Artificial intelligence improves bronchoscopy performance: a randomised crossover trial

Kristoffer Mazanti Cold[1,2], Kaladerhan Agbontaen[3], Anne Orholm Nielsen[1,2,4], Christian Skjoldvang Andersen[5], Suveer Singh[3,6,7] and Lars Konge[1,2]

[1]Copenhagen Academy for Medical Education and Simulation, Rigshospitalet, The Capital Region of Denmark, Copenhagen, Denmark. [2]University of Copenhagen, Copenhagen, Denmark. [3]Chelsea and Westminster Hospital, Chelsea, London, UK. [4]Bispebjerg Hospital, Department of Pulmonary Medicine, Copenhagen, Denmark. [5]North Zealand Hospital, Department of Pulmonary Medicine, Copenhagen, Denmark. [6]Royal Brompton Hospital, Chelsea, London, UK. [7]Faculty of Medicine, Imperial College London, London, UK.

Corresponding author: Kristofer Mazanti Cold (kristoffer.mazanti.cold.01@regionh.dk)

Shareable abstract (@ERSpublications)
AI improves bronchoscopy performance across all experience levels by helping the endoscopists to visualise more segments in a more structured order https://bit.ly/3XyxdXu

Cite this article as: Cold KM, Agbontaen K, Nielsen AO, *et al*. Artificial intelligence improves bronchoscopy performance: a randomised crossover trial. *ERJ Open Res* 2025; 11: 00395-2024 [DOI: 10.1183/23120541.00395-2024].

This article has an editorial commentary: https://doi.org/10.1183/23120541.00931-2024

Received: 16 April 2024
Accepted: 17 Aug 2024

Abstract

Rationale Flexible bronchoscopy is an operator-dependent procedure. An automatic bronchial identification system based on artificial intelligence (AI) could help bronchoscopists to perform more complete and structured procedures through automatic guidance.

Methods 101 participants were included from six different continents at the European Respiratory Society annual conference in Milan, 9–13 September 2023. Participants were split into three groups based on experience: novices (0 bronchoscopies), intermediates (1–249 bronchoscopies) and experienced (⩾250 bronchoscopies). The participants performed two bronchoscopies on a realistic physical phantom, one with AI (AmbuBronchoSimulatorTrainingGUIDEv.0.0.1, Prototype version, Ambu) and one Standard procedure. The F1-group received AI guidance for their first procedure, the F2-group for their second. A crossover randomisation controlled for learning by testing. All procedures were automatically rated according to the outcome measures: inspected segments, structured progressions and procedure time.

Results AI guidance caused the participants to inspect more segments (mean difference, paired t-test: +6.0 segments, p<0.001), perform more structured progressions (+5.2 progressions, p<0.001) and spend more time on the procedure (+72 s, p<0.001) compared to their standard procedures. The effects of AI guidance on inspected segments and structured progression were highest for novices but significant for all experience groups: novices (+8.2 segments, p=0.012 and +6.6 progressions, p<0.001), intermediates (+5.7 segments, p=0.006 and +5.1 progressions, p<0.001) and experienced (+4.3 segments, p=0.006 and +3.8 progressions, p<0.016).

Conclusions AI guidance helped bronchoscopists of all experience levels to inspect more segments in a more structured order. Clinical implementation of AI guidance could help ensure and document more complete bronchoscopy procedures in the future.

Introduction

Lung cancer is the leading cause of cancer mortality globally [1]. Lung cancer mortality has been decreasing in Europe [2], but it still ranks the highest [3]. Several countries have implemented screening programmes to decrease the mortality [4], and guidelines towards standardisation have been recommended by the European Respiratory Society (ERS) [5]. With increasing screening programmes, more lesions in their preliminary stages will be detected, increasing the number of flexible bronchoscopies as the primary sampling technique [6]. Bronchoscopy is an operator-dependent procedure that should entail a full and structured inspection of the bronchial tree in patients suspected of lung cancer, as some tumours are not visible on low-dose computed tomography [7], and tumours can even appear in patients not suspected of lung cancer [8]. Navigation through the bronchial maze where the segments look alike can cause confusion, resulting in missed segments. Therefore, the most frequently used assessment tool for bronchoscopy competence is diagnostic completeness (DC), *i.e.*, inspected segments [9]. Novices tend to have a lower DC than more experienced [10], but even experienced bronchoscopists with >500 bronchoscopies performed miss segments [10, 11]. Additionally, a systematic inspection order (structured progression, SP) and procedure efficiency (mean intersegmental time, MIT) have been proposed and shown discriminatory effect of competence [12–14]. Systematic approaches to teach and train in bronchoscopy have been suggested [15–17], but widespread implementation of individual teaching preferences beyond the traditional apprenticeship model of "see one, do one, teach one" is difficult. A recent study found that novices benefitted significantly from training using a novel bronchial lumen identification system based on artificial intelligence (AI) [18]. AI could ensure a full, systematic and efficient inspection of the bronchial tree that could help even experienced bronchoscopists perform more thorough and structured procedures.

The objective of this study was to test the impact of AI-driven anatomical guidance on bronchoscopy performance of endoscopists with varying experience.

Material and methods

Study site, equipment and settings

The study was conducted in a simulated setting at the Clinical Skills Zone at the ERS Annual Conference 2023, 9–13 September in Milan, Italy, to provide a standardised setting with the inclusion of a high number of participants with varying experience levels. The primary investigators (KC and KA) enrolled participants from their booth using two identical set-ups (figure 1) with the following settings: a single use scope (aScop 5 Broncho HD 5.0/2.2, Ambu, Ballerup, Denmark) connected to a portable monitor (aView 2 Advance, Ambu, Ballerup, Denmark) and further displayed on a 27 inch monitor with the AI interface turned on or off (supplementary material 1 – video, Ambu Broncho Simulator, Prototype version, AmbuBronchoSimultatortrainingGUIDEv.0.0.1, Ambu). All procedures were performed on a realistic phantom mimicking the first four to five divisions of the bronchial tree (Bronchoscopy Training Model LM-092, Koken Co., LTH, Tokyo, Japan). The Ambu Broncho Simulator prototype was developed using a neural network for visually recognising anatomical locations using an aScope 5 Broncho HD, in order to provide the user with real-time positioning support during a bronchoscopy training session on a standardised typical anatomical model (Bronchoscopy training model LM-092, Koken Co.).

Participation was voluntary and all participants who came by and wanted to participate were enrolled. When multiple participants wanted to participate, they waited for their turn on a first-come-first-served basis. All participants gave oral and written informed consent before participating in the study and provided their demographic information (supplementary material 2).

FIGURE 1 The Ambu AI feedback with its three feedback features: on screen labels, lung tree diagram and structured progression score.

Study design, outcome measures, randomisation and rating

The study was designed as a randomised crossover trial and adhered to the CONSORT statement for simulation-based trials (figure 2, supplementary material 3) [18]. The participants were instructed to perform two bronchoscopies by the primary investigators. To accommodate for a "learning-by-testing" effect, we randomised the participants into two groups. The F1-group received AI-driven guidance for their first bronchoscopy (AI-First), but not their second bronchoscopy (Standard-Second). The F2-group received AI-driven guidance for their second bronchoscopy (AI-Second) but not their first bronchoscopy (Standard-First). The participants were randomised using Sealed Envelope (Sealed Envelope, London, UK). The participants were stratified by sex as previous research has indicated that sex can affect skill acquisition in the initial stages of the learning process [19], and into three groups by experience level (novices: 0 bronchoscopies; intermediates: 1–249 bronchoscopies; and experienced: ⩾250 bronchoscopies).

The participants were given a brief introduction about the interface of the AI. They were instructed to visualise all 18 segments of the phantom (10 segments on the right side and eight segments on the left with segment 1 and 2 fused together and no segment 7). They could start either on the left or right lung, but were instructed to proceed according to the segment numbering (for the right side: RB1→RB2→RB3) in as little time as possible but with no time constraint, as they were rated by the AI according to the following four outcome measures, which were chosen as they all showed discriminatory abilities in a previous randomised controlled trial using the AI [20], and have been used in previous simulation-based trials [12–14]:

Diagnostic completeness (DC): This was measured as the number of entered bronchial segments. Maximum score for DC was 18 segments.

Structured progress (SP): A point was rewarded every time the bronchoscopist passed from a segment to the immediately succeeding segment (RB1→RB2=1 point, RB2→RB1=0 points) [13].

Procedure time (PT): A timer was started automatically when the main carina was visualised and ended with extraction of the scope.

Mean intersegmental time (MIT): To have a measure for time spent navigating between the segments, PT was divided with DC and referred to as MIT, as PT is an indicator of competence, but does not hold a direct translation to the efficacy of the procedure.

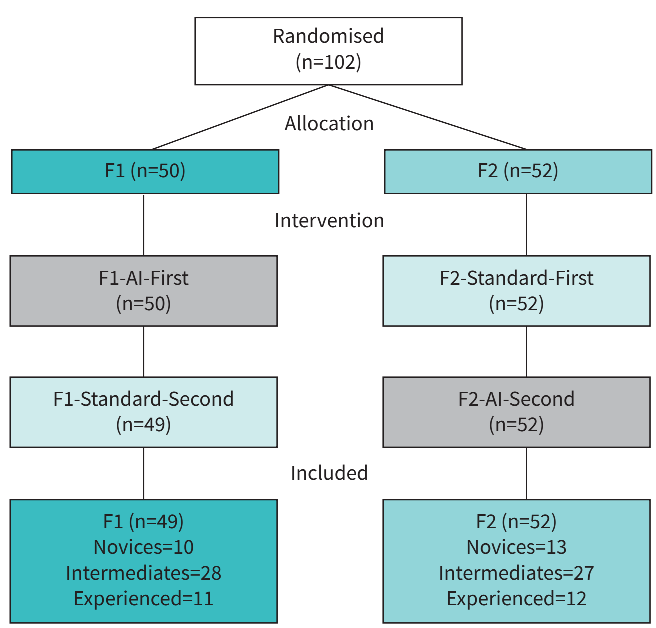

FIGURE 2 Flowchart of the study. Standard procedure: the bronchoscopists did not receive feedback from the artificial intelligence (AI); AI procedure: the bronchoscopists received feedback from the AI. One participant dropped out after the F1-AI-First procedure.

The primary investigators did not provide the participants with any feedback doing the procedures but helped to interpret the scoring by the AI when the procedure was finalised. The AI automatically rated all procedures, as it is an unbiased and automatic rating tool with a high level of validity evidence [21].

Statistical analysis by effect of AI

All statistical testing was conducted for all participants and their subgroups for all four outcome measures (DC, SP, PT and MIT). Statistical testing was completed in R (R version 4.1.2, Vienna, Austria) by K. Cold and in Statistical Package for the Social Sciences (SPSS) version 27 (PASW v27.0; SPSS Inc., Chicago, IL, USA) by L. Konge to ensure statistical robustness and that the test's assumptions were being met. To ensure transparency, all data are available as supplementary material 4.

To test for the effect of AI guidance, all procedures using AI (AI-First and AI-Second) were compared to the procedures without AI (Standard-First and Standard-Second) using paired t-tests. This was done for all participants and for each of the three experience groups. To test for the learning-by-testing effect of a single procedure with AI guidance, Standard-Second procedures were compared to Standard-First procedures using independent samples t-test. Conversely, to test for the learning-by-testing effect of a single procedure without AI guidance, AI-Second procedures were compared to AI-First procedures using independent samples t-test. To test for the validity of using SP, its correlation to DC was calculated using Pearson's correlation, r.

Results

101 participants completed the study (23 novices, 55 intermediates and 23 experienced) from 38 countries of six continents (participant demographic, table 1). One participant was excluded due to dropping out during the first procedure (p_031 F1-AI-First, supplementary material 4). Another participant had their second procedure interrupted by system shut down but redid the second procedure from the beginning (p_046, F2-AI-Second, supplementary material 4).

Comparing AI procedures to standard procedures, the participants visualised more of the bronchial tree; DC (mean difference, +6.0 segments, $p<0.001$), in a more structured order; SP (+5.2 progressions, $p<0.001$), spending longer time doing so; PT (+72 s, $p<0.001$), however, with a higher level of efficiency; MIT (−10.3 s, $p<0.001$). Table 2 shows that these findings were also significant within each experience group (table 2).

TABLE 1 Participant demographics

	Group	F1	F2
Participants, n (female sex)	Total	49 (15)	52 (13)
	Novice	10 (3)	13 (2)
	Intermediate	28 (10)	27 (11)
	Experienced	11 (2)	12 (0)
Age years	Total	36.8±10.0	36.5±9.1
	Novice	29.1±3.8	34.0±10.2
	Intermediate	36.4±9.8	34.4±6.4
	Experienced	44.5±8.8	43.9±9.9
Countries represented[#]**, n**	Total	24	27
	Novice	6	10
	Intermediate	19	16
	Experienced	10	8
Bronchoscopies performed	Total	242±540	373±1415
	Novice	0±0	0±0
	Intermediate	83±78	75±64
	Experienced	868±905	1446±2764
Bronchoscopies performed within last 6 months	Total	39±97	33±49
	Novice	0±0	0±0
	Intermediate	19±22	25±27
	Experienced	124±182	85±71

Values are presented as mean±SD, unless otherwise indicated. Novice (0 bronchoscopies, n=23), intermediate (10–200 bronchoscopies, n=55) and experienced (⩾250 bronchoscopies, n=23). [#]: for countries, the total is lower than the sum of the individual groups, as similar nationalities were represented in different groups.

TABLE 2 Artificial intelligence (AI) procedures (F1-AI-First and F2-AI-Second) *versus* standard procedures (F1-Standard-Second and F2-Standard-First)

AI procedures *versus* standard procedures	AI procedures	Standard procedures	Mean difference	p-value
Total DC, segments	16.2±3.4	10.2±4.8	6.0	<0.001
Total SP, progressions	9.4±5.0	4.2±3.7	5.2	<0.001
Total PT, seconds	346±174	273±128	72	<0.001
Total MIT, seconds	23±14.8	33±23.3	−10	<0.001
Novice DC, segments	16.2±3.4	8.0±3.7	8.2	0.012
Novice SP, progressions	8.4±4.0	1.8±1.7	6.6	<0.001
Novice PT, seconds	467±200	351±169	116	<0.001
Novice MIT, seconds	30±12.6	52±32.3	−22	0.002
Intermediate DC, segments	16.4±3.0	10.7±4.7	5.7	0.006
Intermediate SP, progressions	9.7±4.9	4.6±3.5	5.1	<0.001
Intermediate PT, seconds	323±155	269±104	54	<0.001
Intermediate MIT, seconds	21±14.4	31±17.5	−10	<0.001
Experienced DC, segments	15.8±4.4	11.4±5.3	4.3	0.006
Experienced SP, progressions	9.7±6.0	5.8±4.4	3.8	0.016
Experienced PT, seconds	278±131	205±90	73	<0.001
Experienced MIT, seconds	21±16.4	19±8.6	2	0.53

All 101 AI procedures (F1-AI-First and F2-AI-Second) are pairwise compared for each participant to all 101 standard procedures (F1-Standard-Second and F2-Standard-First) using a paired t-test and for the three subgroups alike: novice (0 bronchoscopies, n=23), intermediate (1–249 bronchoscopies, n=55) and experienced (⩾250 bronchoscopies, n=23). Values are presented as mean±SD. DC: diagnostic completeness; SP: structured progress; PT: procedure time; MIT: mean intersegmental time.

Comparing standard procedures between the groups, the F1-Standard-Second visualised more segments than the F2-Standard-First; DC (mean difference +3.7 segments, p<0.001) in a more structured order; SP (+2.2 progressions, p=0.002), however without increasing the PT significantly (+20 s, p=0.45), but decreasing MIT (−11 s, p=0.016). This was the same pattern for both experienced and intermediate participants but not for novices (table 3).

For the AI procedures (F1-AI-First and F2-AI-Second), none of the outcome measures between the two groups were statistically significant (supplementary material 5).

DC correlated significantly with SP (Pearson's r=0.76, p<0.001) in total and for each subgroup: novices (0.84, p<0.001), intermediates (0.73, p<0.001) and experienced (0.77, p<0.001).

TABLE 3 Difference in performance between standard procedures (F1-Standard-Second and F2-Standard-First)

Standard procedures	F1-Standard-Second	F2-Standard-First	Mean difference	p-value
Total DC, segments	12.2±4.1	8.4±4.7	3.7	<0.001
Total SP, progressions	5.4±3.8	3.2±3.8	2.2	0.002
Total PT, seconds	283±108	264±145	19	0.45
Total MIT, seconds	28±17.2	39±27.0	−11	0.016
Novice DC, segments	8.5±3.0	7.6±4.3	0.9	0.56
Novice SP, progressions	1.8 ±1.4	1.8±1.9	−0.1	0.94
Novice PT, seconds	318±152	376±183	−57	0.42
Novice MIT, seconds	41±22.4	60±36.9	−19	0.16
Intermediate DC, segments	12.6±3.6	8.7±4.9	3.94	0.001
Intermediate SP, progressions	5.5±3.2	3.6±3.5	1.9	0.037
Intermediate PT, seconds	293±93	243±110	50	0.07
Intermediate MIT, seconds	27±14.5	36±19.4	−9	0.05
Experienced DC, segments	14.4±4.2	8.6±4.9	5.6	0.007
Experienced SP, progressions	8.3±4.3	3.7±3.7	4.6	0.011
Experienced PT, seconds	225±81	188±99	38	0.33
Experienced MIT, seconds	18±11.0	20±5.6	−2	0.57

All standard procedures (F1-Standard-Second and F2-Standard-First) are compared, using independent samples t-test in total and for the subgroups: novice (0 bronchoscopies, n=23), intermediate (10–200 bronchoscopies, n=55) and experienced (⩾250 bronchoscopies, n=23). Values are presented as mean±SD. DC: diagnostic completeness; SP: structured progress; PT: procedure time; MIT: mean intersegmental time.

Discussion

This study is the first to explore the efficacy of AI-driven guidance in bronchoscopy across all experience levels (table 1).

When comparing the AI-guided procedures with the standard procedures without AI guidance, the mean difference for DC was +6.0 segments (p<0.001) and for +SP 5.2 progressions (p<001), indicating both a more thorough and a more structured inspection of the bronchial tree when using AI. Novices benefitted most (DC=+8.2 segments, SP=+6.6 progressions) from AI-driven guidance; however both Intermediates (DC=+5.7 segments, SP=+5.1 progressions) and even experienced bronchoscopists (DC=+4.3 segments, SP=+3.8 progressions) benefitted significantly from it (table 2).

A bronchoscopy should entail a full inspection of the bronchial tree [22, 23], and therefore DC is the most frequently used assessment measure for bronchoscopy performance [9]. Experienced bronchoscopists have a higher level of diagnostic yield [24], but the experienced bronchoscopists did not visit all segments although they were instructed to do so. This finding is in accordance with other simulation-based bronchoscopy studies [11–14, 25, 26]. One clinical study analysed 27 bronchoscopy recordings from nine experienced consultants with a mean of 1506 bronchoscopies performed previously; with an average of one missed lobe per procedure [10]. A learning curve study in bronchoscopy illustrated a mean asymptotic level at the 75th bronchoscopy, but that some trainees need up to 250 bronchoscopies to reach their asymptotic level. This is in accordance with another learning curve study in bronchoscopy that identified fast and slow learners [27]. The improvement of DC is therefore important even for intermediates and experienced bronchoscopists, and the AI has potential to do so. No large-scale studies exploring DC for clinical bronchoscopies exist, but the AI should be tested in a clinical setting both to assess DC and whether AI-driven guidance can improve it.

After just one procedure with the AI, the F1-Standard-Second performed better compared to the F2-Standard-First (table 3). This could be due to learning by testing as the F1-group might be more familiar with the phantom for their second procedure than for the F2-group for their first procedure. However, when doing the opposite comparison by testing the AI procedures, no significant differences regarding the outcome measures were found (supplementary material 5), indicating that the F2-group did not gain an advantage by performing a standard procedure before using the AI. Additionally, an increase for the novices was not detectable with just one procedure with the AI, as F1-Standard-Second did not perform better than F2-Standard-First (table 3). This finding could be attributed to the novices' sparse skill level and very limited anatomical knowledge. Since the AI improved the performance of all groups, a training effect could be possible for intermediates and experienced as well, as observed for novices in a previous trial, where training with the AI made novice participants perform better, even with the AI turned off [20]. Future training and clinical implementation studies using the AI for intermediates and experienced should be performed.

All participants were instructed to progress according to the SP score. The most recent systematic review on the effectiveness of simulation-based bronchoscopy training found the score to have a high degree of validity evidence and suggested implementing it into future studies [9]. SP correlated significantly with DC (Pearson's r, p-value: 0.76, p<0.001) in total and for each subgroup: novices (0.84, p<0.001), intermediates (0.73, p<0.001) and experienced (0.77, p<0.001). These findings supply further validity evidence for the SP score as those bronchoscopists who progress according to the score visit more segments. However, SP should be interpreted with caution. Experienced bronchoscopists might develop their own route for navigating the bronchial tree, and if done systematically progressing from RB3→RB2→RB1 can be as useful as progressing according to the SP score. Different bronchial segment nomenclatures, like the numerical used in this study or the anatomical (R1=right upper lobe, apical segment, *etc.*), are used. All intermediates and experienced were familiar with the numerical nomenclature used for the SP score and were asked to progress according to it. Unfortunately, we did not register which nomenclature the participants used in their daily practice, which could have impacted their performance. As the only automatic measure of structuredness and thereby anatomical knowledge, we propose to train and assess novice bronchoscopists according to the SP score like done in this study.

The AI increased the performance regarding DC and SP, however also increasing PT. PT is widely used [9], as it is very intuitively understood and easy to monitor. However, it holds no intrinsic value towards the efficacy of the procedure. MIT is a direct indicator of the efficacy of the procedure by adjusting to DC (*i.e.* PT/DC) [13]. Procedures with AI-driven guidance had a lower MIT than standard procedures (mean difference, paired t-test: −10.3 s, p<0.001). The AI therefore helped the participants to spend a significantly shorter amount of time navigating between the segments, and therefore the increase in PT *per se* should not be regarded as an undesired side-effect of the AI. We therefore recommend future studies to add MIT when using PT as an outcome measure, as an indication towards the efficacy of the procedure.

The longer PT might also be due to getting familiar with a new system, as the participants were tested by their first procedure with the AI. Training with the AI decreased PT for novices in a previous trial [20]. This study only tested the use of AI for one bronchoscopy, and it must be assumed that for a potential second and third procedure with the AI, participants will be more familiar with it and be able to interact more efficiently with it. Future studies should therefore examine the impact of training with the AI on consecutive procedures for the assessment of learning curves.

Our study has several strengths. This is the largest study exploring AI in bronchoscopy, with 101 participants ranging all experience levels and spanning six continents, providing a high level of external validity. We used a randomised cross-over design to assess bronchoscopy performance with and without the AI in accordance with current guidelines for randomised controlled trials [18]. We used the AI to assess bronchoscopy performance, which is an automatic and unbiased tool with solid validity evidence [21]. We did not perform a power calculation but included every participant with interest in participating in the study throughout the conference. Only one participant dropped out and only one had their second procedure interrupted by system shut down. The study has the highest number of participants of any study exploring bronchoscopy performance and achieved a high level of participants compared to other simulation-based studies in general [28].

There are however limitations in the study. Framework and reporting tools in endoscopy exist when developing AI [29]. Private companies do not publicly report on the development and internal validation when developing AI. We therefore only tested the AI application, as the training and validation is a company secret kept by the provider, Ambu A/S. However, the Ambu A/S informed us that the AI was developed by their AI team based on machine learning using a neural network. We did not test how the AI would function on different phantoms with different anatomical variations, and therefore do not know if the AI system would be compatible with different bronchoscopy models. We performed the study in a simulation-based setting for standardisation of the procedure and environment and only tested for standard navigational skills; however, bronchoscopy is a more complex procedure with biopsies, patient monitoring, coughing and tumour detection, which the phantom does not facilitate. Extrapolations of our result must be done with absolute caution, *e.g.* that novices with AI seem to approximate intermediate performance without AI regarding procedure efficiency (MIT, mean±SD): 30±12.6 *versus* 31±17.5 s, and intermediates with AI to approximate expert performance without AI regarding MIT: 21±14.4 *versus* 19±8.6 s (table 2). As clinical impact of the AI was not studied, transfer or testing in a clinical setting should be conducted for future studies to test direct clinical impact [22, 30]. We only tested two bronchoscopies for each participant and not how AI impacted consecutive procedures. Additionally, the AI does not test for wall collisions; however an AI co-pilot for safe steering has been developed [31, 32], highlighting how AI might help assist different aspects of the bronchoscopy procedure [33]. We did not systematically ask participants if they believed that the AI system could improve their or their colleagues' performance. Such information could have proven useful, but due to the design of the study at a conference with limited time for each participant, we decided not to hand out questionnaires.

In conclusion, bronchoscopists of all experiences perform better with AI than without, by visiting more segments, in a more structured order, and more efficiently.

Provenance: Submitted article, peer reviewed.

Author contributions: K.M. Cold is the guarantor of this manuscript, and takes responsibility for the content of the manuscript including the data and analysis. Conceptualisation and design: all authors. Administrative support: all authors. Provision of study materials: all authors. Data analysis and interpretation: all authors. Manuscript writing: all authors. Final approval of manuscript: all authors. Agreement to be accountable for all aspects of the work: all authors.

Conflict of interest: K.M. Cold received funding from Ambu regarding The CoRS-feedback study in colonoscopy (NCT04862793). S. Singh has received funding from Ambu A/S. L. Konge has annotated clinical bronchoscopy videos for Ambu's development of the AI system. The other authors have no conflicts of interest to disclose.

References

1 Bray F, Ferlay J, Soerjomataram I, *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394–424.
2 Jani C, Marshall DC, Singh H, *et al.* Lung cancer mortality in Europe and the United States between 2000 and 2017: an observational analysis. *ERJ Open Res* 2021; 7: 00311-2021.
3 Ferlay J, Colombet M, Soerjomataram I, *et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer* 2018; 103: 356–387.

4 Baldwin D, O'Dowd E, Tietzova I, *et al.* Developing a pan-European technical standard for a comprehensive high-quality lung cancer CT screening program. An ERS technical standard. *Eur Respir J* 2023; in press [https://10.1183/13993003.00128-2023].
5 Kauczor H-U, Bonomo L, Gaga M, *et al.* ESR/ERS white paper on lung cancer screening. *Eur Respir J* 2015; 46: 28–39.
6 Mhanna L, Guibert N, Mazieres J, *et al.* Diagnostic accuracy of flexible bronchoscopy for lung cancer diagnosis in an expert center. *Eur Respir J* 2018; 52: Suppl. 62, PA2767.
7 van't Westeinde SC, Horeweg N, Vernhout RM, *et al.* The role of conventional bronchoscopy in the workup of suspicious CT scan screen-detected pulmonary nodules. *Chest* 2012; 142: 377–384.
8 Gottlieb M, Clementsen PF, Hakansson K, *et al.* [Surprising squamous cell carcinoma in the right upper lobe at bronchoscopy in a heavy smoker]. *Ugeskr Laeger* 2014; 176: V05140262.
9 Gerretsen ECF, Chen A, Annema JT, *et al.* The effectiveness of flexible bronchoscopy simulation-based training: a systematic review. *Chest* 2023; 164: 952–962.
10 Konge L, Larsen KR, Clementsen P, *et al.* Reliable and valid assessment of clinical bronchoscopy performance. *Respiration* 2012; 83: 53–60.
11 Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. *Respiration* 2008; 76: 92–101.
12 Colt HG, Crawford SW, Galbraith O 3rd. Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest* 2001; 120: 1333–1339.
13 Cold KM, Svendsen MBS, Bodtger U, *et al.* Using structured progress to measure competence in flexible bronchoscopy. *J Thorac Dis* 2020; 12: 6797–6805.
14 Konge L, Arendrup H, von Buchwald C, *et al.* Using performance in multiple simulated scenarios to assess bronchoscopy skills. *Respiration* 2011; 81: 483–490.
15 Cold KM, Vamadevan A, Nielsen AO, *et al.* Systematic bronchoscopy: the four landmarks approach. *JoVE* 2023; 196: e65358.
16 Murgu SD, Kurman JS, Hasan O. Bronchoscopy education: an experiential learning theory perspective. *Clin Chest Med* 2018; 39: 99–110.
17 Davoudi M, Quadrelli S, Osann K, *et al.* A competency-based test of bronchoscopic knowledge using the Essential Bronchoscopist: an initial concept study. *Respirology* 2008; 13: 736–743.
18 Cheng A, Kessler D, Mackinnon R, *et al.* Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Simul Healthc* 2016; 11: 238–248.
19 Ali A, Subhi Y, Ringsted C, *et al.* Gender differences in the acquisition of surgical skills: a systematic review. *Surg Endosc* 2015; 29: 3065–3073.
20 Cold KM, Xie S, Nielsen AO, *et al.* Artificial Intelligence improves novices' bronchoscopy performance: a randomized controlled trial in a simulated setting. *Chest* 2024; 165: 405–413.
21 Cold KM, Agbontaen K, Nielsen AO, *et al.* Artificial Intelligence for automatic and objective assessment of competencies in flexible bronchoscopy. *J Thorac Dis* 2024; 16: 5718–5726.
22 Cold KM, Konge L. Simulation-based training in flexible bronchoscopy: best practices and future directions. *Chest* 2023; 134: 820–821.
23 Herth FJF. Bronchoscopic techniques in diagnosis and staging of lung cancer. *Breathe (Lausanne, Switzerland)* 2011; 7: 324–337.
24 Kemp SV, El Batrawy SH, Harrison RN, *et al.* Learning curves for endobronchial ultrasound using cusum analysis. *Thorax* 2010; 65: 534–538.
25 Cold KM, Svendsen MBS, Bodtger U, *et al.* Automatic and objective assessment of motor skills performance in flexible bronchoscopy. *Respiration* 2021; 100: 347–355.
26 Ost D, DeRosiers A, Britt EJ, *et al.* Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med* 2001; 164: 2248–2255.
27 Mema B, Mylopoulos M, Tekian A, *et al.* Using learning curves to identify and explain growth patterns of learners in bronchoscopy simulation: a mixed-methods study. *Acad Med* 2020; 95: 1921–1928.
28 Cook DA, Hatala R. Got power? A systematic review of sample size adequacy in health professions education research. *Adv Health Sci Educ Theory Pract* 2015; 20: 73–83.
29 Parasa S, Repici A, Berzin T, *et al.* Framework and metrics for the clinical use and implementation of artificial intelligence algorithms into endoscopy practice: recommendations from the American Society for Gastrointestinal Endoscopy Artificial Intelligence Task Force. *Gastrointest Endosc* 2023; 97: 815–824.e1.
30 Cold KM, Konge L. Response. *Chest* 2024; 165: e61.
31 Zhang J, Liu L, Xiang P, *et al.* AI co-pilot bronchoscope robot. *Nat Commun* 2024; 15: 241.
32 Duan X, Xie D, Zhang R, *et al.* A novel robotic bronchoscope system for navigation and biopsy of pulmonary lesions. *Cyborg Bionic Syst* 2023; 4: 0013.
33 Ishiwata T, Yasufuku K. Artificial intelligence in interventional pulmonology. *Curr Opin Pulm Med* 2024; 30: 92–98.