-management checkpoint

In a future pathway, a confirmed and clinically contextualized spirometric result could prompt medication reconciliation, assessment of inhaler technique, symptom and exacerbation review, evaluation of inhaled corticosteroid indications, smoking-cessation and vaccination support, and follow-up planning. Linked dispensing data could flag discontinuation or low treatment coverage. These actions and their effects were not evaluated in this study.

The OsMed findings identify a national adherence gap, although asthma and COPD are combined.[23] Our dataset contains no prescribing or dispensing information and cannot establish improved adherence. Poor inhaler adherence is associated with exacerbation risk, supporting prospective evaluation of this proposed mechanism.[24]

Interpretation in relation to previous evidence

The observed κ reflects a narrow classification task under supported conditions. It should not be compared directly with studies of full respiratory diagnosis, which require symptoms, exposure, post-bronchodilator physiology and exclusion of alternatives. Longitudinal primary-care data indicate that access to diagnostic spirometry and correct interpretation can improve over time but remain incomplete.[29] The findings should be interpreted in the context of continued review and feedback rather than attributed to training alone. The high negative predictive agreement describes correspondence with the reviewer label, not exclusion of clinical COPD. The greater number of GP-positive/reviewer-negative records identifies a potential focus for future review, but the extract cannot establish whether disagreements arose from borderline physiology, technical quality or interpretation.

Limitations

The main limitation is constructing validity: the study measured agreement on stored labels, not diagnostic accuracy against an independent clinical reference. Raw spirometry, reference equations, pre/post-bronchodilator status and technical-quality grades were unavailable; physiological thresholds could not be checked against individual recordings. The true denominator for bronchodilator testing was unknown.

Central review was concentrated under one account, which

Table 4. Agreement on recorded obstruction severity in 181 paired records, Piedmont spirometry program, 2022-2023.

GP severity	Reviewer mild	Reviewer moderate	Reviewer severe	Reviewer very severe	Total
Mild	61	13	1	0	75
Moderate	4	78	6	0	88
Severe	0	1	17	0	18
Very severe	0	0	0	0	0
Total	65	92	24	0	181

GP, general practitioner. Exact agreement was 86.2% (156/181). Unweighted κ was 0.766 (95% confidence interval 0.664-0.853) and linearly weighted κ was 0.789 (0.699-0.866). Confidence intervals are based on GP-cluster bootstrap resampling. Mild, moderate, severe and very severe were recorded categories corresponding nominally to Global Initiative for Chronic Obstructive Lung Disease grades 1-4. The program framework included all four categories, but neither evaluator recorded a very severe classification. Of the 181 records with paired severity labels, 179 also had positive binary obstruction labels from both evaluators; in two records, the two binary obstruction labels were not both positive. The zero row and column show the absence of very severe labels; agreement estimates use the three observed categories without combining grades.

contributed 397 of 481 paired assessments (82.5%). The pooled agreement estimates are consequently dominated by classifications recorded under that account and may not represent agreement with a broader reviewer group. Because the extract did not establish whether each account represented a single individual or permit linkage to reviewer-specific qualifications and experience, this concentration can be interpreted only at account level. The direction of any effect on agreement cannot be determined. Separately recorded classifications do not establish reviewer blinding, and inter-reviewer reproducibility was not evaluated. GP-cluster resampling accounted for dependence within GPs only; it neither estimated reviewer-level variability nor established reproducibility across reviewers.

The complete-pair analysis excluded 177 of 658 uploads (26.9%). Completion declined over time, raising concern that analyzed records may differ from excluded uploads. The pattern does not establish the missingness mechanism. Individual reasons for missing labels were not recorded. If completion was related to case complexity, technical quality or interpretation difficulty, agreement could be biased in either direction. Missing patient and technical-quality data prevented adequate assessment or adjustment. Results therefore apply to uploads with paired classifications and may not represent all program examinations.

GPs were instructed to avoid repeat examinations on the same patient, but no formal procedure identified duplicate patients or selected among potentially repeated records, and the extract contained no patient key. Repeat examinations therefore cannot be excluded and could give some patients disproportionate weight. Resampling whole GP clusters retains within-GP dependence, including any repeats within a GP, but does not identify or remove duplicate patients or address repeats across GPs. Agreement was therefore estimated per uploaded record; the number of unique patients was unknown.

Only 25 GP clusters contributed to the primary analysis, with unequal sizes of 1-57 records. GPs with more records contributed more to the pooled estimates. Cluster-bootstrap CI account for within-GP dependence, but their precision remains constrained by the limited number and heterogeneity of clusters; 50,000 resamples improve numerical stability without increasing the amount of independent information. Practitioner-level estimates and the exploratory GEE analysis require caution. The supported regional setting also limits generalizability to practices without comparable training, equipment, and central-review access.

No very severe classifications were recorded, so agreement cannot be extrapolated to GOLD 4 obstruction. The absence of demographics, symptoms, exposures, treatment and hospital data also precludes conclusions about patient case mix or clinical benefit.

Priorities for prospective evaluation

Future evaluation should use a prespecified post-bronchodilator protocol, ATS/ERS quality grading, identified reviewers and an independent adjudication process. A pragmatic cluster or stepped-wedge design could compare the supported pathway with usual care. Patient-level linkage to pharmacy, emergency-department and hospital data would allow assessment of diagnostic clarification, repeat testing, referral appropriateness, treatment changes, inhaler adherence, exacerbations and hospital use. These outcomes are needed to test the proposed hospital-community applications.

Conclusions

Trained GPs and central reviewers showed high agreement on recorded spirometric obstruction among uploads with complete paired classifications in a supported primary-care program. These findings do not establish technical quality, independent diagnostic accuracy or benefit to patients. Selective referral, post-discharge reassessment and inhaler-management support require prospective evaluation. Hospitalizations, treatment changes, adherence and other clinical outcomes were not measured.

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for prevention, diagnosis and management of COPD: 2026 report. Available from: https://goldcopd.org/2026-gold-report-and-pocket-guide/
2. Stanojevic S, Kaminsky DA, Miller MR, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. Eur Respir J 2022;60:2101499.
3. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-year age range: the Global Lung Function 2012 equations. Eur Respir J 2012;40:1324-43.
4. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of spirometry 2019 update: an official American Thoracic Society and European Respiratory Society technical statement. Am J Respir Crit Care Med 2019;200:e70-88.
5. Levy ML, Quanjer PH, Booker R, et al. Diagnostic spirometry in primary care: proposed standards for general practice compliant with American Thoracic Society and European Respiratory Society recommendations. Prim Care Respir J 2009;18:130-47.
6. White P, Wong W, Fleming T, Gray B. Primary care spirometry: test quality and the feasibility and usefulness of specialist reporting. Br J Gen Pract 2007;57:701-5.
7. Schermer TRJ, Crockett AJ, Poels PJP, et al. Quality of routine spirometry tests in Dutch general practices. Br J Gen Pract 2009;59:e376-82.
8. Hegewald MJ, Gallo HM, Wilson EL. Accuracy and quality of spirometry in primary care offices. Ann Am Thorac Soc 2016; 13:2119-24.
9. van de Hei SJ, Flokstra-de Blok BMJ, Baretta HJ, et al. Quality of spirometry and related diagnosis in primary care with a focus on clinical use. NPJ Prim Care Respir Med 2020;30:22.
10. Rothnie KJ, Chandan JS, Goss HG, et al. Validity and interpretation of spirometric recordings to diagnose COPD in UK primary care. Int J Chron Obstruct Pulmon Dis 2017;12:1663-8.
11. Schermer TRJ, Jacobs JE, Chavannes NH, et al. Validity of spirometric testing in a general practice population of patients with chronic obstructive pulmonary disease. Thorax 2003;58: 861-6.
12. Schneider A, Gindner L, Tilemann L, et al. Diagnostic accuracy of spirometry in primary care. BMC Pulm Med 2009;9:31.
13. Holton C, Crockett A, Nelson M, et al. Does spirometry training in general practice improve quality and outcomes of asthma care? Int J Qual Health Care 2011;23:545-53.
14. Schermer TRJ, Akkermans RP, Crockett AJ, et al. Effect of e-learning and repeated performance feedback on spirometry test quality in family practice: a cluster trial. Ann Fam Med 2011;9: 330-6.

OPEN ACCESS

15. Parsons R, Schembri D, Hancock K, et al. Effects of the spirometry learning module on the knowledge, confidence, and experience of spirometry operators. NPJ Prim Care Respir Med 2019; 29:30.
16. Averame G, Bonavia M, Ferri P, et al. Office spirometry can improve the diagnosis of obstructive airway disease in primary care setting. Respir Med 2009;103:866-72.
17. Burgos F, Disdier C, de Santamaria EL, et al. Telemedicine enhances quality of forced spirometry in primary care. Eur Respir J 2012;39:1313-8.
18. Marina Malanda N, Lopez de Santa Maria E, Gutierrez A, et al. Telemedicine spirometry training and quality assurance program in primary care centers of a public health system. Telemed J E Health 2014;20:388-92.
19. Thijssing L, van der Heijden JP, Chavannes NH, et al. Telepulmonology: effect on quality and efficiency of care. Respir Med 2014;108:314-8.
20. Jankowski P, Gorska K, Mycroft K, et al. The use of mobile spirometry with feedback quality assessment in primary care: a nationwide cross-sectional feasibility study. Respir Med 2021; 184:106472.
21. Poot CC, Meijer E, Kruis AL, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2021;9: CD009437.
22. Patel K, Pye A, Edgar RG, et al. Cluster randomised controlled trial of specialist-led integrated COPD care (INTEGR COPD). Thorax 2024;79:209-18.
23. Osservatorio Nazionale sull'impiego dei Medicinali. L'uso dei farmaci in Italia. Rapporto Nazionale Anno 2024. Materiale supplementare, Sezione 6, Tabelle 3.7.1.s2-s3. Rome, Italy: Agenzia Italiana del Farmaco; 2025.
24. Vauterin D, Van Vaerenbergh F, Grymonprez M, et al. Medication adherence to inhalation therapy and the risk of COPD exacerbations: a systematic review with meta-analysis. BMJ Open Respir Res 2024;11:e001964.
25. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37:360-3.
26. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96-106.
27. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573-7.
28. Wu CX, Hwang CH, Tan WS, et al. Effectiveness of a chronic obstructive pulmonary disease integrated care pathway in a regional health system: a propensity score matched cohort study. BMJ Open 2018;8:e019425.
29. Athlin A, Lisspers K, Hasselgren M, et al. Diagnostic spirometry in COPD is increasing: a comparison of two Swedish cohorts. NPJ Prim Care Respir Med 2023;33:23.

Received: 16 July 2026; Accepted: 2 October 2026; Early view: 23 October 2026.

program data and manuscript development. Daniela Ferrante: supervision of the statistical methodology. All authors: interpretation, critically revision of the manuscript, approval of the final version.

Conflict of interest: the authors declare no conflicts of interest.

Ethics approval and consent to participate: not required for this retrospective secondary analysis of an operational extract generated during clinical care. No research-specific intervention or patient contact occurred, and the study team received no patient identifiers or re-identification key.

Informed consent: not applicable.

Availability of data and materials: the de-identified operational dataset is not deposited in a public repository. Requests may be addressed to the corresponding author and are subject to data-controller and governance requirements.

Conference presentation: program results were presented at the Scuola Piemontese di Medicina Generale meeting in Turin on 30 September 2023.

Funding: the original clinical implementation program, the secondary statistical analysis and the preparation of this manuscript received no specific external funding. Access to ECWMed was available free of charge to all GPs in Piedmont.

Acknowledgments: the authors thank the participating general practitioners and central reviewers for their contribution to the clinical program.

Declaration of generative artificial intelligence and artificial intelligence-assisted technologies in the writing process: an OpenAI language model was used to support English-language editing, manuscript restructuring, literature-search organisation, consistency checks and preparation of the revised manuscript and responses to reviewers. No generative artificial intelligence was used to generate or alter study data. The authors reviewed the output and retain responsibility for the final content.

Publisher's note: all claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).