Diagnostic performance of a machine-learning algorithm (Asthma/COPD Differentiation Classification; AC/DC) tool versus primary care physicians and pulmonologists in asthma, COPD and ACO

Janwillem W.H. Kocks, MD, PhD[a,b,c] Hui Cao, MD, PhD[d] Björn Holzhauer, PhD[e] Alan Kaplan, MD[f] J. Mark FitzGerald, MD[g†] Konstantinos Kostikas, MD, PhD[h] David Price, FRCGP[c,i] Helen K. Reddel, MBBS, PhD[j] Ioanna Tsiligianni, MPH, PhD[k] Claus F. Vogelmeier, MD[l] Sebastien Bostel, PhD[e] Paul Mastoridis, PharmD[d*]

[a]General Practitioners Research Institute, Groningen, The Netherlands; [b]University of Groningen, University Medical Centre Groningen, GRIAC Research Institute, Groningen, The Netherlands; [c]Observational and Pragmatic Research Institute, Singapore; [d]Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States; [e]Novartis Pharma AG, Basel, Switzerland; [f]Family Physician Airways Group of Canada, University of Toronto, Toronto, ON, Canada; [g]Department of Medicine, Division of Respiratory Medicine, The University of British Columbia, Vancouver, Canada; [h]Respiratory Medicine Department, University of Ioannina School of Medicine, Ioannina, Greece; [i]Centre of Academic Primary Care, Division of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom; [j]The Woolcock Institute of Medical Research and The University of Sydney, Sydney, NSW, Australia; [k]Health Planning Unit, Department of Social Medicine, Faculty of Medicine, University of Crete, Heraklion, Crete, Greece; [l]Department of Medicine, Pulmonary and Critical Care Medicine, University Medical Centre Giessen and Marburg, Philipps-Universität Marburg, Member of the German Centre for Lung Research (DZL), Marburg, Germany

[†]Author is deceased at the final stage of manuscript development.

***Corresponding author**

Paul Mastoridis

Novartis Pharmaceuticals Corporation,
East Hanover, NJ, United States
Telephone: 1-847-899-9334
Email: paul.mastoridis@novartis.com

Funding

The study was funded by Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States.

Declaration of interest

Authors received no compensation related to development of the manuscript. Janwillem W.H. Kocks reports grants, personal fees and non-financial support from AstraZeneca, grants, personal fees and non-financial support from Boehringer Ingelheim, grants and personal fees from Chiesi Pharmaceuticals, grants, personal fees and non-financial support from GSK, grants and personal fees from Novartis, grants from Mundi Pharma, grants from TEVA, outside the submitted work, all paid to his institution; and Janwillem W. H. Kocks holds 72.5% of shares in the General Practitioners Research Institute. Alan Kaplan is a medical advisor or is on the speaker's bureau for AstraZeneca, Behring, Boehringer Ingelheim, Cipla, Covis, GSK, Merck Frosst, Pfizer, Novartis, NovoNordisk, Sanofi, Eisai, Teva, Trudel and Valeo. J. Mark FitzGerald has received honoraria as a member of speaker bureaus and advisory boards for AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, and Sanofi Regeneron, outside the area of work.

Konstantinos Kostikas has received honoraria for presentations and consultancy fees from AstraZeneca, Boehringer Ingelheim, CSL Behring, Chiesi, ELPEN, GSK, Menarini, Novartis and Sanofi (paid to the University of Ioannina). His department has received funding and grants from AstraZeneca, Boehringer Ingelheim, Chiesi, Innovis, ELPEN, GSK, Menarini, Novartis and NuvoAir (paid to the University of Ioannina). David Price has board

membership with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan, Mundipharma, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceuticals, Thermofisher; consultancy agreements with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Mylan, Mundipharma, Novartis, Pfizer, Teva Pharmaceuticals, Theravance; grants and unrestricted funding for investigator-initiated studies (conducted through Observational and Pragmatic Research Institute Pte Ltd) from AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan, Merck, Zentiva (Sanofi Generics), Mundipharma, Novartis, Pfizer, Regeneron Pharmaceuticals, Respiratory Effectiveness Group, Sanofi Genzyme, Teva Pharmaceuticals, Theravance, UK National Health Service; payment for lectures/speaking engagements from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, Kyorin, Mylan, Mundipharma, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceuticals; payment for the development of educational materials from Mundipharma, Novartis; payment for travel/ accommodation/meeting expenses from AstraZeneca, Boehringer Ingelheim, Mundipharma, Mylan, Novartis, Thermofisher; funding for patient enrolment or completion of research from Novartis; stock/stock options from AKL Research and Development Ltd which produces phytopharmaceuticals; owns 74% of the social enterprise Optimum Patient Care Ltd (Australia and UK) and 74% of Observational and Pragmatic Research Institute Pte Ltd (Singapore); 92.61% of Observational and Pragmatic Research Institute Pte Ltd (Singapore); 5% shareholding in Timestamp which develops adherence monitoring technology; is peer reviewer for grant committees of the Efficacy and Mechanism Evaluation programme, and Health Technology Assessment; personal fees from Airway Vista Secretariat, EPG Communication Holdings Ltd, FIECON Ltd, Fieldwork International, OM Pharma SA, PeerVoice, Phadia AB, Spirosure Inc, Strategic North Limited, Synapse Research Management Partners S.L, Talos Health Solutions, WebMD Global LLC, outside the

submitted work; and was an expert witness for GlaxoSmithKline. Helen K. Reddel has provided independent advice on advisory boards/steering committees for AstraZeneca, Chiesi, GSK, Novartis, and Sanofi/Genzyme; has provided independent medical education at symposia funded by AstraZeneca, Boehringer Ingelheim, Chiesi, Getz, GSK, Novartis, Sanofi/Genzyme and Teva; has provided independent consulting for AstraZeneca and Novartis; and her institute has received unrestricted research grants from AstraZeneca, GSK, and Novartis. Ioanna Tsiligianni has received fees for consultancies from Novartis, Boehringer Ingelheim, AstraZeneca and GSK and grants for research activities from the Greek, GSK, AstraZeneca, Boehringer Ingelheim and ELPEN. Claus F. Vogelmeier gave presentations at symposia and/or served on scientific advisory boards sponsored by AstraZeneca, Boehringer Ingelheim, CSL Behring, Chiesi, GSK, Grifols, Menarini, Novartis, Nuvaira, and Aerogen. Hui Cao, Björn Holzhauer, Sebastien Bostel and Paul Mastoridis are employees of Novartis and hold stock in Novartis Pharma AG.

Abstract

Background: Differential diagnosis of asthma and chronic obstructive pulmonary disease (COPD) poses a challenge in clinical practice, and their misdiagnosis results in inappropriate treatment, increased exacerbations and potentially even death.

Objective: To investigate the diagnostic accuracy of the Asthma/COPD Differentiation Classification (AC/DC) tool compared with primary care physicians and pulmonologists in asthma, COPD, and asthma-COPD overlap (ACO).

Methods: The AC/DC machine learning-based diagnostic tool was developed using 12 parameters from electronic health records of >400,000 patients aged ≥35 years. An expert panel of 3 pulmonologists and 4 general practitioners from 5 countries evaluated 119 patient cases from a prospective observational study and provided a confirmed diagnosis (n=116) of asthma (n=53), COPD (n=43), ACO (n=7) or other (n=13). The cases were then reviewed by 180 primary care physicians and 180 pulmonologists from 9 countries and by AC/DC tool, and diagnostic accuracies were compared with reference to the expert panel diagnoses.

Results: Average diagnostic accuracy of the AC/DC tool was superior to primary care physicians (median difference, 24%; 95% posterior credible interval [CrI]: 17–29%; P<0.0001) and was non-inferior and superior (median difference,12%; 95% CrI: 6–17%; P<0.0001 for non-inferiority and P=0.0006 for superiority) to pulmonologists. The average diagnostic accuracies were 73%, 50% and 61% by AC/DC tool, primary care physicians, and pulmonologists versus expert panel diagnosis, respectively.

Conclusion: The AC/DC tool demonstrated superior diagnostic accuracy compared with primary care physicians and pulmonologists in diagnosis of asthma and COPD in patients aged ≥35 years and has the potential to support physicians in the diagnosis of these conditions in clinical practice.

Highlights Box

What is already known about this topic? Misdiagnosis of asthma and COPD can have many negative health consequences. Machine learning is playing an increasing role in diagnostic medicine and has potential use for healthcare professionals in accurate diagnosis of chronic respiratory diseases
What does this article add to our knowledge? The Asthma/COPD Differentiation Classification (AC/DC) machine learning-based diagnostic tool demonstrated superior diagnostic accuracy compared with primary care physicians and pulmonologists in the diagnosis of asthma and COPD in patients aged ≥35 years
How does this study impact current management guidelines? The AC/DC tool has the potential to be an aid in the differential diagnosis of patients with asthma or COPD and provide a valuable additional resource to supplement the decision-making of practicing physicians

Keywords: Asthma, COPD, Differential diagnosis, machine learning, AC/DC tool, asthma-COPD overlap, Primary care physician, Pulmonologist, Accuracy

Abbreviations

AC/DC: Asthma/COPD Differentiation Classification

ACO: Asthma-COPD Overlap

ACQ: Asthma Control Questionnaire

AI: Artificial Intelligence

CCQ: Clinical COPD Questionnaire

COPD: Chronic Obstructive Pulmonary Disease

CrI: Posterior Credible Interval

MI: Multiple Imputations

PCP: Primary Care Physician

Introduction

Asthma and chronic obstructive pulmonary disease (COPD) are heterogeneous chronic respiratory diseases that have overlapping diagnostic criteria and sometimes similar clinical presentations, which pose a challenge in their differential diagnoses, especially in smokers, ex-smokers and older adults.[1-4] Asthma-COPD overlap (ACO) comprises patients with characteristics of both asthma (e.g. variability of airway limitation, allergies) and COPD (e.g. age of onset ≥40 years, chest X-ray with severe hyperinflation).[1, 4, 5] Chronic respiratory diseases are major causes of morbidity and mortality, and incorrect diagnosis may lead to negative consequences in disease management;[6] for example, underdiagnosis of asthma leads to increased hospitalisations, emergency room visits, risk of death and healthcare resource costs.[7-9] Misdiagnosis can result in adverse events due to incorrect treatment (particularly if asthma is treated with long-acting bronchodilators alone) and increased treatment costs.[7-12] Overlapping diagnosis of asthma and COPD was reported to be 15% to 32%.[4] Hence, accurate diagnosis is key for therapeutic decision-making.

Artificial intelligence, especially machine learning, is playing an increasing role in diagnostic medicine and might be useful for primary care physicians (PCPs) and other healthcare professionals in accurate and differential diagnosis of chronic respiratory diseases.[13-16] The Asthma/COPD Differentiation Classification (AC/DC) tool employs a machine learning-based algorithm and was developed to aid PCPs and other physicians in fast and accurate diagnosis of asthma, COPD or ACO, in conjunction with spirometry, and to reduce any delay in symptomatic patients receiving appropriate therapy.[17] This study investigated the diagnostic accuracy of the AC/DC tool compared with PCPs and pulmonologists in the differential diagnosis of asthma, COPD, ACO and other respiratory diseases using patient cases from a prospective observational study in general practice.[17] Pulmonologists and PCPs were selected as the medical professions to be evaluated as they were the professionals most

likely to initially interact with patients with a respiratory disease, they manage patients with respiratory diseases, which is not be the case with other groups such as allergists for COPD, and are potentially the primary users of the AC/DC tool.

Methods

Study design

This was a non-interventional, multinational, multiple-rater, multiple-case study that utilised de-identified patient cases from a prospective observational study (FOCUS), which recorded data for patients presenting to general practices in the Netherlands with respiratory symptoms.[18] Cases were included if patients were aged ≥35 years at the time of data collection and if the key data required for the AC/DC tool had been recorded. Further details about methods are provided in the Online Repository Text.

The AC/DC tool was initially developed using the clinical characteristics of >400,000 patients aged ≥35 years with diagnoses of asthma, COPD or ACO by specialists (pulmonologists/allergists), as identified from Optum® de-identified electronic health records dataset between 2010 and 2017 (for index date definitions please see Online Repository Text). In an internal validation, the model achieved a sensitivity of 0.98, 0.98 and 0.78, a precision of 0.97, 0.97 and 0.92, and a F1-score of 0.98, 0.98 and 0.84, in diagnosing asthma, COPD and ACO, respectively (Online Repository Text).[17] From the >400,000 patients' data, 12 variables were identified as the most impactful and hence utilised by the AC/DC tool (**Table 1)**.

The performance (external validation) of the AC/DC tool (utilising the 12 most impactful variables) was then compared with pulmonologists and PCPs in the diagnosis of patients from the FOCUS study, and the findings are reported here.

Written informed consent obtained from each patient during the observational study [18] permitted secondary use of their data for this study. The study protocol was reviewed by an independent ethics committee or institutional review board and was conducted in accordance with the International Conference on Harmonization Guidelines for Good Clinical Practice and the Declaration of Helsinki. The study essentially comprised four steps, as shown in **Figure 1.**

Step 1. Expert panel diagnosis of each case (gold standard): A panel of seven experts (comprising 3 PCPs and 4 pulmonologists from 5 countries, who were also involved in the development of the AC/DC tool) reviewed the clinical data of 119 de-identified (eligible, n=116) patient cases from the observational study [18], and each expert determined a diagnosis of either asthma, COPD, ACO or "other disease than asthma, COPD and ACO" for each patient, and recorded the difficulty of diagnosis on a 6-point Likert scale from 0–5, with 0–1 described as "easy to diagnose" and 4–5 as "hard to diagnose" for each. The observational study database included variables such as patients' demographics and baseline clinical characteristics; current inhaled medication (yes/no); medical history questionnaire including Medical Research Council (MRC) dyspnoea scale, the Asthma Control Questionnaire (ACQ-7, 0–6) and the Clinical COPD Questionnaire (CCQ, 0–6); and spirometry results. Two symptom definitions were used; symptom definition #1: symptoms present in previous 7 days if ACQ Q4 score (shortness of breath) >0, ACQ Q5 (wheeze) >0 or CCQ Q5 (cough) >0; symptom definition #2: ACQ Q4 score >1 or ACQ Q5 >1 or CCQ Q5 >1. Symptoms (yes/no) were fed into the algorithm and shown to the physicians.

To be considered an expert panel diagnosis, five out of seven experts had to provide the same diagnosis. The primary case set included patients with a diagnosis of asthma, COPD or ACO; the exploratory case set included patients with a diagnosis of asthma, COPD, ACO and "other disease than asthma, COPD and ACO".

Step 2. Diagnosis of the clinical cases by PCPs and pulmonologists: PCPs and pulmonologists were recruited from 9 countries (United States, Canada, United Kingdom, France, Germany, Spain, Australia, China, and India). Participating PCPs and pulmonologists were included if they were licensed and practicing at the time of study with ≥3 years in practice and ever diagnosed or treated ≥1 patient with a respiratory disease.

Each physician reviewed 30 expert panel diagnoses of combined primary and exploratory clinical cases (24 cases and 6 re-reviews to assess intra-rater variability) and assigned a diagnosis of asthma, COPD, ACO or "other", together with their level of confidence in the diagnosis from 1 ("not at all confident") to 7 ("very confident"), using a cross-sectional, 60-minute web-based electronic case review system.

Step 3. Diagnosis of the clinical cases by the AC/DC tool: For the AC/DC tool, a total of 100 algorithms were trained, each with recall (true positive diagnosis) and precision (% of true positive diagnosis) ≥80% for asthma, COPD and ACO, and overall accuracy of ≥95% to fully characterise the stability of the model training process and the model performance.

The AC/DC tool assessed each of the expert panel diagnosis cases and either rejected the expert-assigned diagnosis or assigned a probability to diagnoses of asthma, COPD or ACO. The algorithm rejected cases if clinical characteristics were beyond the range on which the algorithm was trained. The diagnosis assigned by the AC/DC tool was the diagnosis with the highest predicted probability. Refer to **Figure E1** in the **Online Repository** for the confusion matrix of panel diagnosis vs diagnosis by algorithms and physicians.

Step 4. Outcome

The primary objective was to compare the average diagnostic accuracy of the AC/DC tool with those of PCPs and pulmonologists when evaluating clinical cases of asthma, COPD or ACO in the primary case set. The diagnostic accuracy of the AC/DC tool, PCPs and pulmonologists was defined as the correct diagnoses of the clinical cases, expressed as a

227 percentage, when compared with the confirmed diagnoses assigned by the expert panel (gold
228 standard). The differences in the average overall diagnostic accuracy between AC/DC tool
229 and PCPs and between AC/DC tool and pulmonologists in reference to the expert panel
230 diagnoses were analysed.
231 Secondary objectives included: (i) comparison of diagnostic accuracy (sensitivity [recall],
232 precision [positive predictive value], F1 score [harmonic mean of sensitivity and precision],
233 negative predictive value and specificity of the AC/DC tool in diagnosing asthma, COPD and
234 ACO cases compared with PCPs and pulmonologists in the primary case set; (ii)
235 determination of inter-rater and intra-rater agreement among PCPs and pulmonologists in the
236 primary case set by using Fleiss' kappa. The F1-score is the harmonic mean of precision and
237 recall; therefore, this score takes both false positives and false negatives into account.
238 Key exploratory objectives included: (i) examination of diagnostic accuracy of AC/DC tool
239 compared with PCPs and pulmonologists in diagnosing asthma, COPD and ACO in the
240 primary case set subgroups based on the expert panel scores for difficulty of diagnosis (lower
241 tertile [easy]; middle tertile [moderately hard]; and upper tertile [hard]); (ii) examination of
242 the diagnostic accuracy of PCPs and pulmonologists in diagnosing cases in the exploratory
243 case set.
244 ***Statistical analysis***
245 The expert panel diagnosis cases were divided into the primary case set (asthma, COPD, and
246 ACO cases), and the exploratory case set (including cases of other diseases than asthma,
247 COPD, and ACO). The primary outcome was analysed using a Bayesian model that jointly
248 modelled each patient's true disease status (i.e. expert panel diagnosis) as a categorical
249 random variable, and the diagnoses given for each patient by each physician or algorithm
250 (determined using multinomial logistic regression). The multinomial logistic regression
251 model included a separate intercept term for each combination of disease (asthma, COPD,

and ACO) and group (PCPs, pulmonologists, and AC/DC tool), as well as a random case and random rater effect. The primary analysis included the first diagnosis for each case by the physicians. Any repeated diagnoses by the same physician of the same patient were excluded but were considered for estimation of intra-rater reliability.

The key objective of a superiority trial is to demonstrate that a new treatment or device is better than an active control or placebo or a conventional method, while a non-inferiority trial is designed to show that treatments are not unacceptably worse than, or 'non-inferior' to, the comparator.[19] A machine learning tool could be of value to PCPs if it is superior to PCPs without necessarily needing to be superior to pulmonologists. Hence, superiority of AC/DC tool was tested against PCPs. However, after testing superiority to PCPs, non-inferiority to pulmonologists was tested followed by superiority to pulmonologists. The null hypothesis was tested against the alternative hypothesis for superiority of AC/DC tool versus PCPs and pulmonologists for the primary outcome, and a 10% non-inferiority margin versus pulmonologists, was used. A similar margin (10%) has been used previously in the literature.[20] The main analysis used the primary symptom definition #1. A sensitivity analysis was also performed using a symptom definition #2 for the primary outcome. Point estimates and their 95% credible intervals for the primary analysis were obtained as the medians, and 2.5^{th} and 97.5^{th} percentiles of the posterior distribution for the average diagnostic accuracy of, as well as for the difference in average accuracy between, AC/DC tool, PCPs, and pulmonologists. Calculation of differences and their 95% credible intervals allows quantification of the uncertainty around the diagnostic performance of each group (AC/DC, PCPs and pulmonologists) and judgement as to whether between-group differences are likely due to chance or not.

Power calculations were based on simulations in which it was assumed that pulmonologists provide 60% correct diagnoses and the AC/DC tool ≥82% – with the assumption that

pulmonologists would perform better than PCPs. With at least 30, 30, and 20 patient cases with a panel diagnosis of asthma, COPD, and ACO, respectively, approximately 90% power was achieved for a comparison of the algorithms compared with 50 pulmonologists. As all pulmonologists could not review all clinical cases, the number of pulmonologists was increased to achieve a similar total number of reviewed cases. The number of PCPs chosen was the same as the number of pulmonologists. Statistical analyses were conducted in R version 3.6.1 using the RStan R package for the primary analysis.[21-23] The details of missing data imputation are presented in the Online Repository Text..

Results

Baseline demographics and clinical characteristics

This analysis included 116 cases (asthma, n=53; COPD, n=43; ACO, n=7; other, n=13) who were assigned an expert panel diagnosis; i.e. n=103 (53+43+7) in the primary case set (diagnosis of other not included) and 116 in the exploratory case set; consensus was not achieved for 3 cases (**Figure 1**). Baseline demographics and clinical characteristics of the cases are presented in **Table 2**.

Out of the 116 patients with an expert panel diagnosis used to evaluate the AC/DC tool, 95 (82%) had no missing data for the 12 variables, while out of the remaining 21 (18%) patients 12 (10%) had missing information on dyspnea, 13 (11%) missing information on wheeze, and 8 (7%) missing cough symptom information (some of the 21 had missing information on more than one symptom). Other variables such as demographic information, spirometry results, smoking information, and comorbidity information were completely available for all of the 116 patients.

In total, 360 physicians (180 PCPs and 180 pulmonologists) from 9 countries (20 PCPs and 20 pulmonologists from each country) were included with mean post-residency practice times comparable between PCPs (8.4–27.3 years) and pulmonologists (9.7–22.3 years).

Average diagnostic accuracy

Average diagnostic accuracy of AC/DC tool was superior to PCPs (median difference 24%; 95% posterior credible interval [CrI]: 17–29%; P<0.0001) and was non-inferior and superior to pulmonologists (median difference 12%; 95% CrI: 6–17%; P<0.0001 for non-inferiority and P=0.0006 for superiority) in correct diagnosis of asthma, COPD and ACO (based on expert panel diagnosis). The average diagnostic accuracy of pulmonologists was superior to that of PCPs (median difference 12%, 95% CrI: 8–15%) (**Figure 2**). Sensitivity analyses showed similar results to the main analysis (AC/DC tool vs PCPs, median difference 24%, 95% CrI: 17–30%; AC/DC tool vs pulmonologists, median difference 11%, 95% CrI: 5–17%; pulmonologists vs PCPs, median difference 12%, 95% CrI: 9–16%).

Secondary measures of diagnostic performance

For sensitivity (percentage of "true positive" diagnoses (based on expert panel) made from all diagnoses with each disease), the AC/DC tool correctly identified higher proportions of asthma and COPD patients, while PCPs and pulmonologists correctly diagnosed more ACO patients. The precision (percentage of true positive diagnoses made from the total positive diagnoses made) results for diagnosis of asthma was similar between AC/DC tool and pulmonologists and only slightly lower for PCPs, while precision results for COPD and ACO were higher for PCPs and pulmonologists than the AC/DC tool. The F1-score (a measure of accuracy that combines sensitivity and precision) for the AC/DC tool was higher than that for PCPs and pulmonologists for the diagnosis of asthma, better than PCPs and similar to pulmonologists for diagnosing COPD, and less than PCPs and pulmonologists for ACO. The negative predictive value (percentage of true negative diagnoses given from the negative diagnoses made) was higher for the AC/DC tool for asthma and COPD than those for PCPs and pulmonologists, while the values for PCPs and pulmonologists were slightly higher than

the AC/DC tool for ACO. Specificity (percentage of true negative diagnoses made from all diagnoses who did not have each diagnosis) values were similar for the AC/DC tool, PCPs, and pulmonologists for diagnosis of asthma, lower for the AC/DC tool versus PCPs and pulmonologists for the diagnosis of COPD, and higher for the AC/DC tool versus PCPs and pulmonologists for the diagnosis of ACO (**Table 3**).

Diagnostic accuracy by case difficulty

The proportion of cases that PCPs and pulmonologists correctly diagnosed declined with increasing case difficulty as assessed by the expert panel (**Figure 3**). The AC/DC tool showed a notable much higher percentage of accuracy for the hardest cases across all 3 categories, than PCPs and pulmonologists (**Figure 3**) (note that the study was not powered to determine statistical significance in the diagnostic accuracy of the tool, PCPs and pulmonologists by case difficulty).

Inter- and intra-rater agreement

The Fleiss' kappa for inter-rater agreement for diagnostic consensus was higher among pulmonologists than PCPs across all diagnoses [0.29 (95% CrI: 0.25–0.33) and 0.19 (95% CrI: 0.16–0.22), respectively], as was the intra-rater reliability [0.55 (95% CrI: 0.51–0.59) and 0.48 (95% CrI: 0.44–0.52), respectively]. The inter-rater agreement was high for both definitions used in the AC/DC algorithm (**Figure 4**).

Performance of the AC/DC tool and physicians in exploratory case set

The diagnostic accuracy for both PCPs and pulmonologists in the exploratory case set was the same as that in the primary set, while the diagnostic accuracy of the AC/DC tool was lower in the exploratory case set (**Figure 5**).

Discussion

This multinational, non-interventional, observational study utilised de-identified real-life clinical practice case data to determine the diagnostic accuracy of the AC/DC tool (developed

by machine learning from data in an electronic medical record database) versus PCPs and pulmonologists in diagnosis of asthma, COPD and ACO in patients aged ≥35 years. The primary objective of this study was met; average diagnostic accuracy of the AC/DC tool for these diagnoses was superior to PCPs and non-inferior and superior to pulmonologists. Furthermore, the diagnostic accuracy of pulmonologists was superior to PCPs, as might be expected by virtue of their medical specialisation.

In this study, the AC/DC tool displayed greater sensitivity for diagnosing asthma and COPD cases than PCPs or pulmonologists, with accuracy and precision values for the AC/DC tool being similar to those reported elsewhere for other machine learning models.[17, 24] However, when diagnosing ACO, diagnostic performance of the AC/DC tool was considerably lower than that of PCPs and pulmonologists. While the small sample size (n=7) might have contributed to this finding, there are several other reasons that might also explain these results. Firstly, machine-learning algorithms can struggle when faced with class imbalance, and patients with ACO were the least common class of patients in the training data.[17] Secondly, pulmonologists also had the lowest sensitivity for this diagnosis, which might be because of variations in definitions and perceptions of this disease between physicians and countries [4]. Indeed, some countries do not have a specific definition for ACO in their guidelines, and neither GINA nor GOLD considers ACO to be a specific diagnosis.[3, 4] Thirdly, the features used in development of the AC/DC tool may not be ideally suited for distinguishing ACO from COPD. One of the characteristics that the clinicians in this study were presented was age at onset of respiratory symptoms. A younger age at onset of symptoms is one of the features that, in a patient with persistent airflow limitation, drives a clinical decision towards ACO. However, data for age at onset of respiratory symptoms were not recorded in the Optum® database, and thus were not included in the development of AC/DC. Moreover, while ACO is an interesting construct, there is a lack of double-blind

randomised clinical studies on the treatment of ACO, with current safety recommendations being based on observational studies.[11, 12, 25]

The diagnostic consensus (inter- and intra-rater agreement) was higher among the pulmonologists than PCPs, but both lower than AC/DC tool. As to be expected, it was more likely that two pulmonologists would agree on a diagnosis than two PCPs because of their specialisation. In contrast, the AC/DC tool was extremely consistent as all the algorithms always produced a similar result from the same inputs. In addition, the difference in diagnostic accuracy between AC/DC tool and both PCPs and pulmonologists increased with increasing difficulty of diagnosis, as assessed by the expert panel. These results suggest that the AC/DC tool has the potential to improve accuracy and specificity in the differential diagnosis of asthma and COPD, especially for more difficult to diagnose cases. It should also be noted that these results are from single point in time estimates rather than longitudinal data; the use of longitudinal data in a primary care setting would allow PCPs to determine a response to treatment that could support a clinical diagnosis – this might, in part, explain the diagnostic accuracy of PCPs compared with the AC/DC tool and pulmonologists.

When the AC/DC tool misdiagnosed cases, it categorised patients with asthma as having a high probability of COPD and tended to assign a COPD diagnosis to ACO. However, this was not observed with PCPs or pulmonologists, who more often misclassified patients with asthma as having ACO, and patients with COPD as having ACO or asthma. GINA[4] recommends patients with asthma or ACO should receive an inhaled-corticosteroid-based therapy as it reduces the risk of hospitalisation or death,[11-13, 25] and many COPD patients may be safely treated with bronchodilators alone.[3] Thus, for safety reasons, modifying the training of the algorithm is essential to reflect the consequences of misdiagnosis.

Currently, some biomarkers are used as surrogates for diagnosis of airway diseases due to limitations/availability of spirometry and to guide pharmacotherapy [4]. Analysis of exhaled

nitric oxide and sputum or blood eosinophil count are sometimes used to determine steroid responsiveness and adjustment of anti-inflammatory therapies in patients with asthma. Similarly, club cell secretory protein-16, surfactant protein D and fibrinogen can predict severity and risk of exacerbations in patients with COPD.[26] Blood eosinophil count was evaluated during the development of the AC/DC tool but it did not have any impact on the outcomes so was not included in the tool. In addition, the literature does not report use of any biomarkers in machine learning models for diagnosis of airway diseases.

Is should be noted that the tool is being evaluated here based on current specialist practice rather than diagnostic guidelines, which could, in theory, lead to the reinforcement of common clinical errors in diagnosis. However, this external validation study has allowed assessment of the model performance based on the consensus diagnosis of a panel of experts, who are highly familiar with existing guidelines and highly experienced in the field, which should avoid the routine application of diagnostic criteria and provide the most reliable diagnosis possible given the available information.

The AC/DC tool was designed to aid the physician in asthma and COPD diagnosis with higher accuracy after other diseases have already been ruled out. Thus, this tool is not intended as a standalone model to rule out all diseases and only accept asthma, COPD or ACO. This was particularly evident when the performance of the AC/DC tool was compared between the two (primary and exploratory) case sets. The diagnostic accuracy of the AC/DC tool declined from the primary to exploratory case set because the tool does not have the option to diagnose a patient as "other" and so it either rejects the case or misdiagnoses it as asthma, COPD or ACO. An advantage of the AC/DC tool is it provided a higher inter-rater agreement for the diagnosis of a specific disease across all diagnoses, while there was greater variability in the decisions reported by both PCPs and pulmonologists.

This analysis has several limitations that need to be considered during clinical decision-making processes: (i) the study included only patients aged ≥35 years, and the AC/DC tool cannot be used in younger patients; (ii) the AC/DC tool is not intended to diagnose patients on its own but rather to be used in addition to spirometry to aid physicians in the differential diagnosis of asthma, COPD and ACO, so it might have been worthwhile to include a group of physicians aided by the AC/DC tool in this study; (iii) although an electronic review of a case file may differ substantially from a face-to-face diagnosis in a physician's practice, the performance of the two physician groups aligned reasonably well with the published literature and the expectation that pulmonologists would outperform PCPs;[24, 27] (iv) the assumption that the clinicians have already ruled out all other potential causes of respiratory symptoms may not be clinically relevant, given (for example) the high cost of cardiac investigations for patients presenting with breathlessness; (v) the distinction between "history of allergic rhinitis" and "history of chronic rhinitis", both of which were significant in the development of the AC/DC tool, may not be clear in clinical practice; (vi) limited data on ACO were available in the database to train the AC/DC tool; however, the provision of data such as age of onset and reversibility to PCPs and pulmonologists did not improve their diagnostic accuracy versus the AC/DC tool; (vii) physicians were included in this analysis were from the IPSOS database rather than from random sampling for PCPs and pulmonologists; (viii), PCPs often also rely on social determinants of health, rather than spirometry.

The results of this study should be considered in the context of two assumptions: (i) the expert panel diagnoses of asthma, COPD, ACO and "other diseases" are accurate for each patient, whereas the expert panel members were provided with only brief clinical details for the purpose of assigning a diagnosis; (ii) the AC/DC tool classified each case only to asthma,

COPD or ACO and did not have the "other" diagnostic option – unlike PCPs and pulmonologists for the primary analysis.

Further validation and assessment of the AC/DC tool is required given the lower performance for diagnosing ACO and the risk of hospitalisation and death if patients with asthma or ACO are given a diagnosis of COPD and treated with bronchodilators alone.[11, 12, 25] The AC/DC tool accurately separates asthma from COPD, while ACO diagnosis is not sensitive or specific so should prompt reconsideration by the clinicians to put the patients in the asthma pathway to be safe.

The tool is currently under development and the options for meeting the regulatory requirements to make it available to physicians in the form of "software as a medical device" are currently being evaluated. Current discussions suggest that once the physician rules out other diagnoses and concludes that their patient has either asthma or COPD, they will ask the patient a series of 5 questions about their symptoms that are entered into a smartphone/computer/tablet, together with their spirometry data (using a portable or other spirometer) from which they will then obtain the output of the AC/DC tool, which they can take into account in their diagnosis. It is anticipated that this will take place at the physician's office and take no longer than 3-5 minutes to complete (**Figure 6**).

Subject to the further validation and safety considerations described above, the tool has the potential to support a range of clinicians including nurse practitioners, PCPs, pulmonologists and respiratory experts functioning across healthcare facilities such as mini-clinics, outpatient or satellite care centers, and large hospitals, in distinguishing between asthma and COPD in patients aged ≥35 years in whom other cause of respiratory symptoms have been excluded. The non-invasive tool takes about 3 to 5 minutes to evaluate a patient, thereby benefitting clinicians with busy schedules, and the FEV_1 values generated through any spirometer can

serve as an input for the tool. It could be cost-effective and time-saving as fewer patient visits could be required to arrive at a diagnosis of asthma or COPD.

Conclusions

Overall, the AC/DC tool demonstrated superior diagnostic accuracy compared with PCPs and pulmonologists for correctly diagnosing patients with asthma and COPD, but not patients with ACO, as long as "other" diagnoses can be ruled out before applying the AC/DC tool. The AC/DC tool has the potential to be an aid in the differential diagnosis of patients with asthma and COPD aged ≥35 years and provide a valuable additional source of information to supplement the final decision-making by practicing physicians.

Data sharing

Novartis is committed to sharing access to patient-level data and supporting documents from eligible studies with qualified external researchers. These requests are reviewed and approved by an independent review panel on the basis of scientific merit. All data provided are anonymised to respect the privacy of patients who have participated in the trial in line with applicable laws and regulations.

Declaration of interest

Authors received no compensation related to development of the manuscript. Janwillem W.H. Kocks reports grants, personal fees and non-financial support from AstraZeneca, grants, personal fees and non-financial support from Boehringer Ingelheim, grants and personal fees from Chiesi Pharmaceuticals, grants, personal fees and non-financial support from GSK, grants and personal fees from Novartis, grants from Mundi Pharma, grants from TEVA, outside the submitted work, all paid to his institution; and Janwillem W. H. Kocks holds 72.5% of shares in the General Practitioners Research Institute. Alan Kaplan is a medical advisor or is on the speaker's bureau for AstraZeneca, Behring, Boehringer Ingelheim, Cipla, Covis, GSK, Merck Frosst, Pfizer, Novartis, Novo Nordisk, Sanofi, Eisai, Teva, Trudel and Valeo. J. Mark FitzGerald has received honoraria as a member of speaker bureaus and advisory boards for AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, and Sanofi Regeneron, outside the area of work.

Konstantinos Kostikas has received honoraria for presentations and consultancy fees from AstraZeneca, Boehringer Ingelheim, CSL Behring, Chiesi, ELPEN, GSK, Menarini, Novartis and Sanofi (paid to the University of Ioannina). His department has received funding and grants from AstraZeneca, Boehringer Ingelheim, Chiesi, Innovis, ELPEN, GSK, Menarini, Novartis and NuvoAir (paid to the University of Ioannina). David Price has board membership with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan,

Mundipharma, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceuticals, Thermofisher; consultancy agreements with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Mylan, Mundipharma, Novartis, Pfizer, Teva Pharmaceuticals, Theravance; grants and unrestricted funding for investigator-initiated studies (conducted through Observational and Pragmatic Research Institute Pte Ltd) from AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan, Merck, Zentiva (Sanofi Generics), Mundipharma, Novartis, Pfizer, Regeneron Pharmaceuticals, Respiratory Effectiveness Group, Sanofi Genzyme, Teva Pharmaceuticals, Theravance, UK National Health Service; payment for lectures/speaking engagements from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, Kyorin, Mylan, Mundipharma, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceuticals; payment for the development of educational materials from Mundipharma, Novartis; payment for travel/ accommodation/meeting expenses from AstraZeneca, Boehringer Ingelheim, Mundipharma, Mylan, Novartis, Thermofisher; funding for patient enrolment or completion of research from Novartis; stock/stock options from AKL Research and Development Ltd which produces phytopharmaceuticals; owns 74% of the social enterprise Optimum Patient Care Ltd (Australia and UK) and 74% of Observational and Pragmatic Research Institute Pte Ltd (Singapore); 92.61% of Observational and Pragmatic Research Institute Pte Ltd (Singapore); 5% shareholding in Timestamp which develops adherence monitoring technology; is peer reviewer for grant committees of the Efficacy and Mechanism Evaluation programme, and Health Technology Assessment; personal fees from Airway Vista Secretariat, EPG Communication Holdings Ltd, FIECON Ltd, Fieldwork International, OM Pharma SA, PeerVoice, Phadia AB, Spirosure Inc, Strategic North Limited, Synapse Research Management Partners S.L, Talos Health Solutions, WebMD Global LLC, outside the submitted work; and was an expert witness for GlaxoSmithKline. Helen K. Reddel has

provided independent advice on advisory boards/steering committees for AstraZeneca, Chiesi, GSK, Novartis, and Sanofi/Genzyme; has provided independent medical education at symposia funded by AstraZeneca, Boehringer Ingelheim, Chiesi, Getz, GSK, Novartis, Sanofi/Genzyme and Teva; has provided independent consulting for AstraZeneca and Novartis; and her institute has received unrestricted research grants from AstraZeneca, GSK, and Novartis. Ioanna Tsiligianni has received fees for consultancies from Novartis, Boehringer Ingelheim, AstraZeneca and GSK and grants for research activities from the Greek, GSK, AstraZeneca, Boehringer Ingelheim and ELPEN. Claus F. Vogelmeier gave presentations at symposia and/or served on scientific advisory boards sponsored by AstraZeneca, Boehringer Ingelheim, CSL Behring, Chiesi, GSK, Grifols, Menarini, Novartis, Nuvaira, and Aerogen. Hui Cao, Björn Holzhauer, Sebastien Bostel and Paul Mastoridis are employees of Novartis and hold stock in Novartis Pharma AG.

Author's contribution

The authors contributed to the preparation of the manuscript draft, along with critical review and approval of manuscript for submission to the journal. All authors contributed to intellectual content of the manuscript and approved for publication.

Acknowledgement

The studies were funded by Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States. Under the direction of authors, Rabi Panigrahy, Preethi B and Ian Wright (professional medical writers; Novartis) assisted in the preparation of this article in accordance with the third edition of Good Publication Practice (GPP3) guidelines (http://www.ismpp.org/gpp3).

References

1. Llanos JP, Ortega H, Germain G, Duh MS, Lafeuille MH, Tiggelaar S, et al. Health characteristics of patients with asthma, COPD and asthma-COPD overlap in the NHANES database. Int J Chron Obstruct Pulmon Dis 2018; 13:2859-68.
2. Buist AS. Similarities and differences between asthma and chronic obstructive pulmonary disease: treatment and early outcomes. Eur Respir J Suppl 2003; 39:30s-5s.
3. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2022. Global Strategy for The Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Diseases. Available at: https://goldcopd.org/2022-gold-reports-2/. Accessed: June 01 2022., 2021.
4. Global Initiative for Asthma (GINA) 2022. Global Strategy for Asthma Management and Prevention. Available at: https://ginasthma.org/wp-content/uploads/2022/07/GINA-Main-Report-2022-FINAL-22-07-01-WMS.pdf Accessed: June 01, 2022 2022.
5. Hosseini M, Almasi-Hashiani A, Sepidarkish M, Maroufizadeh S. Global prevalence of asthma-COPD overlap (ACO) in the general population: a systematic review and meta-analysis. Respir Res 2019; 20:229.
6. Ho T, Cusack RP, Chaudhary N, Satia I, Kurmi OP. Under- and over-diagnosis of COPD: a global perspective. Breathe (Sheff) 2019; 15:24-35.
7. Aaron SD, Boulet LP, Reddel HK, Gershon AS. Underdiagnosis and Overdiagnosis of Asthma. Am J Respir Crit Care Med 2018; 198:1012-20.
8. Heffler E, Madeira LNG, Ferrando M, Puggioni F, Racca F, Malvezzi L, et al. Inhaled Corticosteroids Safety and Adverse Effects in Patients with Asthma. J Allergy Clin Immunol Pract 2018; 6:776-81.
9. Heffler E, Crimi C, Mancuso S, Campisi R, Puggioni F, Brussino L, et al. Misdiagnosis of asthma and COPD and underuse of spirometry in primary care unselected patients. Respir Med 2018; 142:48-52.

10. Broder MS, Raimundo K, Ngai KM, Chang E, Griffin NM, Heaney LG. Cost and health care utilization in patients with asthma and high oral corticosteroid use. Ann Allergy Asthma Immunol 2017; 118:638-9.

11. Gershon AS, Campitelli MA, Croxford R, Stanbrook MB, To T, Upshur R, et al. Combination long-acting beta-agonists and inhaled corticosteroids compared with long-acting beta-agonists alone in older adults with chronic obstructive pulmonary disease. JAMA 2014; 312:1114-21.

12. Kendzerska T, Aaron SD, To T, Licskai C, Stanbrook M, Vozoris NT, et al. Effectiveness and Safety of Inhaled Corticosteroids in Older Individuals with Chronic Obstructive Pulmonary Disease and/or Asthma. A Population Study. Ann Am Thorac Soc 2019; 16:1252-62.

13. Kaplan A, Cao H, FitzGerald JM, Iannotti N, Yang E, Kocks JWH, et al. Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. J Allergy Clin Immunol Pract 2021; 9:2255-61.

14. Topalovic M, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. Eur Respir J 2019; 53.

15. Badnjevic A, Gurbeta L, Custovic E. An Expert Diagnostic System to Automatically Identify Asthma and Chronic Obstructive Pulmonary Disease in Clinical Settings. Sci Rep 2018; 8:11645.

16. Abdel-Aal A, Lisspers K, Williams S, Adab P, Adams R, Agarwal D, et al. Prioritising primary care respiratory research needs: results from the 2020 International Primary Care Respiratory Group (IPCRG) global e-Delphi exercise. NPJ Prim Care Respir Med 2022; 32:6.

17. Kaplan A, Cao H, Fitzgerald JM, Yang E, Iannotti N, Kocks JWH, et al. Asthma/COPD Differentiation Classification (AC/DC): Machine Learning to Aid Physicians in Diagnosing Asthma, COPD and Asthma-COPD Overlap (ACO). American Thoracic Society International Conference: Am J Resp Crit Care Med, 2020:A6285-A.

18. van de Hei SJ, Flokstra-de Blok BMJ, Baretta HJ, Doornewaard NE, van der Molen T, Patberg KW, 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.

609 19. Head SJ, Kaul S, Bogers AJ, Kappetein AP. Non-inferiority study design: lessons to be
610 learned from cardiovascular trials. Eur Heart J 2012; 33:1318-24.
611 20. Meaden C, Joshi M, Hollis S, Higham A, Lynch D. A randomized controlled trial comparing
612 the accuracy of general diagnostic upper gastrointestinal endoscopy performed by nurse or
613 medical endoscopists. Endoscopy 2006; 38:553-60.
614 21. R Core Team. R: A language and environment for statistical computing. R Foundation for
615 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2019.
616 22. Stan Development Team. RStan: the R interface to Stan. R package version 2.19.2. Available
617 from: http://mc-stan.org/. 2019.
618 23. Honaker J, King G, M. B. Amelia II: A program for missing data. Journal of statistical
619 software 2011; 45:1-47.
620 24. Spathis D, Vlamos P. Diagnosing asthma and chronic obstructive pulmonary disease with
621 machine learning. Health Informatics J 2019; 25:811-27.
622 25. Suissa S, Ernst P. Observational Studies of Inhaled Corticosteroid Effectiveness in COPD:
623 Lessons Learned. Chest 2018; 154:257-65.
624 26. Leung JM, Sin DD. Biomarkers in airway diseases. Can Respir J 2013; 20:180-2.
625 27. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive
626 pulmonary disease (COPD) in asthma patients using electronic medical records. J Am Med
627 Inform Assoc 2009; 16:371-9.

628

Figure Legends

Figure 1: Study design and patients flow for AC/DC validation study

AC/DC, Asthma/COPD Differentiation Classification; ACO, asthma-COPD overlap; COPD, chronic obstructive pulmonary disease; PCP, primary care physician

Figure 2: a) Overall and b) difference in diagnostic accuracy of AC/DC tool, PCPs, pulmonologists in diagnosis of clinical cases of asthma, COPD and ACO (primary case set)

This analysis is based on primary symptom definition (#1): score >0 on ACQ Q4 (dyspnoea)/ACQ Q5 (wheeze)/CCQ Q5 (cough). The differences in the average overall diagnostic accuracy between AC/DC tool and PCPs, and between AC/DC tool and pulmonologists in reference to the expert panel diagnoses were analysed.
AC/DC, Asthma/COPD Differentiation Classification; ACQ, asthma control questionnaire; CCQ, clinical COPD questionnaire; CrI, posterior credible interval; PCP, primary care physician

Figure 3: Diagnostic accuracy of AC/DC tool, PCPs pulmonologists by case difficulty (assigned by the expert panel) in the primary case set

AC/DC, Asthma/COPD Differentiation Classification; COPD, chronic obstructive pulmonary disease; PCPs, primary care physicians
Based on tertiles of average of difficulty ratings of panel members: 1: Easy, 2: Moderately hard, 3: Hard to diagnose. The diagnosis was based on symptom definition #1 (Def. #1) that includes ACQ Q4 score>0 (dyspnoea), ACQ Q5>0 (wheeze) and CCQ Q5>0 (cough), and symptom definition #2 that includes ACQ Q4 score>1/ACQ Q5>1/CCQ Q5>1

Figure 4: Variation in performance within AC/DC tool (algorithms), pulmonologists, and PCPs in the primary case set

AC/DC, Asthma/COPD Differentiation Classification; PCP, primary care physician

Figure 5: Comparison of performance of the AC/DC tool and physicians in primary and exploratory case sets

Data presented as median; error bars represent CrI values

Combined posterior from 100 MIs, each on the patients accepted by ≥1 algorithm(s) for that MI. This analysis is based on primary symptom definition (#1): score >0 on ACQ Q4 (Dyspnea)/ACQ Q5 (Wheeze)/CCQ Q5 (Cough).

AC/DC, Asthma/COPD Differentiation Classification; CrI, posterior credible interval; MD, median difference; MI, multiple imputations; PCP, primary care physician

Figure 6: Potential clinical utility of AC/DC digital diagnostic tool

AC/DC, Asthma/COPD Differentiation Classification; ACO, Asthma/COPD overlap; COPD, chronic obstructive pulmonary disease; HCP, healthcare provider