Original research

Digital twin-based bronchoscopy simulator improves training performance and skill retention of novices: a randomised controlled study

Mingming Deng,[1] Fajiu Li,[2] Fei Tang,[3] Wei Chen,[4] Feng Wang ,[5] Chun-Li Tang,[6] Run Tong,[1] Zhen Yang,[7] Weidong Xu,[1,8] Nan Zhang,[9] Yang Xia ,[10] Shiyue Li,[11] Felix J F Herth,[12] Gang Hou [1]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/thorax-2025-223147).

For numbered affiliations see end of article.

Correspondence to
Dr Gang Hou;
hougangcmu@163.com

MD, FL, FT, WC, FW, C-LT, RT and ZY contributed equally.

FJFH and GH are joint senior authors.

Received 10 February 2025
Accepted 3 October 2025

ABSTRACT

Rationale Conventional bronchoscopy training often does not ensure lasting skill retention or adaptability to different anatomies, limiting real-world impact. This study used a digital-twin bronchoscopy simulator with various CT-derived bronchial tree models to better train novices.
Objectives To explore training with various anatomically diverse bronchial tree models in novices' bronchoscopy performance.
Methods 60 bronchoscopy-naive participants were randomly assigned to three groups (n=20 each): control (written instruction only), anatomic-uniformity (trained on one standard bronchial model) and anatomic-variety (trained on multiple patient-derived bronchial models). All participants performed two tests: test 1 on a standard model and test 2 on a new CT-derived model. Both tests were repeated 3 months later to assess skill retention. The primary comparison was between the anatomic-variety and anatomic-uniformity groups.
Measurements and main results 60 participants completed tests 1 and 2. 55 returned at 3 months. In test 1, there were no significant differences between the anatomic-variety and anatomic-uniformity groups in diagnostic completeness (DC, 0 segments, p=0.576), structured progress (SP, 1 correct progression, p=0.091) and procedure time (31 s, p=0.831). In test 2, the anatomic-variety group had significantly higher DC (2.5 segments, p<0.001) and SP (9 progression, p<0.001) than the anatomic-uniformity group. At 3 months, the anatomic-variety group retained superior DC and SP scores in both tests despite slight declines.
Conclusions Training with diverse anatomical models significantly enhanced bronchoscopy performance compared with repetitive practice on a single standardised model with partially maintained learning gains at 3 months.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Simulation-based training has proven effective in facilitating learning of bronchoscopy equal to the Halstedian model of 'see one, do one, teach one'. Established simulator training provides only limited bronchial tree models and may overestimate its training effectiveness. Training with varied bronchial tree models could better simulate clinical scenarios, making this approach a potentially valuable training method.

WHAT THIS STUDY ADDS

⇒ Training on different bronchial tree models significantly improved the trainers' end-of-training bronchoscopy performance, especially on the brand-new model. Their skills were also partially maintained learning gains at 3 months.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study represents the first application of digital twin technology in bronchoscopy training. We created multiple bronchial tree models through this technology to better simulate clinical environments. The training programme applied in this study can be translated into clinical practice.

INTRODUCTION

The objective of flexible bronchoscopy is to navigate through the central airways and identify specific bronchial segments for accurate diagnosis and treatment of patients with lung cancer and other respiratory diseases. However, during the early stages of a trainee's learning curve, there is a lower diagnostic biopsy yield, higher complication rates and heightened patient discomfort.[1 2] These observations could be attributed to the conventional apprenticeship model, in which trainees conduct procedures under the supervision of experienced physicians. Simulation-based training, including high-fidelity and low-fidelity simulators, has proven effective in facilitating learning of bronchoscopy and can spare patients from the initial part of the trainee's learning curve.[1–4] Additionally, a standard phantom of the bronchial tree combined with an artificial intelligence (AI) system could improve novices' diagnostic completeness (DC) and structured progress (SP) scores on the bronchial tree.[5] However, a common weakness of these studies is the limited numbers of bronchial trees used in the training and final exam; more specifically, in previous simulator studies, the number of anatomical models used for training typically ranged from 1 to 6.[6 7] While simulation training on standardised models may enhance technical proficiency in controlled settings,

Check for updates

© Author(s) (or their employer(s)) 2025. No commercial re-use. See rights and permissions. Published by BMJ Group.

To cite: Deng M, Li F, Tang F, *et al*. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thorax-2025-223147

its ability to foster adaptability to real-world clinical complexity remains limited.[8] Conventional simulators often fail to address anatomical variability, such as congenital branching anomalies or tumour-induced airway distortions, commonly encountered in practice. Consequently, skills acquired through repetitive training on uniform models may not translate to improved bronchoscopy performance in a patient setting (Kirkpatrick levels 3–4), underscoring the need for training paradigms that bridge simulated and clinical environments.

Therefore, we developed a digital-twin bronchoscopy simulator of a patient environment. The purpose of this study was to evaluate whether training with various bronchial tree models in this digital-twin-based bronchoscopy simulator enhances novices' end-of-training bronchoscopy performance compared with the classical high-fidelity simulator. Additionally, the study aimed to determine whether this training curriculum supports the retention of bronchoscopy skills in novices 3 months post-training. Some of the results of these studies have been previously reported in the form of an abstract.[9]

METHODS

This was a multicentre randomised controlled trial conducted in a simulated environment at three centres. This trial adhered to the Consolidated Standards of Reporting Trials (CONSORT) guidelines for simulation-based studies.[10] The completed CONSORT checklist is provided, and the trial protocol is available on request.

Materials

The digital-twin-based simulator: All simulator-based operations are conducted using this device (Simulation AI plus, Zhejiang UE Medical Corp. Zhejiang, China), which comprises the following three core components. The digital-twin bronchoscopy simulator was developed through a multistage process integrating advanced imaging, AI and hardware engineering. First, over 40 000 anonymised chest CT sequences (1 mm slice thickness) were processed using a custom UV-Net deep learning model to achieve precise three-dimensional (3D) segmentation of pulmonary structures: five lung lobes with smooth boundaries, seventh-generation bronchial trees (including 18 segmental and 42 subsegmental bronchi), and differentiated arterial/venous vasculature. These segmented models were converted into interactive 3D meshes via Unity3D, enabling photorealistic rendering of both extraluminal anatomy and endoscopic views. A proprietary bronchoscope handle equipped with inertial and haptic sensors tracked real-time manoeuvres (insertion, rotation, tip articulation), synchronising movements with the virtual model.

Equipment: Bronchoscopy training was performed using a digital-twin-based bronchoscopy simulator, which consisted of an operating device for receiving simulated bronchoscopy operations from the user, a monitor for displaying a graphical interface for interaction with the user, and an AI-based feedback system containing four features: DC, SP, procedure time (PT) and wall contact time.

Screen labels: In the learning mode, the system automatically recognised bronchial segments and displayed them directly on the endoscopic image (figure 1).

Tracheobronchial tree diagram: After importing the CT image, the tracheobronchial tree diagram was automatically generated. The lung tree diagram informed the endoscope of its position in the bronchial tree during training.

The digital bronchial tree models were derived from high resolution CT (HRCT) of the chest performed for clinical indications in patients. HRCT was performed with a CT scanner (GE HiSpeed Advantage CT scanner, GE Medical Systems, Milwaukee, Wise) during a single breath-hold acquisition. Scanning parameters consisted of 2 mm X-ray beam collimation (slice thickness), 6 mm/s table speed and 1 mm reconstruction intervals. HRCT images were then transferred from the scanner to the digital-twin based bronchoscopy simulator. The system reconstructed multilayer views of the thorax and virtual bronchoscopy images. AI automatically identified anatomical structures, including 18 segmental bronchi, as well as blood vessels, nodules and other features.

DC score: The DC score tracked the segments that were entered. If the segment was correctly identified, the stage displayed the correct anatomical name and recorded a score of 1 point; otherwise, it turned grey.

SP score: The SP score tracked the segments that were entered and whether they had been entered according to the SP score. If entered in accordance with the SP score (ie, immediately following the correct preceding segment), 1 point was awarded; otherwise, no points were recorded for that segment. The system automatically recognised and ranked the process by which the participants identified the segments.[5]

Participants

Participants (medical students) were recruited from various local medical schools. The exclusion criteria included prior experience with clinical or simulated endoscopy. The sample size calculation focused on detecting differences between the anatomic-variety and anatomic-uniformity groups (primary comparison). Based on prior data,[5] group B (anatomic uniformity) had an SP score of 14±3.9. We hypothesised that group C (anatomic variety) would achieve near-perfect performance (SP=17.5±1.5).[11] Using a two-sample t-test ($\alpha=0.05$, power=90%), 17 participants per group were required (total n=34). To account for potential attrition and enable three-group descriptive comparisons, we inflated the sample to 20 per group (total N=60).

Randomisation and training

The study contained two parts: training and testing. Initially, all participants watched an instructional video demonstrating the fundamental operation of the bronchoscopy simulator, which was operated and recorded by the first author. A random number table was used to allocate participants at a 1:1:1 ratio across three groups: the control group, the anatomic-uniformity group and the anatomic-variety group. Participants were additionally stratified by sex, as previous studies[12] have identified sex as a factor influencing performance during the initial skill acquisition phase (figure 2).

Anatomic-variety group: After randomisation, the anatomic-variety group watched a 4 min instructional video on how to use the digital-twin-based bronchoscopy simulator. Twenty digital bronchial tree models generated from patients' CT were used for training. After each bronchoscopy, the anatomic-variety group received a report that included the PT and visual segment checklists, such as DC and SP scores generated by the AI system.

Anatomic uniformity group: Following randomisation, the anatomic-uniformity group watched the same instructional video. Trainees were trained exclusively using a standardised digital bronchial tree model within the same simulator mentioned above. The standardised digital bronchial tree model utilised in this study was identical to that reported in a previous study.[5] After each bronchoscopy, the anatomic uniformity group also received the same report.

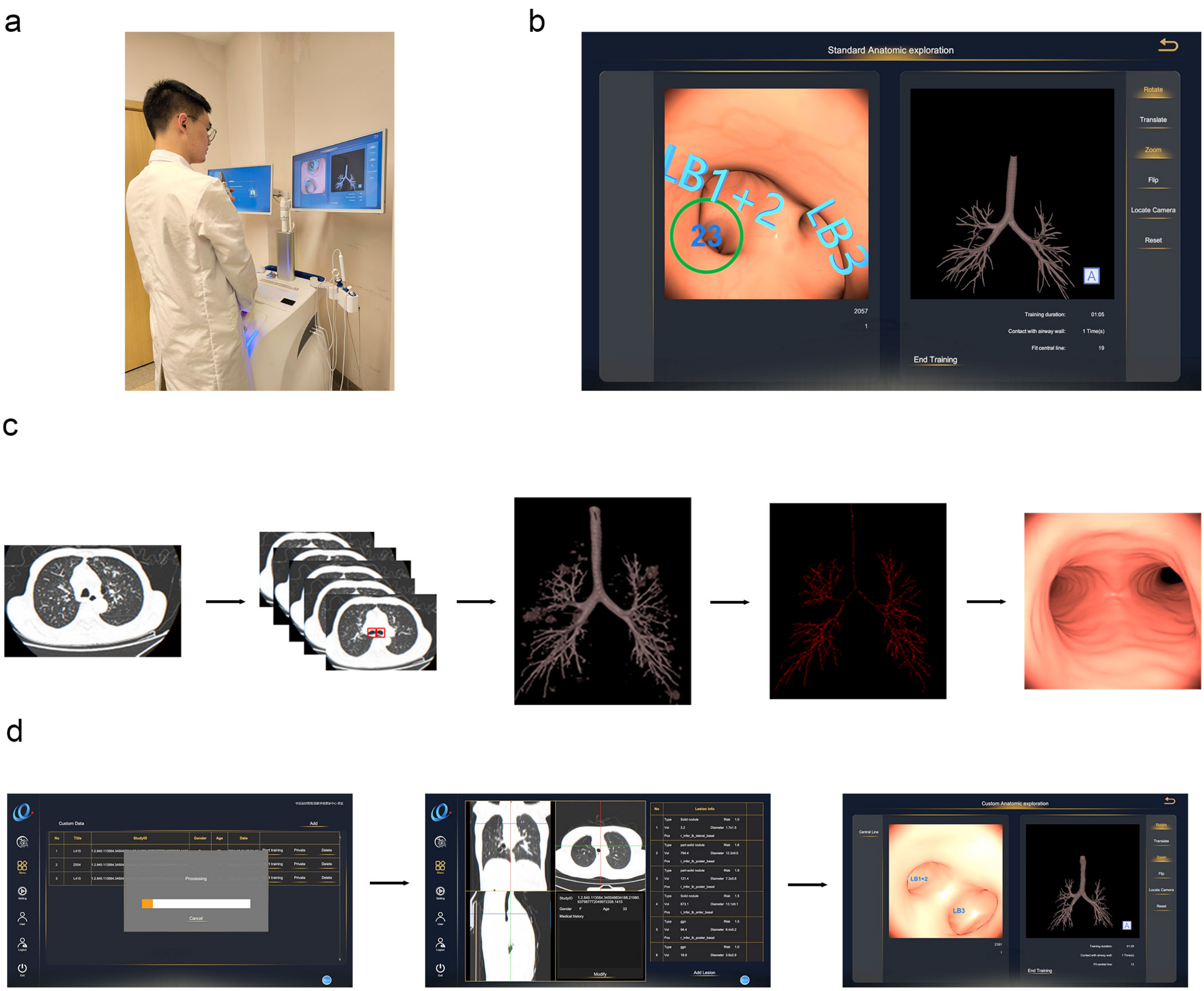

Figure 1 Simulator learning group with the studying mode activated. (a) Participant Training with the Digital-Twin Based Simulator with Artificial Intelligence Activated. (b) Onscreen tab (centre of the screen) displaying the currently visualised segment in the standard model. LB1+2/LB3. (c) Three-dimensional reconstruction of the digital bronchial tree models: 1. Image processing: The acquired CT image data underwent preprocessing, including noise reduction and enhancement operations, to improve image contrast for better visualisation of the bronchial structures. 2. Bronchial segmentation: A combination of threshold segmentation and region-growing methods was used to separate the bronchial structures from the background, forming independent bronchial structures. The segmented data were manually annotated with the anatomical names of the bronchial segments. Once sufficient data had been accumulated, a deep learning model for the automated annotation of bronchial segment names was developed. 3. Three-dimensional reconstruction: The segmentation of all cross-sectional bronchial data was completed and combined with the data layer thickness to perform overlay calculations, forming a three-dimensional (3D) mesh model of the bronchial structures. The mesh model was optimised with postprocessing steps such as detail enhancement and smoothing. 4. Centreline calculation: The vertices of the 3D mesh model were converted into point cloud data, and the central axis of the lumen was extracted based on geometric methods. The central axis data were filtered to remove abnormal branches, ensuring an approximate 2 mm spacing between each central axis point to facilitate subsequent simulation operations. 5. Simulation operations: Data from bronchoscopy handle sensors, central axis data and a 3D airway mesh model were integrated to form a physical motion system. A 3D physics engine based on collision and constraint calculations was used to determine the position and direction of the bronchoscope tip within the 3D airway model, thereby enabling the depiction of images under the bronchoscope. (d) Example of 3D reconstruction in the simulator: 1. CT import and processing. 2. Real-time patient CT display. 3. Airway reconstruction and simulation: The lung tree diagram (right panel) indicates the location of bronchoscopy exploration in the examined area. The lower right panel displays the length of training, number of wall touches and time the bronchoscope remained centred in the airway.

Control group: After randomisation, the control group was shown a 2 min instructional video that introduced them to traditional training methods based on the Four-Landmark Approach.[13] They were also given an instructional booklet explaining the training methods and a poster highlighting the methods and anatomy of the bronchial segments. The Four-Landmark Approach is an instruction on performing structured bronchoscopy that divides the bronchial tree into four landmarks and combines them with endoscopic angles.

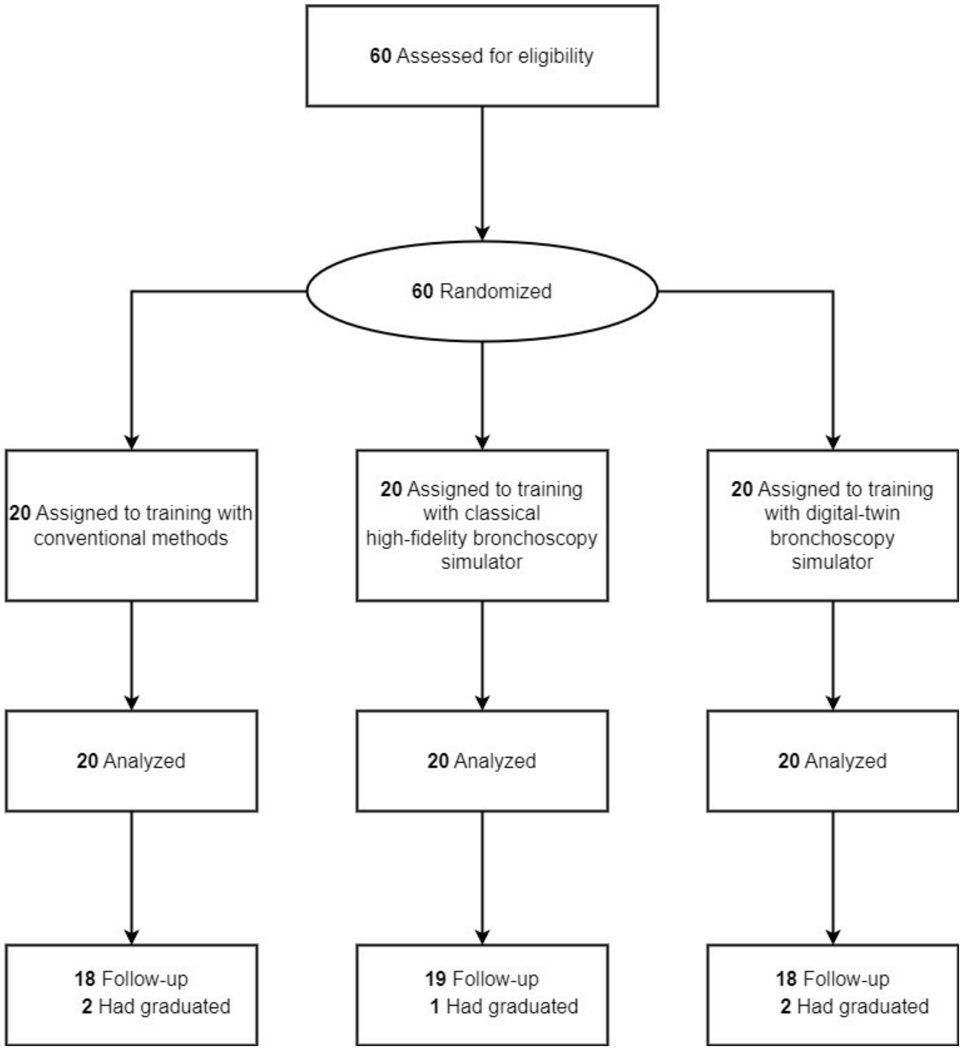

Figure 2 Participant flow diagram. A detailed diagram illustrating the progression of participants throughout the trial.

Training: Participants were allowed unrestricted training time but were constrained to a maximum of 3.5 hours to prevent fatigue during testing. Participants were permitted unrestricted training time to reflect typical clinical learning environments. While mastery learning—a gold standard requiring predefined competency benchmarks—was not employed here, this design choice allowed us to examine naturalistic skill acquisition patterns and trainee self-assessment accuracy.

Test

When the participants no longer felt that they had benefited from the additional training, they took the final test, which was a complete bronchoscopy, with no guidance or assistance. The final test was the same for all groups, with no use of a feedback tool. It involved two tests: Test 1 was a standard digital bronchial tree model (the model once used in the training of the anatomic-uniformity group), and test 2 was a digital bronchial tree model based on a brand-new chest CT. The test could only be performed once, after which the user interface was locked, and the test ended. The system automatically timed the test, starting when the bronchoscope entered the airway and ending when the bronchoscope completely exited the airway.

After the test, all groups completed the Intrinsic Motivation Inventory (IMI) questionnaire.[5] The questionnaire comprised six statements. Each statement was rated using a Likert scale from 1 to 7, with 1 indicating ‘not at all true’ and 7 indicating ‘very true’.

Skill maintenance: 3 months after training, participants repeated the same testing protocol to assess knowledge retention; this involved a standard and a brand-new digital bronchial tree model with the following measures: DC, SP and PTs.

Outcome measures

All participants were informed that they were required to achieve the highest possible score on the examination:

DC: DC was defined as probing and recognising all bronchial segments, with a total of 18 segments: 10 in the right lung and 8 in the left lung; left lungs one and two were fused, and no segment 7.[3 14]

SP: A point was awarded each time a participant moved from one segment to the next.[15] For example, exploring the right upper lobe in the following order: a sequence of trachea/right bronchial segment RB1/RB2/RB3 earned three points, while a sequence of trachea/RB2/RB1/RB3 earned 0 points.

PT: PT was defined as the time spent visualising the carina to extract the scope.[3 14] AI automatically timed the operational process.

Statistical analysis

The ceiling effect and the fact that DC and SP have a maximum score of 18 could make normal distribution infeasible. Consequently, data for DC, SP, PT and training duration are presented as median±IQR, and non-parametric Kruskal-Wallis tests were employed for group comparisons. The χ^2 test was used for categorical variables. The primary analysis compared the anatomic-variety group with the anatomic-uniformity group. All other comparisons were considered secondary analyses. For the primary comparison, the Kruskal-Wallis test was applied. To compare all three groups across DC, SP and PT, Kruskal-Wallis tests were used. When a globally significant difference was detected ($p<0.05$), Dunn's test with Bonferroni correction was used for post hoc pairwise comparisons. Effect sizes were calculated as rank eta squared (ηH^2). For comparisons between groups (ie, anatomic-variety group vs anatomic-uniformity group), effect sizes were calculated using the rank-biserial correlation (r_b). The r_b was interpreted as follows: 0.1=small, 0.3=medium, 0.5=large. Given the bounded, non-normal distribution of the DC and SP scores, a Scheirer-Ray-Hare test (rank-based two-way analysis of variance) was used to evaluate the effects of group and time (post-test vs 3-month retention) and their interaction. Between-group differences are reported as median differences and p values. Statistical significance was set at $p<0.05$.

RESULTS

The digital-twin bronchoscopy simulator of a patient environment that could reconstruct digital bronchial trees generated from diverse patient chest CT data and automatically identify bronchial segments based on AI. This simulator not only mimics the experience of bronchoscopy in a real clinical setting by providing various anatomic models but also provides immediate AI-based training feedback. Training with an extensive range of digital bronchial tree models generated from CT data of various patients closely simulates clinical practice.

In this prospective, randomised controlled trial conducted from 1 June 2024 to 28 September 2024, 60 participants were enrolled, and 20 participants were allocated to each group (table 1).

Post-test

In test 1, the primary comparison showed that there were no significant differences between the anatomic-variety and anatomic-uniformity groups in the three outcomes (median difference, p value): DC (0 segments, p=0.576), SP (1 correct progressions, p=0.091) and PT (31 s, p=0.831). Secondary analyses revealed that both groups exhibited superior training performance compared with the control group, as indicated by the three outcomes (median, p value): DC (18, 18, 13.5 segments, $p<0.001$, $\eta H^2=0.627$), SP (17, 16, 8.5 correct progressions, $p<0.001$, $\eta H^2=0.596$) and PT (475, 444, 913.5 s, $p<0.001$, $\eta H^2=0.416$).

In test 2, the primary comparison demonstrated the anatomic-variety group had significantly higher DC (median difference: 2.5 segments, $p<0.001$, $r_b=0.945$) and SP (median difference: 9 progression, $p<0.001$, $r_b=0.995$) than the anatomic-uniformity group. It is noteworthy that there was no significant difference in PT between the two groups (see table 1). Secondary analyses revealed that both the anatomic-variety and anatomic-uniformity groups demonstrated superior training performance in comparison to the control group (table 1).

Additionally, to evaluate the stability of the bronchoscopy performance for each group, we compared the outcomes in two tests. The anatomic-variety group demonstrated more stable performance in both the DC and SP scores: DC (median difference: –0.5 points, p=0.257) and SP (0 correct progressions, p=0.796). However, the anatomic-uniformity group showed an unstable performance in both DC and SP: DC (–3 points, $p<0.001$, $r_b=-0.78$) and SP (–8 correct progressions, $p<0.001$, $r_b=-1.0$). The PTs were prolonged in both the anatomic-variety (221 s, $p<0.001$, $r_b=0.488$) and anatomic-uniformity groups (346 s, $p<0.001$, $r_b=0.607$). In the control group, the performance was also unstable in both the DC and SP, but the PTs showed no significant change between the two tests.

The training time across the three groups was comparable, but the participants using the digital-twin simulator had higher IMI scores than the control group (18.5 scores, $p<0.001$, $r_b=0.539$) (table 2).

Table 1 Participants' demographic and outcome measures

	Control group (group A, N=20)	Anatomic-uniformity group (group B, N=20)	Anatomic-variety group (group C, N=20)	P value (group A vs B vs C)	P value (group B vs C)
Female sex*	10 (50%)	10 (50%)	10 (50%)	1	1
Age, years	24.3 (0.44)	24.6 (0.57)	24.0 (0.00)	0.640	0.429
Testing with the trained standard digital bronchial tree model					
DC, segments	13.5 (2.5)	18 (2)	18 (1)	< 0.001	0.576
SP, progressions	8.5 (5.5)	16 (3)	17.0 (2.5)	< 0.001	0.091
PT, s	913.5 (249.0)	444 (251.5)	475.0 (268)	< 0.001	0.831
Testing with a digital bronchial tree generated from a brand-new chest CT					
DC, segments	7.0 (4)	15 (2)	17.5 (1)	<0.001	<0.001
SP, progression, s	3.5 (3)	8 (2.5)	17 (1.50)	<0.001	<0.001
PT, s	1007.5 (473.50)	790 (331.50)	696.0 (258)	0.001	0.475
Time spent training, min	55 (29.50)	60 (35)	60 (15.50)	0.842	0.936

Data are presented as median (IQR) and were compared using the Kruskal-Wallis test.
*Data are presented as numbers (percentage) and were compared using the χ^2 test. Statistical significance was set at $p<0.05$.
DC, diagnostic completeness; PT, procedure time; SP, structured progress.

Table 2 Intrinsic Motivation Inventory (IMI)

	Control group (group A, N=20)	Anatomic-uniformity group (group B, N=20)	Anatomic-variety group (group C, N=20)	P value (group A vs B vs C)
I put a lot of effort into this	5 (2)	5 (1)	5.5 (1)	0.031
I think I did pretty well at the final test, compared with the other students	3 (2)	4 (1.5)	5 (2)	0.001
I felt pressured while training*	3.5 (5)	3 (2)	3 (2)	0.793
I think this training session is important to do because it can help me to perform better bronchoscopies	2 (1.5)	7 (1)	7 (1)	<0.001
I would recommend others to train their skills with this system	1 (1)	6.5 (1)	7 (1)	<0.001
I would like to continue to use this training system	1.5 (2)	7 (1)	7 (1)	<0.001
IMI total (total=35)	12(16)	28.5 (4)	30.50(4)	<0.001

Data are presented as median (IQR) and were compared using the Kruskal-Wallis test.
Each statement was rated using a Likert scale from 1 to 7, where one indicated 'not at all true' and seven indicated 'very true'. Number 7 indicates the best score, except for a, because it is a reverse-coded statement in which 1 indicates the best score and 7 the worst score.
*This was not included in the total IMI score.

Skill maintenance

55 of 60 (91.67%) participants returned for the bronchoscopy skill retention test. In test 1, the anatomic-variety group showed better performance maintenance than the other groups in the two outcome measures (median, p value: DC: 17, 17, 12 segments, $p<0.001$, $\eta H^2=0.462$; SP: 15, 7, 6 correct progressions, $p<0.001$, $\eta H^2=0.528$). Among the three groups, there was a trend toward a decrease in test 1 over time. Although a statistically significant difference was seen in SP scores among the groups (anatomic-variety group, $p<0.01$, $r_b=-0.556$; anatomic-uniformity group, $p<0.001$, $r_b=-0.895$; control group, $p<0.05$, $r_b=-0.406$), neither the anatomic-variety group nor the anatomic-uniformity group had significant changes in DC scores. In Test 2, the anatomic-variety group still had the best performance among the groups (DC, 17, 14, 6 segments, $p<0.001$, $\eta H^2=0.365$; SP, 13, 4, 3 correct progressions, $p<0.001$, $\eta H^2=0.738$). The three groups showed no differences in the PT for the retention test (table 3). Both the anatomic-variety and anatomic-uniformity groups showed a statistically significant difference (DC: anatomic-variety group, $p=0.008$, $r_b=-0.364$; anatomic-uniformity group, $p=0.010$, $r_b=-0.645$; control group, $p=0.068$; SP: anatomic-variety group, $p=0.001$, $r_b=-0.593$; anatomic-uniformity group, $p=0.007$, $r_b=-0.548$; control group, $p=0.060$) over time.

To investigate whether the learning gains remained stable over time, a non-parametric repeated measures Scheirer-Ray-Hare test was applied to evaluate within-group changes over time (post-test vs 3-month retention) and between-group differences (table 4). A main effect of group was seen for all measures. Additionally, a main effect of the test was found for 'SP' in test 1 and 'DC and SP' in test 2, indicating that the performance score for these measures deteriorated from post-test to retention test. However, no interaction was found between the three groups for these measures, so the deterioration in performance was found in all three groups.

DISCUSSION

To our knowledge, this is the first study to use a digital-twin bronchoscopy simulator to train and test novice bronchoscopy performance. Our results indicate that the 'anatomic-variety' training approach promoted better, faster and more stable bronchoscopy performance for novices, regardless of whether the standard or brand-new digital bronchial tree models are used. Importantly, the anatomic-variety group demonstrated partial retention of learning gains when reassessed at 3 months, as evidenced by DC and SP scores, as well as PT. However, these gains declined in the absence of repeated exposure to simulation.

Previous studies[3 16 17] have demonstrated that simulation-based training (based on virtual reality simulators or a standard phantom of the bronchial tree) combined with feedback can improve the DC and SP scores of novices, which is consistent with our study findings. In our study, the participants who used the digital-twin-based bronchoscopy simulator with AI

Table 3 Outcome measures of skill maintenance

	Control group (group A, N=20)	Anatomic-uniformity group (group B, N=20)	Anatomic-variety group (group C, N=20)	P value (group A vs B vs C)	P value (group B vs C)
Testing with the trained standard digital bronchial tree model					
DC, segments	12(3)	17±3	17(1)	<0.001	0.214
SP, progressions	6 (2)	7±4	15(6)	<0.001	<0.001
PT, s	731(203)	632(248)	604(227)	0.056	0.504
Testing with a digital bronchial tree generated from a brand-new chest CT					
DC, segments	6 (2)	14(2)	17(3)	<0.001	<0.001
SP, progressions	3 (4)	4 (4)	13(5)	<0.001	<0.001
PT, s	900 (322)	877 (375)	741 (340)	0.326	0.261

Data are presented as median (IQR) and were compared using the Kruskal-Wallis test.
Statistical significance was set at p<0.05.
DC, diagnostic completeness; PT, procedure time; SP, structured progress.

Table 4 Statistical data from repeated measures Scheirer-Ray-Hare test on the measures from the post-test and the retention test

	Test group interaction	Group main effect	Test main effect
Testing with the trained standard digital bronchial tree model			
DC, segments	H=0.233, p=0.89	H=60.62, p<0.001*	H=3.499, p=0.061
SP, progressions	H=4.30, p=0.12	H=51.99, p<0.001*	H=20.64, p<0.001*
PT, s	H=7.47, p=0.02*	H=32.36, p<0.001*	H=0.86, p=0.350
Testing with a digital bronchial tree generated from a brand-new chest CT			
DC, segments	H=0.32, p=0.85	H=84.05, p<0.001*	H=4.92, p=0.030*
SP, progressions	H=0.24, p=0.89	H=69.62, p<0.001*	H=7.56, p=0.010*
PT, s	H=3.04, p=0.22	H=13.34, p=0.001*	H=0.72, p=0.400

*Statistical significance was set at p<0.05.
DC, diagnostic completeness; PT, procedure time; SP, structured progress.

feedback performed significantly better in terms of DC, SP and PT compared with the control group during standard digital bronchial tree model tests. In contrast to previous studies,[5] we found that novices trained using a standard digital bronchial tree model, such as the classical high-fidelity simulator, showed a reduction in bronchoscopy performance when confronted with another new digital bronchial tree model test. These results suggest that the training effectiveness of the classical simulator is overestimated because it may not lead to stability in bronchoscopy performance. Schmidt and Bjork[18] demonstrated that variable practice—exposing learners to diverse task versions—enhances generalisation to novel scenarios, even at the cost of slower initial learning. Our digital-twin simulator operationalises this principle by reconstructing anatomically diverse bronchial trees from patient CT data, thereby bridging the gap between simulated and clinical environments. By replicating the core design of the earlier study by Cold *et al*[5] we ensured methodological consistency, allowing robust evaluation of how anatomical diversity in training affects skill acquisition. This approach underscores the translational potential of digital-twin technology while validating prior findings in a new context.

A plateau in the learning curve of bronchoscopy can be reached with at least ten different patients,[19 20] but the same bronchial model trained ten times may not achieve the same results. Classical bronchoscopy simulators typically offer a limited number of cases for training and testing,[7 21] which may be a key factor contributing to the poor performance of participants when tested using the new bronchial tree model after training with traditional simulators. The adaptive expertise approach emphasises deep conceptual understanding by integrating multiple concepts necessary for expertise, the discovery of new solutions through struggle and failure, and the ability to train in a variety of contexts.[19 20] However, our digital twin-based bronchoscopy simulator is pretrained on an extensive CT dataset, allowing it to automatically and accurately reconstruct lung structures, including anatomical variants and lesions, facilitating the generation of a comprehensive bronchial tree model based on patient data. This simulator offers diverse scenarios to replicate realistic clinical environments, such as various anatomical variations and the presence of tumours. Although our study did not directly measure patient-level outcomes, prior systematic reviews suggest that improved procedural adaptability correlates with reduced complications (eg, fewer mucosal injuries during bronchoscopy). Future work should evaluate whether anatomic-variety training translates to measurable clinical benefits, such as shorter PTs or lower rates of diagnostic errors.

Additionally, the simulator-based training group achieved significantly higher scores on the IMI compared with the control group in terms of confidence and overall experience. These findings suggest that the instructional method enhances learners' confidence in performing bronchoscopy and improves their overall learning experience and performance. The goal of learning is not merely to acquire procedural skills but also to work effectively in dynamic and ever-changing environments, which is a defining characteristic of medical practice.[19 22 23] Therefore, approaches that emphasise ability (rather than mere performance), as well as strategies focused on preparing for lifelong learning and developing adaptive expertise, are crucial.

Skill maintenance is crucial for basic clinical skills, particularly during short residency rotations and continuing medical education.[24 25] Nonetheless, high-quality research on this topic remains limited. Previous studies have demonstrated that both low-fidelity 3D-printed airway models and classical high-fidelity bronchoscopy simulators significantly enhance students' bronchoscopy performance and help sustain learning gains when integrated into the curriculum.[26 27] In our study, the digital-twin bronchoscopy simulator also demonstrated superior training outcomes, both immediately following the training and during the retention test conducted 3 months later. The enhanced skill retention and adaptability observed in the anatomic-variety group resonate with Schmidt and Bjork's[18] concept of contextual interference. By training on diverse bronchial models, novices faced increased cognitive demands during practice, requiring them to continuously adapt navigation strategies. This 'desirable difficulty'—though initially slowing skill acquisition—strengthened long-term retention and transferability. Moreover, using the anatomic-uniformity group as a reference, the digital-twin-based bronchoscopy simulator also had better performance in the brand-new digital bronchial tree model test. However, the retention test performance with the digital model revealed a general decline in ability compared with the postsimulation scores; this observation is in line with similar previous reports.[26 27] The implication of these results is that this digital-twin-based bronchoscopy simulator should be considered for trainees who experience a significant gap in exposure to bronchoscopy.

This study has some limitations. First, the accessibility of the equipment is a concern given that it is currently not widely available. This new technology requires continuous improvement to enhance its feasibility for widespread adoption. Second, we reported high retention scores for bronchoscopy performance in the digital-twin simulator group, but the retention rates of technical skills remained unclear. In the future, we will explore how frequently or what type of content training must occur to improve the knowledge retention rate. We will also develop a structured training programme. Third, the impact of training on future clinical practice performance is particularly significant.[20 25] Our results suggest that this novel simulator could improve novice bronchoscopy performance (navigation through the bronchial tree and lung segment recognition ability); however, its ability to shorten the learning curve of novices remains unknown. Therefore, this should be explored further in the future. This study focused on simulated performance (Kirkpatrick level 2), but real-world translation requires validation through longitudinal clinical trials.[28] For example, tracking trainees' patient outcomes (level 4) could quantify the simulator's impact on diagnostic accuracy or complication rates. Fourth, while replicating the core design of Cold *et al*[5] strengthened the comparative validity for assessing anatomical diversity, this approach inherently limits the exploration of other potentially influential variables unique to our simulator or training paradigm that might differ from the referenced AI system. Finally, our self-regulated training design, while ecologically valid, introduces variability in skill attainment compared with mastery learning. Systematic reviews[8 29] emphasise that mastery learning—with its emphasis on deliberate practice and objective benchmarks—reduces performance variance and ensures baseline competency.[11] In contrast, our anatomic-variety group's superior retention despite self-paced training suggests that anatomical diversity may partially compensate for unstructured practice, a hypothesis requiring further validation. Future studies should integrate mastery criteria with variable anatomical training to optimise both consistency and adaptability.

CONCLUSIONS

Training with a digital twin simulator enables novices to achieve more structured bronchoscopy performance, as evidenced by superior SP scores.

Author affiliations

[1]National Center for Respiratory Medicine, State Key Laboratory of Respiratory Health and Multimorbidity, National Clinical Research Center for Respiratory Diseases, Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Department of Pulmonary and Critical Care Medicine, China-Japan Friendship Hospital, Beijing, China
[2]Department of Pulmonary and Critical Care Medicine, Affiliated Hospital of Jianghan University, Wuhan, Hubei, China
[3]Department of Interventional Pulmonology and Endoscopic Diagnosis and Treatment Center, Anhui Chest Hospital, Hefei, Anhui, China
[4]Department of Pulmonary and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[5]Department of Respiratory and Critical Care Medicine, Affiliated Beijing Chaoyang Hospital of Capital Medical University, Beijing, China
[6]State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China
[7]Department of Respiratory and Critical Medicine, First Medical Center, Chinese PLA General Hospital, Beijing, China
[8]China-Japan Friendship Hospital (Institute of Clinical Medical Sciences), Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
[9]Department of Pulmonary and Critical Care Medicine 2, Emergency General Hospital, Beijing, China
[10]Department of Respiratory and Critical Care Medicine, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[11]State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China
[12]Department of Pneumology and Critical Care Medicine, Thoraxklinik, University of Heidelberg and Translational Lung Research Center, Heidelberg, Germany

Contributors FJFH and GH conceived the manuscript and were responsible for conceptualisation and study design. MD, FL, FT, C-LT, RT, WC and ZY conducted the database search and data extraction. MD, WX and WC performed the statistical analysis. MD, FL, FT, C-LT and YX contributed to the manuscript writing. ZY, FW, NZ, SL and FJFH contributed to study evaluation. GH is the guarantor of the article, taking responsibility for the integrity of the work as a whole. All authors contributed to the article, revised it and approved the submitted version.

Funding This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0528900 to G Hou, 2024ZD0528902 to G Hou), CAMS Innovation Fund for Medical Sciences (CIFMS, Grant No. 2024-I2M-ZH-022, 2022-I2M-1-025) and Excellence & Innovation Initiative of China-Japan Friendship Hospital (NO.ZRZC2025-XYA01).

Competing interests GH and MD are the original inventors of the reported bronchoscopy simulator system. This system provides simulated bronchoscopy procedures and airway reconstruction.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was conducted from June to July 2024, approved by the ethics committees of all participating centres (China-Japan Friendship Hospital: 2024-KY-095) and written informed consent was obtained from all patients for the designated CT data. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs

Feng Wang https://orcid.org/0000-0002-2429-5662
Yang Xia https://orcid.org/0000-0003-2487-2244
Gang Hou https://orcid.org/0000-0003-3438-1764

REFERENCES

1 Stather DR, MacEachern P, Chee A, *et al*. Trainee impact on procedural complications: an analysis of 967 consecutive flexible bronchoscopy procedures in an interventional pulmonology practice. *Respiration* 2013;85:422–8.
2 Ouellette DR. The safety of bronchoscopy in a pulmonary fellowship program. *Chest* 2006;130:1185–90.
3 Colt HG, Crawford SW, Galbraith O. Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest* 2001;120:1333–9.
4 New ML, Huie TJ, Claar D, *et al*. Virtual Reality Anatomy Trainer Turns Teaching Endobronchial Ultrasound Inside-Out. *Chest* 2025;167:1440–50.
5 Cold KM, Xie S, Nielsen AO, *et al*. Artificial Intelligence Improves Novices' Bronchoscopy Performance: A Randomized Controlled Trial in a Simulated Setting. *Chest* 2024;165:405–13.
6 Blum MG, Powers TW, Sundaresan S. Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Ann Thorac Surg* 2004;78:287–91.
7 Ost D, DeRosiers A, Britt EJ, *et al*. Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med* 2001;164:2248–55.
8 Gerretsen ECF, Chen A, Annema JT, *et al*. Effectiveness of Flexible Bronchoscopy Simulation-Based Training: A Systematic Review. *Chest* 2023;164:952–62.
9 Xu W, Hou G, Deng M. RCT abstract - a novel artificial intelligence feedback system boosts novice bronchoscopy performance based chest-ct simulated tracheal models: a randomized controlled trial. ERS Congress 2024 abstracts; September 14, 2024:RCT998.
10 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;11:238–48.
11 Cold KM, Wei W, Agbontaen K, *et al*. Mastery Learning Guided by Artificial Intelligence Is Superior to Directed Self-Regulated Learning in Flexible Bronchoscopy Training: An RCT. *Respiration* 2025;104:206–15.
12 Ali A, Subhi Y, Ringsted C, *et al*. Gender differences in the acquisition of surgical skills: a systematic review. *Surg Endosc* 2015;29:3065–73.
13 Cold KM, Vamadevan A, Nielsen AO, *et al*. Systematic Bronchoscopy: the Four Landmarks Approach. *J Vis Exp* 2023;2023:196.

14 Konge L, Arendrup H, von Buchwald C, *et al*. Using performance in multiple simulated scenarios to assess bronchoscopy skills. *Respiration* 2011;81:483–90.
15 Cold KM, Svendsen MBS, Bodtger U, *et al*. Using structured progress to measure competence in flexible bronchoscopy. *J Thorac Dis* 2020;12:6797–805.
16 Gopal M, Skobodzinski AA, Sterbling HM, *et al*. Bronchoscopy Simulation Training as a Tool in Medical School Education. *Ann Thorac Surg* 2018;106:280–6.
17 Siow WT, Tan G-L, Loo C-M, *et al*. Impact of structured curriculum with simulation on bronchoscopy. *Respirology* 2021;26:597–603.
18 Schmidt RA, Bjork RA. New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychol Sci* 1992;3:207–18.
19 Voduc N, Adamson R, Kashgari A, *et al*. Development of Learning Curves for Bronchoscopy: Results of a Multicenter Study of Pulmonary Trainees. *Chest* 2020;158:2485–92.
20 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;95:1921–8.
21 Colt HG, Williamson JP. Training in interventional pulmonology: What we have learned and a way forward. *Respirology* 2020;25:997–1007.
22 Brady AK, Town JA, Robins L, *et al*. Bronchoscopy Teaching Without a Gold Standard: Attending Pulmonologists' Assessment of Learners, Supervisory Styles, and Variation in Practice. *Chest* 2021;160:1799–807.
23 Cold KM, Svendsen MBS, Bodtger U, *et al*. Automatic and Objective Assessment of Motor Skills Performance in Flexible Bronchoscopy. *Respiration* 2021;100:347–55.
24 Routt E, Mansouri Y, de Moll EH, *et al*. Teaching the Simple Suture to Medical Students for Long-term Retention of Skill. *JAMA Dermatol* 2015;151:761–5.
25 Ernst A, Wahidi MM, Read CA, *et al*. Adult Bronchoscopy Training: Current State and Suggestions for the Future: CHEST Expert Panel Report. *Chest* 2015;148:321–32.
26 Bjerrum AS, Hilberg O, van Gog T, *et al*. Effects of modelling examples in complex procedural skills training: a randomised study. *Med Educ* 2013;47:888–98.
27 Feng DB, Yong YH, Byrnes T, *et al*. Learning Gain and Skill Retention Following Unstructured Bronchoscopy Simulation in a Low-fidelity Airway Model. *J Bronchology Interv Pulmonol* 2020;27:280–5.
28 Clausen AF, Sperling S, Jensen RD, *et al*. In-situ simulation-based team training reduces incidence of negative events during bronchoscopy. A prospective educational intervention cohort study. *Respir Res* 2025;26:133.
29 Kennedy CC, Maldonado F, Cook DA. Simulation-based bronchoscopy training: systematic review and meta-analysis. *Chest* 2013;144:183–92.