



ANNALS

OFFICIAL JOURNAL OF THE ACADEMY OF MEDICINE, SINGAPORE

VOLUME 55

NUMBER 7

JULY 2026



Artificial intelligence tools in dermatology education: A scoping review on their application, efficacy, and limitations

This scoping review charts the evolving role of artificial intelligence from the classroom to the clinic, highlighting priorities for safe and ethical integration. (See full article, p. 370)

Illustration by Maria De-Castro

Also in this issue

A multicomponent control programme in nursing homes in Singapore

Burden trends of Alzheimer's disease and other dementias in China, 1990–2023: Sex, age, and risk-attributable patterns

Interpreting multivariable regression coefficients in observational clinical research: The Table 2 fallacy

Editor-in-Chief

Raymond Seet

Deputy Editors

Deidre Anne De Silva

Beng Yeong Ng

Associate Editors

Brian Goh

Li Yang Hsu

Board Members

Ravindran Kanesvaran

Felix Keng

Mariko Koh

Alfred Kow

Jan Hau Lee

Joseph Ng

Andrew Ong

Dujeepa Samarasekera

John Soong Tshon Yit

Mythily Subramaniam

Clement Tan

Tjun Yip Tang

Editorial Fellows

Dedrick Chan Kok Hong

Andrew Ho Fu Wah

Emeritus Editors

Vernon MS Oh

Eng King Tan

Immediate Past Editor

Erle Lim

Manager

Wen Shan Leong

Senior Editorial Executives

Nuraiziah Johari

Diane Mendez Pulvera

OPEN ACCESS

Annals is an open access journal, where our articles may be used according to the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License (CC BY-NC-SA 4.0). You may share, distribute, remix, tweak and build upon the work non-commercially, while ensuring appropriate credit is given and that new creations are licensed under the identical terms.

DISCLAIMER

All articles published, including editorials, original articles, commentaries, letters and reviews, represent the opinion of the authors and do not necessarily reflect the official policy of the Academy of Medicine, Singapore.

ISSN 2972-4066

Annals, Academy of Medicine, Singapore

Volume 55 | Number 7 | July 2026

EDITORIAL

Strengthening infection control in Singapore's nursing homes: Lessons, limitations, and the road ahead

Meng Ying Sim, Ray Junhao Lin 347

ORIGINAL ARTICLES

A multicomponent control programme in nursing homes in Singapore

Louisa Sun, Nares Smitasin, Brenda Mae Alferez Salada, Cathrine Teo, Garvi Pandya, Xin Hu, Dale Andrew Fisher.....350

Burden trends of Alzheimer's disease and other dementias in China, 1990–2023: Sex, age, and risk-attributable patterns

Yao Li, Shuai Wang, Kaiyue Xie, Siyv Qian, Mingyue Ma, Haining Zhang.....358

REVIEW ARTICLE

Artificial intelligence tools in dermatology education: A scoping review on their application, efficacy, and limitations

Chun Hian John Yap, Ziying Vanessa Lim 370

COMMENTARY

Interpreting multivariable regression coefficients in observational clinical research: The Table 2 fallacy

John Rong Hao Tay, Nasir Z Bashir, Upul Cooray 383

LETTERS TO THE EDITOR

Transcatheter aortic valve implantation for severe aortic regurgitation with Trilogy system: Initial Southeast Asia experience

Jonathan Yap, Samuel Ji Quan Koh, Zameer Abdul Aziz, Sivaraj Govindasamy, Victor Tar Toong Chao, Mohammed Rizwan Amanullah, See Hooi Ewe, Kay Woon Ho 388

Real-world drug utilisation of nintedanib and pirfenidone for interstitial lung disease in Singapore

Tingfeng Lee, Yi Jing Lim, Andrew Yunkai Li, Ziqin Ng, Zhikui Chrystal Huang, Gin Tsen Chai.....392

A primary care perspective on gender-focused mental health services in Singapore: Correspondence

Jeremiah Pereira.....396

"Gazing into the crystal ball of mortality prediction in conservative kidney care": Correspondence

Norhasyimah Tamin, Fazean Idris, Shin Bey Ho, Vui Heng Chong, Jackson Tan.....398

Strengthening infection control in Singapore's nursing homes: Lessons, limitations, and the road ahead

Meng Ying Sim¹ MRCP (UK), Ray Junhao Lin¹ MMed (Int Med)

In this issue of the *Annals*, Sun et al. contribute to a growing body of evidence supporting preventive and systems-based interventions in nursing homes (NHs).¹ As populations age and healthcare systems shift towards community-based and preventive care models, NHs have become an increasingly important component of healthcare delivery. In Singapore, where the proportion of residents aged 65 years and above is projected to rise from approximately 16% to one-quarter of the population by 2030, this trend carries particular urgency. Strengthening care in these settings has implications not only for patient outcomes but also for health system sustainability, particularly through reducing healthcare-associated infections, avoidable hospital admissions, and unnecessary utilisation of acute hospital resources.

Existing literature has consistently shown that proactive interventions, including standardised care processes and coordinated service delivery between NHs and acute care systems, can reduce avoidable transfers to hospitals. In a systematic review by Graverholt et al., interventions such as advance care planning, enhanced access to specialist services, and structured clinical pathways for lower respiratory tract infections were associated with reduced hospital admissions.² In Singapore, similar principles have guided efforts to strengthen integration between acute hospitals and long-term care facilities. The Empower Nursing Homes And improve staNdsards of CarE (ENHANCE) programme by Sengkang General Hospital in Singapore demonstrate how structured partnerships can improve preventive care delivery in the community.³ ENHANCE incorporated tools including the National Early Warning Score, teleconsultation, clinical care pathways, and staff training aimed at improving NH staff capacity to assess and manage acutely unwell residents on site, thereby reducing hospital transfers.

Sun et al. extend previous work through emphasis on infection prevention and control (IPC) as a core strategy for reducing infectious disease burden and avoidable hospital transfers in NHs. Through improved access to specialist support, structured IPC measures, and standardised clinical practices, the Infectious

Diseases Community Programme (IDCP) was associated with improved NH IPC capacity and reductions in methicillin-resistant *Staphylococcus aureus* (MRSA) acquisition and inappropriate fever-related transfers. These findings are consistent with international evidence demonstrating that effective IPC programmes in long-term care facilities require multicomponent approaches incorporating staff education, surveillance systems, hand hygiene, environmental cleaning, antimicrobial stewardship, and strong organisational support.⁴ This is important in NHs, where communal living environments and frequent healthcare encounters as well as limitations in staffing, infrastructure, and IPC expertise increase the risk of infectious disease transmission and inconsistent implementation of IPC practices.

The reported 33.6% relative reduction in new MRSA acquisitions ($P < 0.001$) is both statistically significant and clinically meaningful given the burden of multidrug-resistant organisms (MDROs) in long-term care settings and the frequent movement of residents between hospitals and NHs, which influences transmission dynamics across the wider healthcare system.⁵ Underpinning these outcomes was a structured, 3-tiered training programme in which 1260 staff across all 20 participating NHs were trained across 37 sessions, incorporating a train-the-trainer component designed to sustain IPC capacity beyond the programme period. Notably, influenza vaccination coverage exceeded 90% of residents, with 7 of 9 NHs achieving at least 80% coverage among both residents and healthcare workers. This represents a substantial improvement from baseline and reflects the programme's concurrent work to address consent barriers, knowledge gaps, and cost-related obstacles to vaccine uptake.

Of note, 2 facilities, labelled NH E and NH F, demonstrated post-implementation increases in MRSA acquisition rates. The authors attribute these findings to limited baseline data, later programme onboarding, smaller sample sizes, and unmeasured confounders including differing resident case-mix and staffing variability. These discordance highlight that IPC programme effectiveness is not uniform across facilities,

The *Annals* is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Department of Infectious Diseases, Woodlands Hospital, Singapore

Correspondence: Dr Meng Ying Sim/Dr Ray Junhao Lin, Department of Infectious Diseases, Woodlands Hospital, 17 Woodlands Drive 17, Singapore 737628. Email: sim.meng.ying@nhghealth.com.sg/ray.lin@nhghealth.com.sg

Accepted: 3 June 2026

Published Online First: 9 June 2026

and early, comprehensive programme integration is key to achieving consistent outcomes. Identifying and addressing such heterogeneity will be essential as similar programmes are considered for broader implementation.

Beyond individual IPC programme components, the findings align with growing evidence supporting horizontal IPC strategies, which aim to reduce transmission broadly across multiple organisms rather than targeting a single pathogen.^{4,6} Such strategies include hand hygiene, environmental cleaning, staff education, antimicrobial stewardship, surveillance, and standardised barrier precautions—all of which feature as components of the IDCP. These approaches are relevant in NHs, where residents have increased infection risks and MDRO exposures due to multiple comorbidities, frequent healthcare exposure, and prolonged stays. The influential cluster-randomised PROTECT trial by Huang et al. demonstrated that a multicomponent IPC intervention combining targeted decolonisation strategies, chlorhexidine bathing, enhanced barrier precautions, and staff education significantly reduced MDRO prevalence and hospitalisations in NHs.⁶ While decolonisation represents a targeted vertical IPC strategy rather than a strictly horizontal approach, the PROTECT trial reinforces the complementary value of bundling universal precautions with staff education—an approach that mirrors the logic of the IDCP.

In Singapore, Healthier SG reflects a broader shift in national strategy towards preventive care, integrated population health, and chronic disease management.⁷ However, while Healthier SG has focused primarily on the above, long-term care facilities have received comparatively less attention despite accounting for substantial healthcare resource utilisation.⁸ NHs increasingly care for residents with greater clinical complexity and higher-acuity needs, making IPC capacity in these settings critical to broader health system resilience. Yet, sustaining IPC programmes at NHs require long-term institutional investment in workforce development, surveillance systems, institutional infrastructure, and IPC governance—strategies outlined by the World Health Organization's Global Report on Infection Prevention and Control.⁹ Notwithstanding the challenges of competing operational demands, the role of NHs within the healthcare system must therefore be more explicitly recognised. Central to this is the Agency for Integrated Care (AIC), which plays a pivotal role in overseeing quality standards, funding, and workforce development in Singapore's NH sector. Given that most of Singapore's NHs are operated by voluntary welfare organisations—which face distinct staffing constraints and funding

structures compared to hospital-linked facilities—any strategy for sustaining and scaling programmes such as the IDCP must account for these heterogeneous operating environments. Embedding IPC outcomes as measurable quality indicators within AIC's accreditation and oversight framework would represent a meaningful and sustainable institutional step forward.

The IDCP also incorporated an antimicrobial stewardship (AMS) component, through which an antimicrobial-choice guideline was developed collaboratively with primary care physicians, infectious disease specialists, and pharmacists. This aligns with Singapore's National Action Plan on Antimicrobial Resistance and reflects growing international evidence demonstrating that AMS in long-term care settings can be feasible and clinically beneficial.¹⁰ The programme's Fever and Desaturation Pathway additionally incorporated assessment of Advance Care Plans (ACPs) and Preferred Plans of Care in hospital transfer decisions—a dimension that intersects with Singapore's National Strategy for Palliative Care and highlights the importance of expanding ACP penetration in NHs as a complementary strategy to reduce inappropriate transfers.¹¹

Beyond clinical interventions, sustainable IPC programmes depend on robust surveillance and data systems. The study demonstrates how routinely collected healthcare data can be leveraged to support IPC monitoring and evaluation, including detecting trends and assessing intervention outcomes. By using data already captured through electronic medical records and national reporting systems, the study illustrates how existing data can generate meaningful process and outcome indicators for quality improvement. Looking ahead, integration of NH IPC surveillance data with Singapore's National Electronic Health Record, could enable more timely outbreak detection and real-time monitoring of IPC quality indicators at a system level.

The study highlights several important directions for future research. Formal health economic evaluation is needed to assess the cost-effectiveness and long-term financial sustainability of the IDCP model, which is essential for sustainable programme adoption in resource-constrained settings. Research employing implementation science frameworks—such as stepped-wedge cluster-randomised designs—could seek to identify which components of the multicomponent programme contribute most to improved outcomes, enabling resource-limited facilities to prioritise effectively. Cross-cluster evaluation across Singapore's 3 regional healthcare clusters would assess the generalisability and scalability of the hospital-NH partnership model. Finally, studies examining the

impact of staff turnover on IPC programme sustainability, and evaluating interventions, such as digital microlearning modules and cascade training approaches, are essential to ensuring durable implementation beyond the programme period.

Overall, the study by Sun et al. provides timely evidence supporting multicomponent IPC interventions in NHs. The findings underscore the value of preventive and collaborative approaches to infection control in long-term care facilities and highlight the central role NHs play within integrated healthcare systems in ageing societies. As Singapore navigates the epidemiological and fiscal challenges of a rapidly ageing population, the IDCP model offers a pragmatic and scalable framework—one that warrants broader national attention, formal health economic evaluation, and integration within Singapore's broader health policy architecture.

Ethics statement

Not applicable, as no study participants were recruited.

Declaration

The authors declare they have no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Keywords: antimicrobial stewardship, infection prevention and control, infectious diseases, nursing home, public health

REFERENCES

1. Sun L, Smitasin N, Salada BMA, et al. A multicomponent control programme in nursing homes in Singapore. *Ann Acad Med Singap* 2026;55:OnlineFirst.
2. Graverholt B, Forsetlund L, Jamtvedt G. Reducing hospital admissions from nursing homes: a systematic review. *BMC Health Serv Res* 2014;14:36.
3. Liew YH, Yang Y, Lim SXY, et al. Enhancing care in nursing homes: Qualitative insights from the ENHANCE programme. *Ann Acad Med Singap* 2024;53:758-61.
4. Smith PW, Bennett G, Bradley S, et al. SHEA/APIC guideline: infection prevention and control in the long-term care facility. *Am J Infect Control* 2008;36:504-35.
5. Chow A, Lim VW, Khan AS, et al. MRSA transmission dynamics among interconnected acute, intermediate-term, and long-term healthcare facilities in Singapore. *Clin Infect Dis* 2017;64:S76-81.
6. Huang SS, Septimus E, Kleinman K, et al. Targeted infection prevention in nursing homes and reductions in multidrug-resistant organisms and hospitalisations. *N Engl J Med* 2023;388:1403-14.
7. Ministry of Health, Singapore. Healthier SG White Paper. September 2022. <https://www.healthiersg.gov.sg/resources/white-paper>. Accessed 5 June 2026.
8. Quah WC, Leong CJ, Chong E, et al. Unplanned hospitalisations among subsidised nursing home residents in Singapore: insights from a data linkage study. *Ann Acad Med Singap* 2024;53:657-69.
9. World Health Organization. Global Report on Infection Prevention and Control. 23 May 2022. <https://www.who.int/publications/i/item/9789240051164>. Accessed 5 June 2026.
10. Ministry of Health, Singapore. National Action Plan on Antimicrobial Resistance 2023–2027. https://isomer-user-content.by.gov.sg/18/47d5e456-b123-45e6-a744-8afc2d98d5af/NSAPv2_Final_12Nov2025_for%20publication.pdf. Accessed 5 June 2026.
11. Ministry of Health, Singapore. 2023 National Strategy for Palliative Care. 1 July 2023. <https://www.moh.gov.sg/others/resources-and-statistics/reports-2023-national-strategy-for-palliative-care-report>. Accessed 5 June 2026.

A multicomponent control programme in nursing homes in Singapore

Louisa Sun^{1,3} MBBS, Nares Smitasin^{1,2} MD, Brenda Mae Alferez Salada¹ MD, Cathrine Teo¹ BSc, Garvi Pandya⁴ MBBS, Xin Hu¹ CIC, Dale Andrew Fisher^{1,2} MBBS

ABSTRACT

Introduction: Nursing home (NH) residents are vulnerable to infections and avoidable hospital transfers. The Infectious Diseases Community Program was developed as a multicomponent intervention to strengthen infection prevention and control (IPC), improve vaccination uptake, and support clinical decision-making in NHs.

Methods: The authors conducted a retrospective cohort study in 6 NHs in western Singapore with pre- and post-implementation data, in collaboration with an acute hospital. A total of 4801 admissions were screened, of which 2045 fever-related admissions were included. The intervention comprised IPC training, policy and process reviews, vaccination support, antimicrobial guidance, and implementation of a structured fever and desaturation pathway to standardise hospital transfer decisions.

Results: New methicillin-resistant *Staphylococcus aureus* (MRSA) acquisitions declined from 14.6% to 9.7% (relative reduction 33.6%, 95% confidence interval [CI] -6.8 to -3.1, $P < 0.001$). Inappropriate fever-related transfers decreased from 12.0% (108/898) to 7.3% (61/839), a relative reduction of 39.2% (absolute reduction 4.8%, 95% CI -7.5 to -2.0, $P < 0.001$). Influenza vaccination coverage exceeded 90% by 2022. Staff surveys indicated sustained adoption, with 84% reporting regular use of the pathway.

Conclusion: A collaborative model integrating acute hospital expertise with NH teams was associated with improved IPC outcomes, increased vaccination coverage, and reduced inappropriate transfers, supporting safer care delivery in long-term care.

Keywords: infection prevention and control, nursing homes, hospital transfers, antimicrobial stewardship, vaccination

CLINICAL IMPACT

What is New

- Multicomponent infection control programme in nursing homes (NHs) was associated with reduced MRSA acquisition and fewer inappropriate hospital transfers.
- Influenza vaccination coverage exceeded 90%, with sustained staff adoption of structured clinical pathways.

Clinical Implications

- Findings highlight the potential of collaborative models between acute hospitals and NHs to strengthen infection prevention and support safer, more appropriate care delivery.

INTRODUCTION

Nursing home (NH) residents are particularly vulnerable to infections due to advanced age, frailty, comorbidities, and the institutional nature of care. Hospital transfers for acute illnesses, particularly infections, increase the risk of delirium, nosocomial complications, functional decline, and caregiver

distress, while imposing additional strain on healthcare resources.¹⁻⁵ Many of these admissions are preventable through strengthened infection prevention and control (IPC) practices, and improved assessment and decision-making on-site in the NHs.⁶⁻⁹

As of 2024, Singapore has 88 NHs and other long-term care facilities providing for a small percentage of the population aged 65 years and above.^{10,11} This proportion of older adults is projected to rise to one-quarter by 2030, leading to anticipated corresponding challenges in optimising care quality and healthcare utilisation.¹² Strengthening IPC capacity and enhancing NH staff competencies to manage residents in situ are critical strategies to mitigate avoidable hospital transfers.^{13,14}

Infections remain among the most common causes of hospital admission from NHs, particularly respiratory and urinary tract infections.¹⁵ Robust IPC systems and vaccination programmes can reduce such events, yet many NHs face implementation barriers including limited resources, competing priorities, as well as consent or cost issues.^{16,17} For NH residents who

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Division of Infectious Diseases, Department of Medicine, National University Hospital, Singapore

² Yong Loo Lin School of Medicine, National University of Singapore, Singapore

³ Division of Infectious Diseases, Department of Medicine, Alexandra Hospital, Singapore

⁴ Division of Acute Internal Medicine, Department of Medicine, National University Hospital, Singapore

Correspondence: Prof Dale Andrew Fisher, Division of Infectious Diseases, Department of Medicine, National University Hospital, NUHS Tower Block 1E Kent Ridge Road, Singapore 119228. Email: mdcfda@nus.edu.sg

Accepted: 12 May 2026

Published Online First: 25 May 2026

develop infections, studies have shown that selected conditions, including pneumonia, can be effectively managed in long-term care settings without transfer when structured assessment and treatment protocols are available.^{18,19} For example, Naughton et al. demonstrated comparable outcomes when pneumonia was treated within NHs versus in the hospital.²⁰

While studies have shown that up to 40% of transfers from NHs may be preventable with structured clinical pathways and enhanced staff training,²¹ NH staff often lack decision-making frameworks and medical support to assess acutely ill residents confidently.¹⁹

Recognising these challenges, the Infectious Diseases Community Program (IDCP) was introduced as a collaborative initiative between an acute hospital and 20 participating NHs in western Singapore. The multidisciplinary IDCP team included infectious disease physicians, IPC nurses, and pharmacists who worked closely with NH nursing and managerial leads. The programme aimed to strengthen IPC systems, increase vaccination uptake, and support appropriate clinical decision-making for infectious diseases.

This study evaluated the effectiveness of the IDCP programme using routinely available data from 6 representative participating NHs with complete pre- and post-intervention datasets, assessing its impact on infection prevention outcomes, vaccination coverage, and the appropriateness of hospital transfers.

METHODS

Study design and setting

A retrospective cohort study was conducted to evaluate the multipronged IDCP intervention. While the IDCP programme was implemented across 20 NHs, the analysis focused on 6 facilities with complete pre- and post-implementation data, representing a range of sizes, staffing models, and medical coverage. Two periods were compared: pre-implementation (2015–2017) and post-implementation (2019–2021), with 2018 designated as a washout year to allow routine embedding of programme components. Although defined as the pre-implementation period, selected IDCP components (including policy review and early phases of staff training) were progressively introduced from 2016 onwards. These early interventions were not uniformly implemented across all facilities and were limited in scope; however, their presence may have attenuated observed differences between pre- and post-implementation outcomes.

Intervention: Infectious Diseases Community Program (IDCP)

The IDCP comprised several coordinated components designed to strengthen IPC governance, workforce capacity, vaccination uptake, and antimicrobial stewardship.

As an initial and central component of the IDCP, the team collaborated with each NH to review and align policy documents, identifying outdated guidance, gaps, and overlapping content. Revised recommendations were made to standardise policies across facilities and align them with the Ministry of Health's National Infection Prevention and Control Guidelines for Long-Term Care Facilities.²² Key areas addressed included hand hygiene, isolation or cohorting of residents with known multidrug-resistant organism (MDRO) carriage, environmental disinfection and cleaning, and reporting procedures for infections with outbreak potential. A follow-up review in 2018 demonstrated notable improvements in governance structures and adherence to standardised policies.

To strengthen workforce capacity, a 3-tiered staff training programme was developed and progressively rolled out between 2016 and 2018.

- (1) Tier 1 was a half-day foundational course (approximately 4 hours) for all staff, delivered on-site or virtually, incorporating a train-the-trainer component to enable in-house continuity.
- (2) Tier 2 was a full-day (8-hour) course for supervisory and audit staff, focusing on hand hygiene auditing, environmental cleaning, and outbreak response.
- (3) Tier 3 was an advanced 2.5-day (20-hour) immersion programme conducted at the acute hospital, combining didactic sessions, case-based discussions, and practical ward-based attachments on surveillance, MDRO management, and leadership in IPC.

Training was adapted for virtual delivery during the COVID-19 period. Across all participating NHs, 37 training sessions were conducted for 1260 participants, and by programme completion, all 20 NHs had staff certified through tiers 1 to 3.

Vaccination strategies

Vaccination for the elderly and medically vulnerable individuals is an important preventive measure against infections and reduces hospital admissions.^{22–24} IDCP engagement sessions identified gaps relating to low vaccine uptake by NH residents and healthcare workers. These included cost, limited knowledge, fear of side effects, and challenges in obtaining consent from residents, staff, or next-of-kin.

To address these barriers, the IDCP team developed A Guide to Vaccination of Residents and Staff in Long-Term Care Facilities in Singapore (internal document), outlining procurement steps, counselling frameworks, and common misconceptions. It was disseminated prior to the post-implementation period and adapted to individual facility needs. Funding from the JurongHealth Fund enabled provision of up to 2500

free pneumococcal vaccinations, primarily administered during the washout period in 2018 and early post-implementation period. While vaccines were largely utilised at the facility level, centralised records of exact uptake and timing were incomplete, precluding precise quantification of coverage and impact on hospital admissions.

Fever and desaturation pathway

A key component of the IDCP was the fever and desaturation pathway (Fig. 1), developed to guide NH staff in assessing residents with fever or respiratory symptoms and standardising decisions on hospital transfer. Early iterations were refined with staff input. The pathway outlined vital-sign monitoring, supportive measures, and recognition of deterioration requiring escalation. Assessments included blood pressure, respiratory rate, oxygen saturation, and mental status using the Glasgow Coma Scale (GCS), supported by targeted training from IPC nurses. Residents meeting severity criteria were transferred, while stable residents received supportive care within the NH. By early 2019, the pathway had been adopted in all participating NHs, with COVID-19-specific elements added subsequently.

Structured clinical pathway used to guide NH staff in assessing residents with fever or desaturation,

standardising decisions on supportive management within the facility versus transfer to acute hospital care.

Training comprised a combination of didactic teaching, case discussions, and simulation. Introductory modules covered fundamental IPC practices, while advanced modules built competence in clinical assessment, recognition of deterioration, and understanding of advance care plans (ACPs) to ensure alignment with residents' goals of care.

Antimicrobial stewardship component

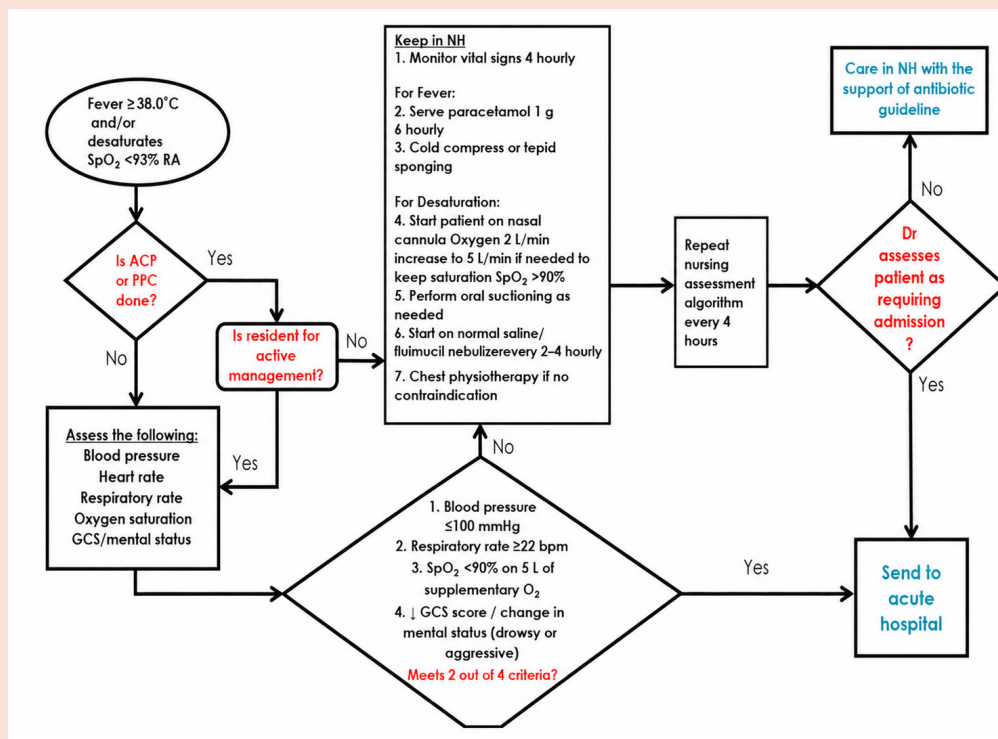
An antimicrobial-choice guideline was developed with primary care physicians, infectious disease specialists, and pharmacists. Introduced prior to the post-implementation period, it addressed an existing gap in national guidance, and enabled facilities to establish in-house formularies for timely access. Its specific impact was not evaluated in this study, and may represent an unmeasured co-intervention.

Data and outcome measures

Patient data were retrieved from hospital electronic medical records, NH transfer forms, and internal IDCP logs. Admissions were screened for any mention of fever as a leading or contributing factor.

To evaluate IPC initiatives, methicillin-resistant *Staphylococcus aureus* (MRSA) acquisition rates

Fig. 1. The Infectious Diseases Community Program's fever and desaturation pathway.



ACP: advance care plan; GCS: Glasgow Coma Scale; NH: nursing home; PPC: preferred plans of care; RA: room air

were used as a surrogate measure. Acquisition was defined as the detection of MRSA at hospital admission in residents whose prior screening results were negative or unknown. During this period, all adult admissions to Singapore's acute hospitals underwent MRSA active surveillance using standard culture-based methods (nares, axilla, and groin swabs), with discharge screening for those initially negative.²⁵ No major changes in sampling sites, laboratory techniques, or surveillance protocols occurred during the study period. Residents who were negative on 1 admission but positive on a subsequent admission were classified as having NH-acquired MRSA, providing an objective, though imperfect, measure of NH IPC performance. MRSA infections (distinct from colonisation) were not analysed separately. As residents may have been admitted to other acute healthcare facilities between admissions to the study site, interim MRSA acquisition events may not have been captured, introducing potential misclassification. However, as the study hospital is the primary referral centre for participating NHs, such events were likely limited and non-differential across study periods. MRSA acquisition was assessed at the level of admission episodes rather than unique residents. As patient-level linkage across admissions and institutions was not available, repeat admissions by the same resident could not be excluded, and the proportion of frequently readmitted individuals could not be determined. This approach was applied consistently across study periods, and any resulting bias is likely non-differential.

To assess the impact of the fever and desaturation pathway, hospital transfers were considered appropriate if residents met at least 2 of the following criteria: systolic blood pressure ≤ 100 mmHg, respiratory rate ≥ 22 breaths/min, oxygen saturation $\leq 90\%$, or a decrease of 2 or more points on the GCS from baseline. Documentation of an ACP was also reviewed. During the COVID-19 pandemic, the pathway was adapted to include suspected or confirmed COVID-19 cases as legitimate indications for hospital admission, consistent with evolving national guidance. Transfers were classified as inappropriate if residents failed to meet these criteria, and had no other urgent medical indication for hospitalisation.

An anonymous survey of NH staff was also conducted to capture qualitative insights into adherence to the pathway and perceptions of whether the approach reduced unnecessary admissions. Surveys were conducted at 2 time points: 2018 (washout year) and 2022 (post-implementation).

Statistical analysis

Data were analysed at both the individual NH and aggregate level, excluding facilities with incomplete

pre-intervention data. Differences in proportions between pre- and post-implementation periods were compared using chi-square tests. All data analyses were performed using R, version 2024.09.0+375 (R Foundation for Statistical Computing, Vienna, Austria), within the RStudio environment (Posit Software, Boston, MA, US).

RESULTS

Over the study period, 4801 hospital admissions of NH residents were screened, of which 2045 admissions involved fever as a primary or contributory reason. Screening for MRSA on hospital admission demonstrated a significant decline in new acquisitions, from 14.6% (329 of 2247) in the baseline period to 9.7% (253/2613) after programme implementation (Table 1). This represented a relative reduction of 33.6% (absolute reduction 5.0 percentage points, 95% confidence interval [CI] -6.8 to -3.1, $P < 0.001$). The decline coincided with greater uptake of standard precautions, regular staff training, and the establishment of multidisciplinary IPC committees within most participating NHs. All established facilities recorded meaningful improvements, with relative reductions ranging from 30.7% to 48.9%. NH C achieved the largest improvement, from 19.0% to 9.7%—a 48.9% relative reduction, although CIs for individual facilities were wide due to year-to-year variability (Table 1).

Percentages represent the proportion of new MRSA-positive cases among residents screened in the respective period. Facility-level data were available only as aggregated percentages derived from institutional reports, and corresponding numerator and denominator values were not consistently available.

Training outcomes reflected broad engagement across NH staff. A total of 20 tier 1 sessions, including 10 face-to-face, 5 virtual, and 5 train-the-trainer modules, trained 959 participants. Ten tier 2 sessions trained 170 staff in supervisory roles, while 8 tier 3 sessions reached 131 staff identified for IPC leadership responsibilities. Most training occurred during the washout year (2018), with additional sessions conducted through 2019 to accommodate new facilities and staff turnover.

Analysis of inappropriate transfers focused on 4 NHs (A to D) with complete pre- and post-intervention data. Across these facilities, the overall proportion of inappropriate fever-related transfers decreased significantly from 12.0% (108/898) before implementation to 7.3% (61/839) afterwards, representing a 39.2% relative reduction (absolute reduction 4.8 percentage points, 95% CI -7.5 to -2.0, $P < 0.001$; Table 2). Improvements were consistent across most facilities, with the exception of NH A, which showed no significant change. Two additional NHs (E and F),

Table 1. Changes in new MRSA acquisition rates pre- and post-implementation of the Infectious Diseases Community Program (2015–2017 versus 2019–2021).

Nursing home	Pre-Intervention, %	Post-intervention, %	Change in MRSA acquisition, % (95% CI)
NH A	14.0	9.7	-4.3 (-55.9 to 47.2)
NH B	20.7	13.7	-7.0 (-67.1 to 53.1)
NH C	19.0	9.7	-9.3 (-64.9 to 46.2)
NH D	14.3	9.3	-5.0 (-56.5 to 46.5)
NH E ^a	4.7	7.7	+3.0
NH F ^a	4.0	7.3	+3.3
Overall ^b	14.6	9.7	-5.0 (-6.8 to -3.1) ($P<0.001$)

CI: confidence interval; MRSA: methicillin-resistant *Staphylococcus aureus*; NH: nursing home; P : P value.

^a For NH E and NH F, there were only had data for 2017 in the pre-intervention period.

^b Overall analysis based on total MRSA screenings and acquisitions across all facilities.

which lacked full pre-intervention data, achieved post-implementation inappropriate-transfer rates (7.8% and 7.9%) within the same range as those observed in the 4 fully evaluated facilities.

Staff perceptions similarly reflected these quantitative findings. An anonymous survey conducted in 2018 across 3 NHs demonstrated high engagement with the fever and desaturation pathway, with between 63.6% and 91% of staff reporting that they “always” used the pathway when assessing residents with fever. More than 95% were familiar with the tool, and only 2.2% reported awareness without use. Regarding impact, between 54.5% and 75.6% of staff reported that the pathway “often” prevented unnecessary transfers, while a further 20% to 45.5% reported that

it prevented transfers “a few times”. A follow-up survey in 2022 confirmed sustained use, with 78% reporting continued adoption and 84% (combining “always” and “very often”) indicating regular use for clinical decision-making. Across participating NHs, all staff surveyed either strongly agreed (58%) or somewhat agreed (42%) that the programme had improved their facility’s capacity to triage fever cases appropriately.

Influenza vaccination uptake trends improved across most participating NHs over the study period. Although coverage varied between facilities, by 2021, 7 of 9 NHs with available data had reached at least 80% influenza vaccination coverage among residents (including 4 with coverage above 90%;

Table 2. Changes in inappropriate fever transfers pre- and post-implementation of the Infectious Diseases Community Program (2015–2017 versus 2019–2021).

Nursing home	Pre-intervention total no. (inappropriate %)	Post-intervention total no. (inappropriate %)	Change in inappropriate transfers, % (95% CI)	P value
NH A	324 (6.8)	292 (9.2)	+2.5 (-1.8 to 6.8)	0.41
NH B	190 (20.0)	167 (7.2)	-12.8 (-19.7 to -5.9)	<0.001
NH C	236 (14.8)	241 (7.1)	-7.8 (-13.3 to -2.2)	<0.001
NH D	148 (8.8)	139 (3.6)	-5.2 (-10.7 to 0.3)	<0.001
NH E ^a	25 (8.0)	129 (7.8)	-0.2 (-11.8 to 11.3)	0.94
NH F ^a	27 (18.5)	127 (7.9)	-10.6 (-26.0 to 4.7)	<0.001
Overall ^b	898 (12.0)	839 (7.3)	-4.8 (-7.5 to -2.0)	<0.001

CI: confidence interval; NH: nursing home

^a For NH E and NH F, there were only data for 2017 in the pre-intervention period.

^b Overall results exclude NH E and NH F due to incomplete pre-intervention data.

Fig. 2), and 6 of 9 had reached the 80% target among healthcare workers (including 3 above 90%). By the end of the programme, more than 90% of residents had received at least 1 influenza vaccination. Additionally, educational sessions and vaccination guidelines were well received, with a majority of 89% of healthcare workers reporting that these resources improved their knowledge and understanding of different vaccines.

Each panel represents 1 NH (NH A–J). Naming is consistent with Tables 1 and 2 for NH A–F; additional NHs (NH G–J) are included in this figure where vaccination data were available but were not part of the MRSA and transfer analyses due to incomplete datasets. Solid orange lines show annual vaccination coverage (%) for NH residents. Dashed line indicates the 80% coverage target. Gaps in trend lines indicate years for which data were not available.

DISCUSSION

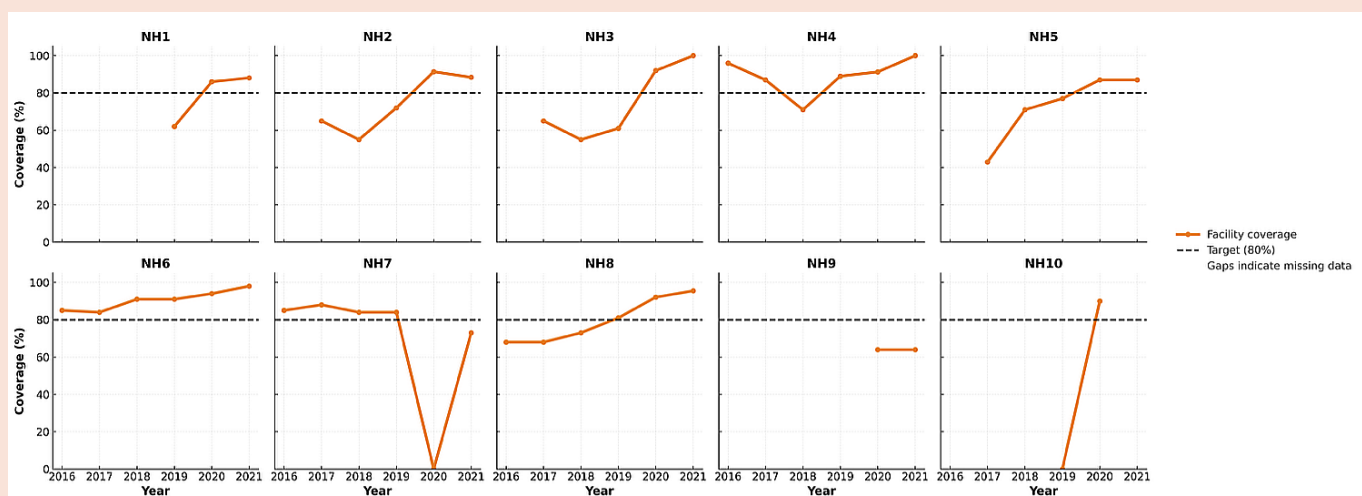
This study demonstrated that a multipronged infectious disease programme in NHs was associated with measurable improvements in IPC and clinical decision-making. New MRSA acquisitions declined significantly across most established facilities, inappropriate fever-related transfers were reduced by nearly 40%, and influenza vaccination coverage exceeded 90% of residents by 2022. However, NH E and NH F demonstrated increases in MRSA acquisition rates in the post-implementation period. These facilities had only 1 year of baseline data, limiting the reliability of pre- and post-implementation comparisons. In addition, later programme onboarding,

smaller sample sizes, and unmeasured factors such as resident case-mix, staffing variability, and referral patterns may have contributed to the observed increases. Staff surveys confirmed high and sustained adoption of the fever and desaturation pathway, highlighting increased confidence and more standardised triage decisions. Together, these outcomes reflect not only adherence to structured IPC processes but also the value of collaboration and sustained engagement between acute hospital IPC teams and NH leadership.

Working with NHs presents inherent challenges. These facilities are often understaffed and underfunded, with leadership balancing competing priorities.^{18,26} Support from acute hospital teams may be perceived as adding workload or introducing standards that do not align with existing systems, particularly when documentation is manual and clinical outcomes are not routinely captured. Despite these barriers, staff enthusiasm in participating in training and assuming IPC leadership roles demonstrated that improvement was both recognised and welcomed.

Measuring IPC outcomes is difficult even in acute care settings. In this real-world intervention, prospective surveillance within NHs was not feasible; however, leveraging acute hospital systems enabled a practical evaluation. Routine screening for MRSA carriage on hospital admission and discharge provided a consistent surrogate of IPC standards. These data were extracted directly from hospital IPC surveillance systems, allowing for consistent and objective measurement across the study period. Although limited by swab sensitivity, and the assumption that a negative discharge swab followed by a

Fig. 2. Annual influenza vaccination rates among residents in participating nursing homes (2016–2021). Data provided by institutional quality reports.



NH: nursing home

positive admission swab reflected NH acquisition, the use of identical processes pre- and post-intervention, alongside consistent declines across multiple facilities, support the validity of the observed improvements. Direct quantitative measures of IPC practices (e.g. hand hygiene compliance rates) were not consistently available across facilities; however, multiple system-level indicators, including policy standardisation, establishment of IPC governance structures, and sustained staff training, support improved adherence to IPC practices over time.

There are also meaningful risks and costs associated with frequent hospital transfers for self-limiting conditions. The structured fever and desaturation pathway reduced admissions that did not meet objective transfer criteria. Staff surveys reinforced that the pathway increased clinical confidence and provided a standardised approach to fever management. Inappropriate transfers from NHs should be regarded as a valuable quality metric, particularly given increasing global pressure on acute care bed availability. Although this study required manual data extraction, real-time monitoring and feedback could more effectively prevent unnecessary transfers and their negative consequences for residents, families, and health systems.

Strengthening medical support within NHs will be essential to build on progress. Some facilities lacked daily medical input, which constrained staff capacity to manage acute changes. Enhanced access to primary care, deployment of advanced practice nurses, and expansion of hospital-at-home services are potential strategies to fill this gap.²⁶⁻²⁸ Staff turnover was another challenge, as new employees were not always adequately trained in the pathway. Limited documentation of ACPs or preferred plans of care also introduced uncertainty, as staff were sometimes unsure whether hospitalisation was consistent with the residents' preferences. Although awareness of advance care planning improved, the absence of clear directives complicated decision-making in some cases.

Several limitations should be considered. As expected in a real-world programme evaluation, this retrospective study relied on routinely collected data and was therefore subject to variability in documentation and completeness. Baseline demographic and clinical characteristics of residents were not consistently available, limiting adjustment for potential confounders. The analysis was conducted within a single regional cluster of NHs, which may affect generalisability to other settings with different resources and care structures. External factors, including the COVID-19 pandemic and the evolving national policies on infection control and hospital transfer criteria, may also have influenced observed outcomes. Formal health economic evaluation of the programme was not conducted. As such, cost-

effectiveness and long-term financial sustainability could not be assessed and represent important areas for future study.

However, the consistency of improvements across multiple facilities and outcome measures, alongside stable data collection processes before and after implementation, supports a meaningful programme effect. These findings should therefore be interpreted as a pragmatic evaluation of a complex, multicomponent intervention, reflecting the realities of implementation in long-term care settings.

CONCLUSION

In conclusion, this study provides evidence that a collaborative strategy encompassing IPC policy reviews, staff education, and vaccination support can strengthen infection control in NHs. The use of a structured clinical pathway improved the appropriateness of hospital transfers, aligning decisions with objective clinical criteria while supporting resident well-being. Although resource-intensive, such work is feasible and sustainable when built on partnership, shared goals, and an appreciation of the unique challenges faced by NHs and their staff.

Supplementary material

Annex S1. STROBE checklist.

Acknowledgements

The authors gratefully acknowledge the contributions of the IDCP team, and the Infection Prevention Team at the National University Hospital whose expertise and commitment were essential in the design and implementation of this work. We thank Mr Thomas Soo, Dr Natasha Bagdasarian, Dr James Molton, Ms Nai Ying Jiin, Ms Hon Qi, Ms Alison Cheng, Ms Bindu Karunakaran, Ms Amira Toriman, Ms Jessica Michaels, and Ms Tan Shire Yang among countless others for their invaluable support.

Ethics statement

Ethics approval for this study was granted by the National Healthcare Group Domain Specific Review Board (2018/00533), with a waiver of individual consent granted on the basis that no identifiable patient information was disclosed.

Declaration

The authors declare there are no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript. The Infectious Diseases Community Program received partial support from the Jurong Health Fund, which enabled pneumococcal vaccination for NH residents. This funding did not support the present study, and the funder had no role in study design, data collection and analysis, or manuscript writing.

Reporting Guidelines

The authors used the STROBE reporting guideline³⁰ to draft this manuscript, and the STROBE reporting checklist³¹ during editing (Supplementary Annex S1).

REFERENCES

- Webb HT, Lieb KM, Stump TE, et al. Describing transfers originating out-of-facility for nursing home residents. *J Am Med Dir Assoc* 2022;23:105-10.
- Quah WC, Leong CJ, Chong E, et al. Unplanned hospitalisations among subsidised nursing home residents in Singapore: insights from a data linkage study. *Ann Acad Med Singap* 2024;53:657-69.
- Pérez-Cruz M, Parra-Anguita L, López-Martínez C, et al. Burden and anxiety in family caregivers in the hospital that debut in caregiving. *Int J Environ Res Public Health* 2019;16:3977.
- Lemoyne SE, Herbots HH, De Blick D, et al. Appropriateness of transferring nursing home residents to emergency departments: a systematic review. *BMC Geriatr* 2019;19:17.
- Bally KW, Nickel C. Acute hospital admissions among nursing home residents—benefits and potential harms. *Praxis (Bern 1994)* 2013;102:987-91.
- Ouslander JG, Lamb G, Tappen R, et al. Interventions to reduce hospitalizations from nursing homes: evaluation of the INTERACT II collaborative quality improvement project. *J Am Geriatr Soc* 2011;59:745-53.
- Trahan LM, Spiers JA, Cummings GG. Decisions to transfer nursing home residents to emergency departments: a scoping review of contributing factors and staff perspectives. *J Am Med Dir Assoc* 2016;17:994-1005.
- Saliba D, Kington R, Buchanan J, et al. Appropriateness of the decision to transfer nursing facility residents to the hospital. *J Am Geriatr Soc* 2000;48:154-63.
- Ouslander JG, Lamb G, Perloe M, et al. Potentially avoidable hospitalizations of nursing home residents: frequency, causes, and costs. *J Am Geriatr Soc* 2010;58:627-35.
- Immigration & Checkpoints Authority, Singapore. Report on Registration of Births and Deaths 2020. https://www.ica.gov.sg/docs/default-source/ica/stats/annual-bd-statistics/stats_2020_annual_rbd_report.pdf?sfvrsn=21a3e11e_4. Accessed 6 May 2026.
- Ministry of Health, Singapore. Health Facilities. 26 May 2025. https://www.moh.gov.sg/others/resources-and-statistics/health-facilities/?utm_source=chatgpt.com/. Accessed 14 May 2026.
- World Bank. Population ages 65 and above (% of total population). <https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS>. Accessed 27 August 2025.
- Rolland Y, Mathieu C, Tavassoli N, et al. Factors associated with potentially inappropriate transfer to the emergency department among nursing home residents. *J Am Med Dir Assoc* 2021;22:2579-86.
- Thandar MM, Rahman MO, Haruyama R, et al. Effectiveness of infection control teams in reducing healthcare-associated infections: a systematic review and meta-analysis. *Int J Environ Res Public Health* 2022;19:17075.
- Lee MH, Lee GA, Lee SH, et al. Effectiveness and core components of infection prevention and control programmes in long-term care facilities: a systematic review. *J Hosp Infect* 2019;102:377-93.
- Bloch N, Männer J, Gardiol C, et al. Effective infection prevention and control measures in long-term care facilities in non-outbreak and outbreak settings: a systematic literature review. *Antimicrob Resist Infect Control* 2023;12:113.
- Chambers D, Cantrell A, Preston L, et al. Reducing unplanned hospital admissions from care homes: a systematic review. *Health Soc Care Deliv Res* 2023;11:1-130.
- Basso I, Gonella S, Bassi E, et al. Impact of quality improvement interventions on hospital admissions from nursing homes: a systematic review and meta-analysis. *J Am Med Dir Assoc* 2024;25:105261.
- Høyvik E, Doupe MB, Ågotnes G, et al. Barriers to healthy transitions between nursing homes and emergency departments. *Geriatr Nurs* 2024;59:639-45.
- Naughton BJ, Mylotte JM, Tayara A. Outcome of nursing home-acquired pneumonia: derivation and application of a practical model to predict 30-day mortality. *J Am Geriatr Soc* 2000;48:1292-9.
- Graverholt B, Forsetlund L, Jamtvedt G. Reducing hospital admissions from nursing homes: a systematic review. *BMC Health Serv Res* 2014;14:36.
- Ministry of Health Singapore. The national infection prevention and control guidelines for long-term care facilities. Revised 2021. https://isomer-user-content.by.gov.sg/95/cba04fe1-6ec5-497a-8e4f-de4566ffa6e8/national-ipc-guidelines-for-long-term-care-facilities_2021_revised_2nd-edition_moh-website.pdf. Accessed 18 May 2026.
- Gravenstein S, Dahal R, Gozalo PL, et al. A cluster randomized controlled trial comparing relative effectiveness of two licensed influenza vaccines in US nursing homes: design and rationale. *Clin Trials* 2016;13:264-74.
- Cai S, Temkin-Greener H. Influenza vaccination and its impact on hospitalization events in nursing homes. *J Am Med Dir Assoc* 2011;12:493-8.
- Streeter AJ, Rodgers LR, Masoli J, et al. Real-world effectiveness of pneumococcal vaccination in older adults: cohort study using the UK Clinical Practice Research Datalink. *PLoS One* 2022;17:e0275642.
- Fisher DA, Tambyah PA, Lin R, et al. Sustained MRSA control in a hyper-endemic tertiary acute care hospital with infrastructure challenges in Singapore. *J Hosp Infect* 2013;85:141-8.
- Yang XF, Guo JY, Peng WS, et al. Global perspectives on challenges, coping strategies, and future preparedness of nursing home staff during COVID-19: a systematic review and meta-synthesis. *BMC Health Serv Res* 2025;25:872.
- Sturmberg JP, Gainsford L, Goodwin N, et al. Systemic failures in nursing home care: a scoping study. *J Eval Clin Pract* 2024;30:484-96.
- Senek M, Robertson S, Taylor B, et al. Consequences of understaffing on type of missed community care: a cross-sectional study. *Int J Nurs Stud Adv* 2022;4:100075.
- von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007;147:573-7.
- von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453-7.

Burden trends of Alzheimer's disease and other dementias in China, 1990–2023: Sex, age, and risk-attributable patterns

Yao Li¹ MD, Shuai Wang² MD, Kaiyue Xie¹ MD, Siyv Qian¹ MD, Mingyue Ma¹ MD, Haining Zhang¹ MD

ABSTRACT

Introduction: China's population has aged rapidly in recent decades, accompanied by a steady rise in Alzheimer's disease and other dementias (ADOD). Understanding long-term national trends is important for anticipating future healthcare needs.

Methods: The authors analysed Global Burden of Disease (GBD) 2023 estimates for China from 1990 to 2023, following the Guidelines for Accurate and Transparent Health Estimates Reporting, with STROBE used as a secondary reporting reference. Deaths, disability-adjusted life years (DALYs), incidence, and prevalence were examined together with crude rates, age-standardised rates (ASRs), and average annual percentage changes. Results were evaluated and stratified by sex and broad age group. Modelled risk-attributable burden was assessed for selected GBD risk factors—ambient particulate matter pollution, high fasting plasma glucose, smoking, and household air pollution from solid fuels.

Results: Over the study period, the number of people affected increased markedly. Deaths rose from 122,411 in 1990 to 586,002 in 2023, while DALYs and prevalent cases more than tripled. In contrast, ASR showed only modest upward changes. Women had higher ASRs in 2023, although mortality- and DALY-related rates increased more quickly among men. The burden remained concentrated in those aged 75 years and older. Among the selected exposures, ambient particulate matter pollution accounted for the largest attributable share in 2023, whereas household air pollution declined over time.

Conclusion: The marked increase in absolute ADOD burden, together with smaller changes in ASR, is consistent with demographic ageing being an important contributor to rising dementia-related service needs in China. The risk-attributable estimates suggest changing population-level patterns for selected environmental and metabolic exposures, but they should not be interpreted as individual-level causal effects.

Keywords: ageing, Alzheimer's disease and other dementias, China, disease burden, risk factors, trend analysis

CLINICAL IMPACT

What is New

- This study provides an updated national profile of the burden of Alzheimer's disease and other dementias in China from 1990 to 2023, based on Global Burden of Disease 2023 estimates.
- By comparing absolute counts, crude rates, and age-standardised rates, this study distinguishes health-system service demand from age-adjusted epidemiological change.

Clinical Implication

- The study summarises sex, age, and selected risk-attributable patterns to support dementia surveillance, ageing-adapted care planning, and prevention-oriented policy discussion.

INTRODUCTION

Alzheimer's disease and other dementias (ADOD) are now central to the global non-communicable disease agenda, because they combine high late-life prevalence, prolonged disability, substantial caregiver burden, and major health-system costs.^{1,2} As life expectancy rises and fertility declines, population age structures shift towards older age groups, and the number of people living with cognitive disorders is expected to keep growing in many countries, including China.^{3,4}

China represents a critical setting for ADOD burden monitoring for several reasons. First, the speed and scale of demographic transition are historically unusual, with rapid expansion of the elderly population. Second, regional heterogeneity in socioeconomic development, healthcare access, vascular risk profiles, and environmental exposures may shape both diagnosis and disease progression. Third, shifts in

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Department of Neurology and Neuroscience Centre, The First Hospital of Jilin University, Changchun, China

² Department of Gastric and Colorectal Surgery, General Surgery Center, The First Hospital of Jilin University, Changchun, China

Correspondence: Dr Haining Zhang, Department of Neurology and Neuroscience Centre, The First Hospital of Jilin University, Changchun 130021, China. Email: zhanghaining@jlu.edu.cn.

Accepted: 15 July 2026

Published Online First: 22 July 2026

major risk factors, including ambient particulate exposure, metabolic risk, and smoking, may alter future trajectories even when age-standardised rates (ASR) remain relatively stable.⁵⁻⁷ In this context, separating absolute burden growth from standardised rate change is essential for interpretation and policy planning.

Current evidence indicates that dementia burden is influenced by both non-modifiable and modifiable factors. Age is the strongest determinant, while sex-related biological and social pathways may contribute to differential susceptibility, clinical presentation, and survival.⁸⁻⁹ Major prevention frameworks have highlighted potentially modifiable contributors, including vascular and metabolic factors, smoking, and environmental exposures.¹⁰⁻¹² At the same time, diagnostic criteria and disease conceptualisation have evolved, from syndrome-based to biologically informed frameworks, complicating temporal interpretation across decades.¹³ Robust trend analyses must therefore pair epidemiological indicators with cautious interpretation of changing detection patterns and competing mortality.

Large comparative platforms, such as the Global Burden of Disease (GBD) programme, offer internally consistent time series on incidence, mortality, prevalence, disability, and risk-attributable burden, providing a robust basis for long-horizon national assessment.^{14,15} For policy decisions, however, country-level interpretation remains essential. Interpretation is further complicated by the 2020–2023 COVID-19 period, during which many settings reported reduced access to routine cognitive care, delayed diagnosis, and disruption of community support.¹⁶ Although annual burden models may smooth short-term shocks, these disruptions can still influence recent slope estimates and endpoint levels.

Against this background, the present study provides a policy-oriented, non-causal trend analysis of ADOD in China from 1990 to 2023, integrating 4 dimensions: (1) long-run changes in deaths, disability-adjusted life years (DALYs), incidence, and prevalence; (2) contrasts between absolute burden and ASR; (3) sex-specific and age-group trajectories; and (4) temporal shifts in selected risk-attributable burden (ambient particulate matter pollution, high fasting plasma glucose, smoking, and household air pollution from solid fuels).

METHODS

Study design and analytical scope

This was a nationwide, longitudinal burden analysis of ADOD in China from 1990 to 2023. The analysis focused on temporal changes in incidence, deaths, prevalence, and DALYs, with emphasis on both absolute burden expansion and rate-based epidemiological dynamics. To avoid relying on a single statistical

expression of disease burden, the authors jointly evaluated all-age rates and age-standardised rates, and further examined sex-specific and age-specific heterogeneity.

Data source, reporting framework, and acquisition

This study was reported primarily in accordance with Guidelines for Accurate and Transparent Health Estimates Reporting, with STROBE used as a secondary reference for observational reporting structure.^{17,18} All estimates were extracted from GBD 2023 using the China-specific series for the cause category “Alzheimer’s disease and other dementias”.^{14,15} The extracted variables included location, year, sex, age group, cause, measure, metric, point estimate, and 95% uncertainty interval (UI). The measures were incidence, prevalence, deaths, and DALYs; the metrics were numbers, crude rates, and ASR per 100,000 population (age-standardised to the GBD standard population). The extraction date, data filters, and analysis code are provided in Supplementary Appendix S1 to improve reproducibility.

In GBD 2023, ADOD is a modelled population-level cause category that includes Alzheimer’s disease and other dementias rather than biomarker-confirmed Alzheimer’s disease alone.^{13,15,19} The estimates are therefore interpreted as dementia burden estimates at the population level. They do not provide subtype-specific clinical diagnoses and may be influenced by changes in case ascertainment, diagnostic practice, mortality coding, and data availability across calendar years.

GBD estimates are derived from multiple data sources and statistical modelling procedures.^{14,15} For this reason, all results are presented as health estimates with uncertainty rather than as directly observed registry counts. Where possible, the authors report 95% UIs for GBD estimates and 95% confidence intervals (CIs) for joinpoint-derived average annual percentage changes (AAPCs).

Outcomes and epidemiological metrics

The primary burden indicators were incident cases, deaths, prevalent cases, and DALYs. DALYs were used because they combine years of life lost due to pre-mature mortality and years lived with disability, and therefore capture both fatal and non-fatal health loss relevant to dementia service planning. For each indicator, absolute numbers and rates per 100,000 population were reported. Crude rates were used to reflect overall service demand in the population, whereas ASRs were used to describe age-adjusted temporal patterns.^{14,15}

Stratified analysis framework

Sex-stratified analyses were used to describe burden levels and rates of change among females and males

across all 4 outcomes.¹⁹ Age-stratified analyses were used to describe whether burden and trends were concentrated in older groups. The current age grouping of 15–49, 50–74, and ≥75 years was retained to maintain consistency across the extracted outcome and risk-attributable datasets. The authors recognise that more detailed older-age strata, such as 65–74, 75–84, and ≥85 years, would be more informative for geriatric policy interpretation and have noted this as a limitation.

Trend estimation and statistical interpretation

Long-term temporal trends were quantified using log-linear regression in the Joinpoint Regression Program version 5.2.0.0 (National Cancer Institute, Bethesda, US).²⁰ Annual rates or ASRs were used as dependent variables. The maximum number of joinpoints was set to 6 according to the number of annual observations and the default settings of the Joinpoint Regression Program. Join-points were selected using the programme's permutation-based model-selection procedure,²⁰ and AAPCs with 95% CIs were calculated for 1990–2023.²¹ GBD 95% UIs were reported for burden estimates but were not propagated into the joinpoint regression models; therefore, AAPCs should be interpreted as descriptive trend summaries based on point estimates.^{15,21} AAPCs were interpreted descriptively: an AAPC greater than 0 with a CI excluding 0 indicated an increasing trend, an AAPC less than 0 with a CI excluding 0 indicated a decreasing trend, and a CI including 0 indicated no clear long-term directional change.²¹ Because multiple subgroup-specific AAPCs were estimated, statistical significance was not used as confirmatory evidence of causal differences between groups. Data cleaning and figure production were conducted in R version 4.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

Risk-attributable burden assessment

Risk-attributable burden was analysed for 4 selected GBD-attributable risk factors: (1) ambient particulate matter pollution, (2) high fasting plasma glucose, (3) smoking, and (4) household air pollution from solid fuels. These estimates represent modelled population-level attributable burden based on exposure distributions, relative-risk functions, population-attributable fractions, and theoretical minimum risk exposure levels used in the GBD comparative risk assessment framework.²² They should not be interpreted as individual-level causal effects or as evidence that reducing a given exposure would produce proportional reductions in dementia burden. The analysis did not cover all potentially modifiable dementia risk factors.

Data validation and reporting control

All values included in the narrative were checked. Count-level values, rates, percentage changes, and

AAPCs were aligned to table source values to ensure textual consistency with figures and tables. Results are presented as point estimates with 95% UIs or 95% CIs according to the source metrics.

RESULTS

National burden levels and long-term changes

From 1990 to 2023, the absolute burden of ADOD in China increased across all 4 outcomes (Table 1, Fig. 1). Incident cases rose from 755,780 to 3,476,644 (+360.01%), deaths from 122,411 to 586,002 (+378.72%), prevalent cases from 4,352,315 to 20,299,243 (+366.40%), and DALYs from 2,813,268 to 11,646,909 (+314.00%). Standardised indicators also increased over the same period. Age-standardised incidence rate (ASIR) rose from 126.71 to 156.63 per 100,000 (AAPC 0.67, 95% CI 0.53–0.80), age-standardised mortality rate (ASMR) from 27.91 to 28.65 (AAPC 0.27, 95% CI 0.03–0.51), age-standardised prevalence rate (ASPR) from 735.54 to 918.83 (AAPC 0.71, 95% CI 0.55–0.87), and age-standardised DALY rate (ASDR) from 509.08 to 539.50 (AAPC 0.25, 95% CI 0.08–0.42) (Table 1, Fig. 1). At the all-age crude-rate level, mortality increased from 10.38 to 40.96 per 100,000 (AAPC 4.50, 95% CI 4.30–4.69), DALY rate from 238.59 to 814.10 (AAPC 3.92, 3.79–4.05), incidence rate from 64.10 to 243.01 (AAPC 4.14, 3.99–4.29), and prevalence rate from 369.11 to 1418.89 (AAPC 4.18, 4.02–4.35) (Table 2, Fig. 1).

Sex-specific burden trends

In 2023, females showed higher standardised levels than males for all 4 indicators (Table 1, Fig. 2). Female versus (vs) male values were 179.64 vs 129.45 for ASIR, 32.13 vs 23.87 for ASMR, 1059.87 vs 743.39 for ASPR, and 617.86 vs 439.61 for ASDR (per 100,000).

Sex-specific trend slopes differed by outcome. For mortality and DALYs, AAPCs were higher in males than females (ASMR 0.60 [95% CI 0.15–1.06] vs 0.18 [–0.04–0.40]; ASDR 0.54 [0.33–0.74] vs 0.22 [0.06–0.38]). For incidence and prevalence, females showed slightly higher AAPCs than males (ASIR 0.74 [0.58–0.89] vs 0.68 [0.57–0.80]; ASPR 0.81 [0.63–0.98] vs 0.67 [0.54–0.81]) (Table 1).

In absolute terms, 2023 female counts were 2,177,773 incident cases, 378,647 deaths, 12,830,656 prevalent cases, and 7,424,090 DALYs; corresponding male counts were 1,298,871, 207,355, 7,468,587, and 4,222,818 (Table 1).

Age-specific burden trends

Age-stratified analyses showed that burden levels were highest in older adults (Table 2, Fig. 1). In 2023, the ≥75 years group recorded the largest counts in

Table 1. Age-standardised rates and burden of Alzheimer's disease and other dementias in China, 1990–2023.

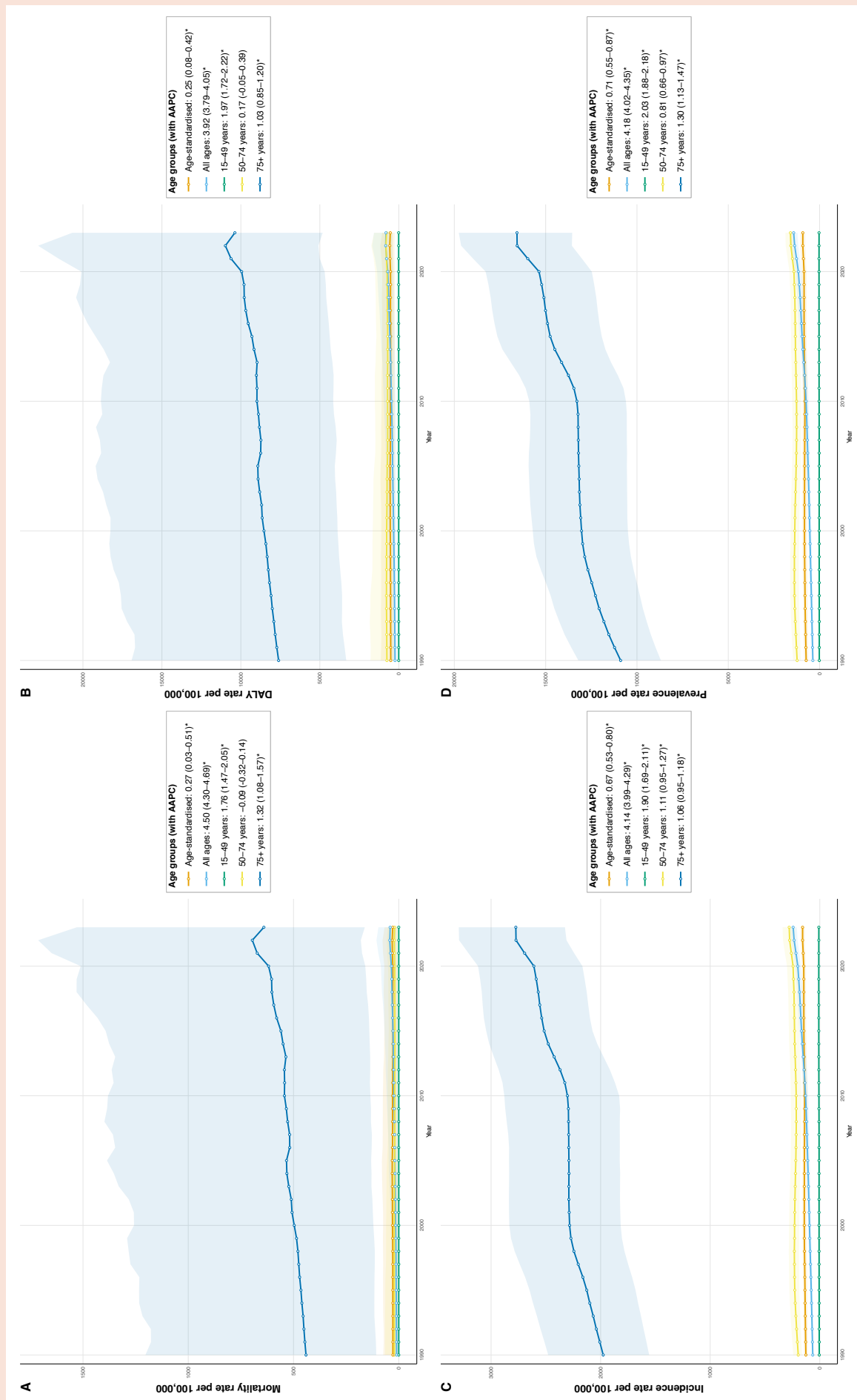
Indicator	Sex	Number (1990, 95% UI)	Number (2023, 95% UI)	Percentage change (%)	ASR (1990, 95% UI)	ASR (2023, 95% UI)	AAPC of ASR (95% CI)
DALYs	Both	2,813,268 (1,276,794–6,411,019)	11,646,909 (5,719,528–22,514,747)	314.00	509.08 (223.96–1146.10)	539.50 (262.25–1051.85)	0.25 (0.08–0.42) ^a
	Female	1,862,725 (825,430–4,077,973)	7,424,090 (3,655,395–14,117,692)	298.56	584.88 (254.25–1315.38)	617.86 (302.41–1171.24)	0.22 (0.06–0.38) ^a
	Male	950,544 (442,767–2,273,633)	4,222,818 (1,933,620–8,413,971)	344.25	386.34 (175.23–868.49)	439.61 (200.23–884.85)	0.54 (0.33–0.74) ^a
Deaths	Both	122,411 (29,124–341,119)	586,002 (146,024–1,385,075)	378.72	27.91 (6.89–73.63)	28.65 (7.10–67.37)	0.27 (0.03–0.51) ^a
	Female	85,376 (20,219–232,011)	378,647 (101,065–894,116)	343.50	31.88 (7.67–84.86)	32.13 (8.53–75.49)	0.18 (–0.04–0.40)
	Male	37,035 (8641–106,083)	207,355 (44,959–494,003)	459.89	20.61 (4.92–57.81)	23.87 (5.15–57.88)	0.60 (0.15–1.06) ^a
Incidence	Both	755,780 (637,685–854,211)	3,476,644 (2,998,723–3,884,422)	360.01	126.71 (108.32–143.87)	156.63 (136.58–175.49)	0.67 (0.53–0.80) ^a
	Female	481,028 (405,077–545,550)	2,177,773 (1,887,505–2,431,095)	352.73	141.86 (121.48–160.77)	179.64 (156.37–200.31)	0.74 (0.58–0.89) ^a
	Male	274,753 (233,423–312,968)	1,298,871 (1,118,042–1,460,288)	372.74	104.24 (88.43–119.41)	129.45 (112.39–146.09)	0.68 (0.57–0.80) ^a
Prevalence	Both	4,352,315 (3,633,202–5,012,549)	20,299,243 (17,341,557–23,309,840)	366.40	735.54 (614.67–845.18)	918.83 (784.59–1058.24)	0.71 (0.55–0.87) ^a
	Female	2,747,076 (2,290,143–3,178,182)	12,830,656 (10,929,366–14,748,557)	367.07	822.38 (687.59–948.16)	1059.87 (905.87–1221.75)	0.81 (0.63–0.98) ^a
	Male	1,605,239 (1,339,491–1,836,615)	7,468,587 (6,362,958–8,563,534)	365.26	598.31 (496.57–690.11)	743.39 (632.99–854.73)	0.67 (0.54–0.81) ^a

AAPC: average annual percentage change; ASR: age-standardised rate; CI: confidence interval; DALY: disability-adjusted life year; UI: uncertainty interval

Rates are per 100,000 population.

^a Indicates that the 95% CI for the AAPC excludes 0.

Fig. 1. Trends in Alzheimer's disease and other dementias indicators in China by age group, 1990–2023. (A) Mortality rate per 100,000 population; (B) DALY rate per 100,000 population; (C) incidence rate per 100,000 population; (D) prevalence rate per 100,000 population.



AAPC: average annual percentage change; DALY: disability-adjusted life year
 Shaded intervals, where shown, represent 95% uncertainty intervals.
 * Indicates that the 95% confidence interval for the AAPC excludes 0.

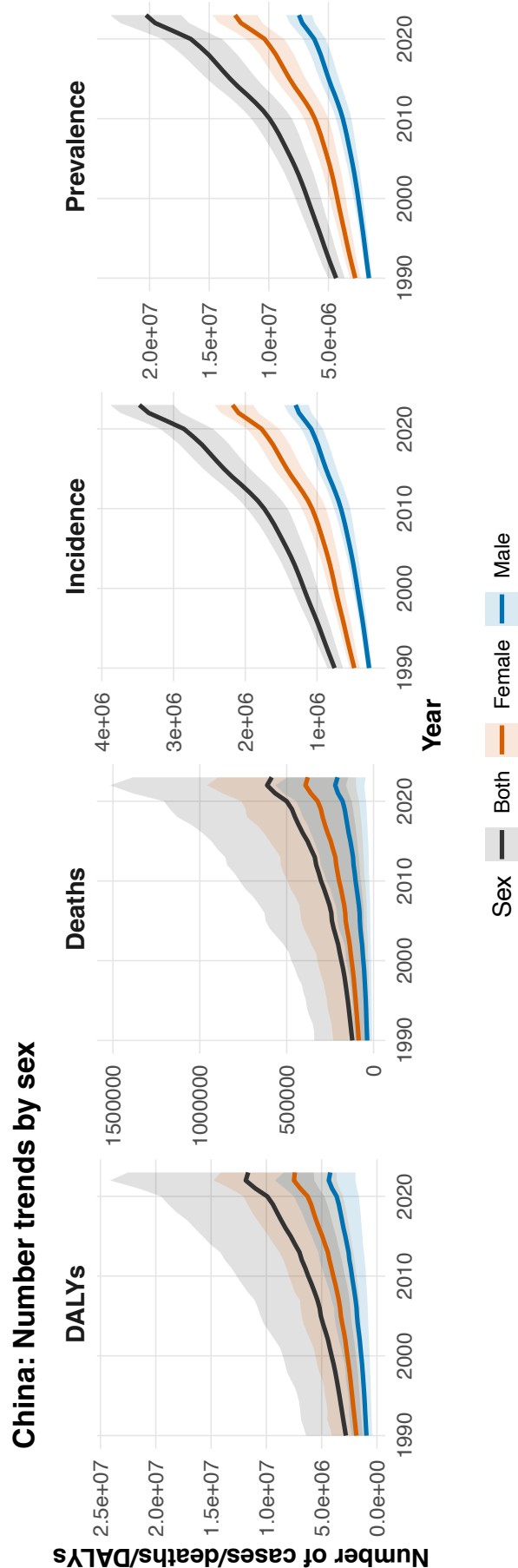
Table 2. Age-specific trends in Alzheimer's disease and other dementias burden in China, 1990–2023.

Metric	Measure	Number (1990, 95% UI)	Number (2023, 95% UI)	Percentage change (%)	Rate (1990, 95% UI)	Rate (2023, 95% UI)	AAPC of rate (95% CI)
Mortality	All ages	122,411 (29,124–341,119)	586,002 (146,024–1,385,075)	378.72	10.38 (2.47–28.93)	40.96 (10.21–96.81)	4.50 (4.30–4.69) ^a
	15–49 years	232 (31–765)	387 (54–1312)	66.38	0.03 (0.00–0.11)	0.06 (0.01–0.20)	1.76 (1.47–2.05) ^a
	50–74 years	36,470 (8267–111,191)	88,993 (19,472–239,427)	144.02	21.07 (4.78–64.24)	19.54 (4.28–52.57)	–0.09 (–0.32–0.14)
	≥75 years	85,708 (20,776–234,089)	496,622 (124,875–1,185,154)	479.43	440.68 (106.82–1203.60)	640.85 (161.14–1529.34)	1.32 (1.08–1.57) ^a
DALY	All ages	2,813,268 (1,276,794–6,411,019)	11,646,909 (5,719,528–22,514,747)	314.00	238.59 (108.28–543.70)	814.10 (399.79–1573.75)	3.92 (3.79–4.05) ^a
	15–49 years	29,006 (14,177–53,153)	51,705 (25,398–89,556)	78.25	4.34 (2.12–7.95)	8.01 (3.93–13.87)	1.97 (1.72–2.22) ^a
	50–74 years	1,304,117 (606,716–3,093,902)	3,552,144 (1,760,535–7,105,274)	172.38	753.49 (350.55–1787.59)	779.98 (386.58–1560.17)	0.17 (–0.05–0.39)
	≥75 years	1,480,145 (647,044–3,290,619)	8,043,060 (3,732,787–16,021,870)	443.40	7610.34 (3326.85–16919.11)	10378.91 (4816.86–20674.91)	1.03 (0.85–1.20) ^a
Incidence	All ages	755,780 (637,685–854,211)	3,476,644 (2,998,723–3,884,422)	360.01	64.10 (54.08–72.44)	243.01 (209.61–271.52)	4.14 (3.99–4.29) ^a
	15–49 years	31,265 (21,765–41,805)	55,221 (39,719–72,454)	76.62	4.68 (3.26–6.25)	8.55 (6.15–11.22)	1.90 (1.69–2.11) ^a
	50–74 years	340,105 (279,784–418,837)	1,271,708 (1,033,726–1,546,066)	273.92	196.51 (161.65–242.00)	279.24 (226.99–339.48)	1.11 (0.95–1.27) ^a
	≥75 years	384,410 (302,822–482,070)	2,149,715 (1,802,110–2,553,325)	459.22	1976.49 (1557.00–2478.62)	2774.03 (2325.48–3294.86)	1.06 (0.95–1.18) ^a
Prevalence	All ages	4,352,315 (3,633,202–5,012,549)	20,299,243 (17,341,557–23,309,840)	366.40	369.11 (308.12–425.10)	1418.89 (1212.15–1629.32)	4.18 (4.02–4.35) ^a
	15–49 years	104,139 (65,485–150,404)	192,464 (121,707–277,120)	84.82	15.57 (9.79–22.49)	29.81 (18.85–42.92)	2.03 (1.88–2.18) ^a
	50–74 years	2,129,515 (1,771,023–2,506,415)	7,266,114 (6,161,748–8,521,135)	241.21	1230.39 (1023.26–1448.15)	1595.49 (1352.99–1871.07)	0.81 (0.66–0.97) ^a
	≥75 years	2,118,661 (1,689,790–2,569,056)	12,840,665 (10,497,187–15,311,086)	506.07	10893.35 (8688.25–13209.10)	16569.83 (13545.76–19757.70)	1.30 (1.13–1.47) ^a

AAPC: average annual percentage change; CI: confidence interval; UI: uncertainty interval

Rates are per 100,000 population.

^a Indicates that the 95% CI for the AAPC excludes 0.



DALY: disability-adjusted life year
The figure shows absolute numbers of DALYs, deaths, incident cases, and prevalent cases among both sexes, females and males.

Fig. 2. Sex-specific number trends in Alzheimer's disease and other dementias burden in China, 1990–2023.

all outcomes, including 496,622 deaths, 8,043,060 DALYs, 2,149,715 incident cases, and 12,840,665 prevalent cases.

For mortality, AAPCs were 1.76 (95% CI 1.47–2.05) in ages 15–49 years, -0.09 (-0.32 to 0.14) in ages 50–74 years, and 1.32 (1.08–1.57) in ages ≥ 75 years. For DALY rates, AAPCs were 1.97 (1.72–2.22), 0.17 (-0.05 to 0.39), and 1.03 (0.85–1.20), respectively. Incidence rates increased across all age groups, with AAPCs of 1.90 (1.69–2.11), 1.11 (0.95–1.27), and 1.06 (0.95–1.18). Prevalence rates also increased across age groups, with AAPCs of 2.03 (1.88–2.18), 0.81 (0.66–0.97), and 1.30 (1.13–1.47), respectively (Table 2, Fig. 1).

Risk-attributable burden profile

In 2023, modelled attributable burden varied across the selected GBD risk factors (Table 3, Fig. 3). Attributable deaths were 190,422 for ambient particulate matter pollution, 103,694 for high fasting plasma glucose, 28,725 for smoking, and 27,759 for household air pollution from solid fuels. The corresponding attributable DALYs were 3,794,481, 2,058,837, 687,815, and 559,148. These estimates should be interpreted as modelled population-level attributable burden among the selected risks rather than as direct causal effects.

From 1990 to 2023, age-standardised attributable death trends differed by risk. AAPCs were 4.29 (95% CI 4.04–4.55) for ambient particulate matter pollution, 0.82 (0.57–1.06) for high fasting plasma glucose, -0.14 (-0.45–0.17) for smoking, and -5.47 (-5.94 to -4.99) for household air pollution from solid fuels. For age-standardised attributable DALY rates, AAPCs were 4.41 (4.24–4.58), 0.74 (0.59–0.89), -0.05 (-0.28–0.19), and -5.41 (-5.87 to -4.96), respectively (Table 3, Fig. 3).

DISCUSSION

This GBD 2023-based analysis describes long-run ADOD burden dynamics in China from 1990 to 2023 using a national health-estimates framework.^{14,15} Across the full period, absolute burden increased across deaths, DALYs, incidence, and prevalence. In contrast, ASR increased more modestly. This contrast is consistent with demographic ageing and population structure change contributing importantly to rising dementia-related service needs, although the present analysis did not formally decompose the increase into population growth, population ageing, and age-specific rate change.^{4,7,19,23}

A central pattern is the divergence between absolute growth and age-standardised growth. Death counts increased markedly over the study window, whereas the corresponding ASMR changed only modestly;

similar contrasts were observed for DALYs, incidence, and prevalence. This pattern is compatible with previous global and China-focused dementia burden studies showing that demographic momentum can substantially increase the number of people affected even when age-adjusted rates change slowly.^{4,7,19,23} Therefore, count-based and rate-based indicators answer different but complementary policy questions.

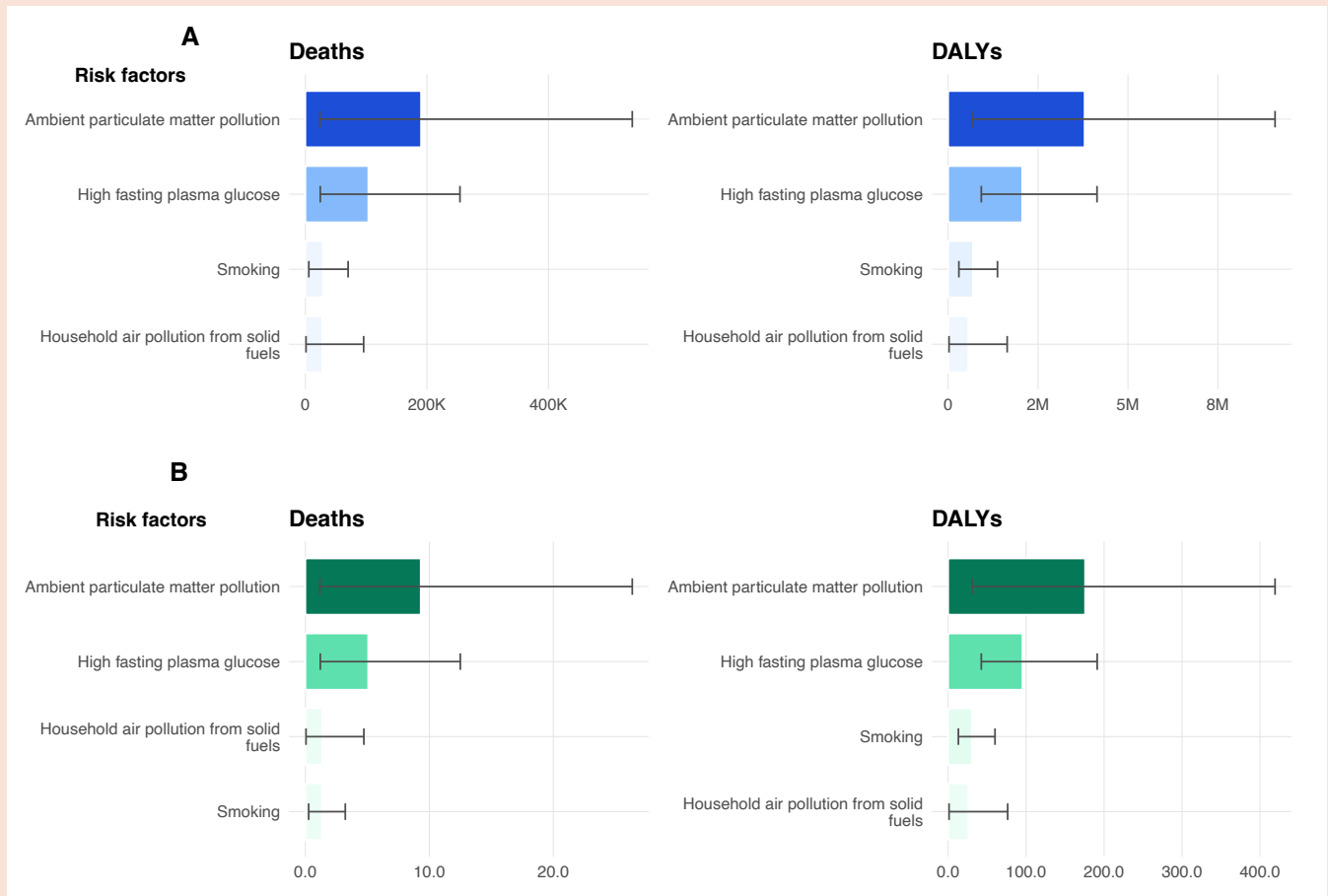
For clinicians, payers, and governments, absolute burden remains the operational denominator for system demand. Memory clinic utilisation, hospital contact volume, community follow-up, long-term care demand, and family caregiving needs are driven primarily by the number of people living with dementia and related disability. Age-standardisation is necessary for temporal and cross-population comparison, but it does not replace count-based planning in a rapidly ageing population.^{5,10}

Sex-specific results show 2 coexisting features. Females had higher levels across major burden indicators in 2023, including both counts and ASR. This is consistent with prior literature indicating that sex and gender differences in ADOD burden may reflect longer female survival, life-course biological factors, and social determinants.^{8,9} However, trend pace differed by endpoint in this dataset, with faster male increases in mortality- and DALY-related standardised trends and slightly faster female increases in incidence- and prevalence-related standardised trends. This pattern indicates that burden level and rate of change do not necessarily move in parallel within sex strata. It also suggests that sex contrasts in ADOD should be evaluated by outcome dimension rather than treated as a single summary effect.

Age-specific trajectories indicate persistent concentration of severe burden in advanced age. The ≥ 75 group contributed the largest numbers of deaths and DALYs in 2023, while younger adult groups carried a smaller but non-zero share of growth across non-fatal outcomes. Near-flat patterns in some middle-age mortality and DALY trends, combined with continued increases in the oldest group, are compatible with an age-gradient structure in which neurodegeneration, frailty accumulation, multimorbidity, and competing-risk dynamics become more prominent late in life.^{10–12,24} Under this pattern, care-system pressure is expected to remain concentrated in very old adults, with implications for continuity of late-life cognitive care, referral pathways, caregiver support, and community-based long-term management.

The risk-attributable component should be interpreted cautiously. Ambient particulate matter pollution had the largest modelled attributable burden among the selected GBD risks in 2023 and showed the steepest increase in age-standardised attributable

Fig. 3. Modelled burden of Alzheimer's disease and other dementias attributable to selected GBD risk factors in China, 2023. (A) Number of attributable deaths and DALYs; (B) age-standardised attributable death and DALY rates per 100,000 population. These estimates are population-level modelled attributable burdens and should not be interpreted as individual-level causal effects.



ASR: age-standardised rate; DALY: disability-adjusted life year

rates. However, these findings reflect GBD comparative risk assessment assumptions and should not be read as direct proof of individual-level causality or proportional preventability.²² Baseline differences between risks may also affect comparisons of relative growth.^{25,26} High fasting plasma glucose also contributed a substantial and increasing modelled burden, which is biologically plausible given previous evidence linking diabetes and chronic hyperglycaemia to cognitive decline and dementia through vascular and metabolic pathways.^{27,28} Smoking-attributable ASRs were broadly stable, while household air pollution from solid fuels declined. These patterns support prevention-oriented surveillance but require cautious translation into policy.²⁹

Practical implications for China and ageing Asian health systems

The findings have several practical implications. First, dementia surveillance should routinely report both

absolute numbers and ASR, because these indicators inform different decisions. Second, primary care and memory-clinic pathways may need expansion to accommodate rising case numbers, particularly among very old adults. Third, community-based long-term care, caregiver support, and payer planning should be strengthened because much of the health loss is concentrated in later life. Fourth, urban–rural and provincial heterogeneity should be assessed in future subnational analyses to avoid masking inequities in diagnosis, access, and long-term care capacity.⁶ Finally, air-quality improvement, diabetes prevention, and tobacco control remain relevant public health priorities, but the present ecological estimates should be used as supportive surveillance evidence rather than causal proof of individual benefit.^{10,22,24}

The 2020–2023 interval overlapped with the COVID-19 pandemic, when disruption of chronic care pathways, delayed diagnosis, and interruption of supportive services were reported in many

Table 3. Risk-attributable burden of Alzheimer's disease and other dementias in China, 1990 and 2023.

MEASURE	Risk factors	Number (1990, 95% UI)	Number (2023, 95% UI)	Percentage change (%)	ASR (1990, 95% UI)	ASR (2023, 95% UI)	AAPC of ASR (95% CI)
Deaths	Ambient particulate matter pollution	10,421 (1241–36,701)	190,422 (23,908–538,041)	1727.24	2.38 (0.29–8.56)	9.30 (1.16–26.36)	4.29 (4.04–4.55) ^a
	High fasting plasma glucose	18,169 (3795–56,479)	103,694 (24,527–254,439)	470.71	4.13 (0.92–12.82)	5.07 (1.20–12.49)	0.82 (0.57–1.06) ^a
	Smoking	7874 (1610–22,863)	28,725 (5701–70,345)	264.81	1.45 (0.33–4.16)	1.33 (0.26–3.22)	-0.14 (-0.45–0.17)
	Household air pollution from solid fuels	40,518 (4953–122,771)	27,759 (1055–95,909)	-31.49	9.23 (1.09–28.83)	1.36 (0.05–4.72)	-5.47 (-5.94 to -4.99) ^a
	Ambient particulate matter pollution	240,041 (31,883–700,695)	3,794,481 (662,388–9,091,509)	1480.76	43.32 (5.68–126.19)	175.68 (30.76–419.35)	4.41 (4.24–4.58) ^a
DALYs	High fasting plasma glucose	425,507 (162,928–1,119,120)	2,058,837 (921,647–4,139,497)	383.86	76.07 (29.87–199.73)	95.29 (42.50–191.14)	0.74 (0.59–0.89) ^a
	Smoking	213,582 (88,244–498,180)	687,815 (297,893–1,371,736)	222.04	31.80 (13.19–74.79)	30.27 (13.10–60.09)	-0.05 (-0.28–0.19)
	Household air pollution from solid fuels	931,579 (171,286–2,457,493)	559,148 (24,278–1,641,389)	-39.98	168.67 (32.37–442.65)	25.89 (1.14–76.31)	-5.41 (-5.87 to -4.96) ^a

AAPC: average annual percentage change; ASR: age-standardised rate; CI: confidence interval; DALY: disability-adjusted life year; UI: uncertainty interval
 Rates are per 100,000 population.

^a Indicates that the 95% CI for the AAPC excludes 0.

settings.^{16,30,31} However, the present study did not formally estimate pandemic-specific effects or compare pre-pandemic and pandemic-era slopes. The COVID-19 period is therefore discussed as a contextual limitation rather than as an explanatory factor for the observed long-term trends.

Overall, the findings are consistent with previous global and China-focused evidence showing rising dementia burden in absolute terms under demographic pressure, while adding an updated GBD 2023 national profile that jointly tracks fatal burden, non-fatal burden, sex and age patterns, and selected risk-attributable estimates.^{4,7,14,15,19,23} This integrated view may support dementia surveillance, ageing-adapted service planning, and cautious prevention-priority setting.

Several methodological features support interpretability, including the long observation window, the use of internally consistent GBD definitions, and the parallel evaluation of multiple burden indicators. Nevertheless, limitations remain. The estimates are modelled rather than derived from complete nationwide adjudicated registry data. Changes in awareness, diagnostic criteria, coding practice, death-certification quality, and data availability may have influenced long-term trends.¹⁹ ADOD subtype granularity was unavailable, limiting disease-specific interpretation. Risk-attributable estimates are ecological and model-based, so ecological fallacy is possible when translating findings to individuals.²² The study did not formally decompose burden growth into population growth, ageing, and age-specific rates, did not perform age-period-cohort modelling, and did not resolve provincial, urban-rural or socioeconomic heterogeneity. Finally, the broad age groups used here may obscure clinically important differences among older adults, particularly those aged 75–84 and ≥85 years.

CONCLUSION

From 1990 to 2023, China experienced large increases in the absolute burden of ADOD, whereas age-standardised increases were smaller. This pattern is consistent with demographic ageing contributing importantly to rising dementia-related service needs. Burden remained higher in females and was concentrated among older adults, especially those aged ≥75 years. Among selected GBD risk factors, ambient particulate matter pollution and high fasting plasma glucose showed substantial modelled attributable burden, while household air pollution from solid fuels declined. These findings support strengthened dementia surveillance, ageing-adapted care capacity, and cautious integration of risk-factor prevention into public-health planning.

Supplementary materials

Appendix S1. Data extraction filters, derived datasets, and analysis code.

Ethics statement

This study used de-identified, aggregated, and publicly available estimates from the GBD Results Tool. The extraction filters, derived datasets, and analysis code are provided as supplementary files to improve reproducibility. No individual-level data were analysed. Thus, no additional ethics approval or informed consent was required.

Declaration

The authors declare that they have no affiliations or financial involvement with any commercial organisation with a direct financial interest in the subject or materials discussed in the manuscript. There is no funding or conflict of interest to declare.

Acknowledgements

The authors thank the Core Facility and Bioinformatics Laboratory of The First Hospital of Jilin University for technical training and methodological support.

REFERENCES

1. Hao M, Chen J. Trend analysis and future predictions of global burden of alzheimer's disease and other dementias: a study based on the global burden of disease database from 1990 to 2021. *BMC Med* 2025;23:378.
2. Li Z, Yang N, He L, et al. Global Burden of Dementia Death from 1990 to 2019, with Projections to 2050: An Analysis of 2019 Global Burden of Disease Study. *J Prev Alzheimers Dis* 2024;11:1013-21.
3. Guo J, Wang P, Gong J, et al. The disease burden, risk factors and future predictions of Alzheimer's disease and other types of dementia in Asia from 1990 to 2021. *J Prev Alzheimers Dis* 2025;12:100122.
4. Li R, Qi J, Yang Y, et al. Disease Burden and Attributable Risk Factors of Alzheimer's Disease and Dementia in China from 1990 to 2019. *J Prev Alzheimers Dis* 2022;9:306-14.
5. Jia L, Du Y, Chu L, et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. *Lancet Public Health* 2020;5:e661-71.
6. Zhou M, Wang H, Zeng X, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019;394:1145-58. Erratum in: *Lancet* 2020;396:26.
7. Liu Y, Wu Y, Chen Y, et al. Projection for dementia burden in China to 2050: a macro-simulation study by scenarios of dementia incidence trends. *Lancet Reg Health West Pac* 2024;50:101158.
8. Nebel RA, Aggarwal NT, Barnes LL, et al. Understanding the impact of sex and gender in Alzheimer's disease: A call to action. *Alzheimers Dement* 2018;14:1171-83.
9. Ferretti MT, Iulita MF, Cavedo E, et al. Sex differences in Alzheimer disease - the gateway to precision medicine. *Nat Rev Neurol* 2018;14:457-69.

10. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020;396:413-Erratum in: *Lancet* 2023;402:1132.
11. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol* 2011;10: 819-28.
12. Norton S, Matthews FE, Barnes DE, et al. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol* 2014;13:788-94.
13. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018;14:535-62.
14. GBD 2023 Causes of Death Collaborators. Global burden of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet* 2025;406: 1811-72.
15. GBD 2023 Disease and Injury and Risk Factor Collaborators. Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet* 2025;406:1873-922.
16. Mok VCT, Pendlebury S, Wong A, et al. Tackling challenges in care of Alzheimer's disease and other dementias amid the COVID-19 pandemic, now and in the future. *Alzheimers Dement* 2020;16:1571-81.
17. Stevens GA, Alkema L, Black RE, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *Lancet* 2016;388:e19-23.
18. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573-7.
19. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019;18:88-106.
20. Kim HJ, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med* 2000; 19:335-51.
21. Clegg LX, Hankey BF, Tiwari R, et al. Estimating average annual per cent change in trend analysis. *Stat Med* 2009;28: 3670-82.
22. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;396:1223-49.
23. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* 2022;7:e105-25.
24. Livingston G, Huntley J, Liu KY, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet* 2024;404:572-628.
25. Peters R, Ee N, Peters J, et al. Air Pollution and Dementia: A Systematic Review. *J Alzheimers Dis* 2019;70:S145-63.
26. Wilker EH, Osman M, Weisskopf MG. Ambient air pollution and clinical dementia: systematic review and meta-analysis. *BMJ* 2023;381:e071620.
27. Biessels GJ, Despa F. Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications. *Nat Rev Endocrinol* 2018;14:591-604.
28. Gudala K, Bansal D, Schifano F, et al. Diabetes mellitus and risk of dementia: A meta-analysis of prospective observational studies. *J Diabetes Investig* 2013;4:640-50.
29. Zhong G, Wang Y, Zhang Y, et al. Smoking is associated with an increased risk of dementia: a meta-analysis of prospective cohort studies with investigation of potential effect modifiers. *PLoS One* 2015;10:e0118333. Erratum in: *PLoS One* 2015;1: e0126169.
30. Wang H, Li T, Barbarino P, et al. Dementia care during COVID-19. *Lancet* 2020;395:1190-1.
31. Hazan J, Liu KY, Isaacs JD, et al. Has COVID-19 affected dementia diagnosis rates in England? *Int J Geriatr Psychiatry* 2023;38:e5976.

Artificial intelligence tools in dermatology education: A scoping review on their application, efficacy, and limitations

Chun Hian John Yap¹ MRCP, Ziying Vanessa Lim² MPH

ABSTRACT

Introduction: Artificial Intelligence (AI) is increasingly explored for medical education, and dermatology, with its visual diagnostic focus, holds promise for AI-enhanced learning. However, evidence on its educational effectiveness remains limited and fragmented. This scoping review aimed to assess the evidence base for applications and limitations of AI-based educational tools in dermatology.

Methods: A systematic search of PubMed, Embase, Web of Science, Scopus, and PsycINFO up to July 2025 was conducted. Data were synthesised narratively, considering types of AI interventions and their assessed outcomes. These were analysed with the Cost, Usability, Credibility, Fairness, Accountability, Transparency, Explainability (CUC-FATE) framework. Study quality was assessed with ROBINS-I and a COSMIN-informed checklist.

Results: A total of 827 records were screened, with 360 duplicates removed, yielding 467 studies. Six full-length studies and 1 conference abstract met inclusion criteria, mostly from 2023 to 2025. These explored AI-generated clinical images, Large Language Model-generated vignettes, intelligent tutoring systems, and clinical decision support tools. Content validation studies generally reported favourable ratings for accuracy, clarity, and educational utility, while intervention studies suggested possible benefits for learning performance and diagnostic accuracy. Usability and credibility were commonly assessed, whereas cost, accountability, fairness, transparency, and explainability were rarely examined.

Conclusions: Most of the studies reviewed had high risk of bias, small sample sizes, and limited methodological rigour, with significant heterogeneity in examined educational outcomes limiting synthesis. While current initial studies on AI hold promise, this scoping review underscores the need for more robust studies with standardised evaluation frameworks, prioritising ethical principles such as fairness, explainability, and accountability for safe integration in training.

Keywords: artificial intelligence, dermatology, large language models, medical education, scoping review

CLINICAL IMPACT

What is New

- This scoping review focuses on the emerging evidence of the direct educational impact of AI-based educational tools in dermatology and identifies major gaps in methodological quality, standardised evaluation, and ethical oversight.
- As the use of AI in healthcare and medical education expands, these findings provide a preview of opportunities and challenges in AI-based dermatology training.

Clinical Implications

- The findings support the need for rigorous and ethically grounded frameworks to guide safe integration of AI into dermatology education, with emphasis on educational effectiveness, fairness, explainability, and accountability.

INTRODUCTION

Artificial Intelligence (AI) technologies—particularly deep learning algorithms and Large Language Model

(LLM) computer systems—have progressed at an unprecedented pace in recent years, driving innovation across a wide range of healthcare applications.¹ These advancements have resulted in a corresponding surge in research on the application of AI in medical education,² including the development of virtual reality simulators³ and intelligent tutoring systems.⁴ Such tools enable the provision of individualised, innovative and adaptive learning experiences, allowing students and physicians to practise medical reasoning.^{5,6} Moreover, integrating AI into medical education allows for distance learning, making medical education more accessible in resource-constrained areas with limited resources⁷ or time due to clinical demands.⁸

In the field of dermatology, its inherent visual diagnostic process makes it an attractive possibility for AI to be applied as an educational tool,⁹ with similar promise being noted in other visually intensive specialties, such as radiology, for the enhancement of clinical competencies and optimising learning outcomes.¹⁰ AI applications—ranging from machine learning-based diagnostic simulators to image-based

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Ministry of Health Holdings, Singapore

² National Skin Centre, Singapore

Correspondence: Dr Ziying Vanessa Lim, National Skin Centre, 1 Mandalay Road, Singapore 308205. Email: drvanessalim@gmail.com

Accepted: 30 June 2026

Published online first: 14 July 2026

learning platforms¹¹—have been proposed to improve diagnostic accuracy, engagement, and learning efficiency.¹² However, the evidence evaluating the effectiveness of these technologies in dermatology education—such as in assessing improvement in diagnostic accuracy, knowledge retention, clinical decision-making—remains fragmented and under-explored, with robust evidence for their use in dermatology education currently lacking.¹³

As such, this scoping review aims to address this gap, examining the current methods by which AI has been applied to dermatological education. By charting the existing research landscape and evaluating current evidence, this scoping review aims to highlight promising applications and identify limitations in existing studies. In doing so, this study aims to suggest future innovations for more efficacious AI usage, including areas of learner and faculty development necessary to integrate AI into dermatological education.

METHODS

The protocol detailing the search strategies of this review was formulated by the authors, who registered this on the International Prospective Register of Systematic Reviews (PROSPERO, registration number CRD420251058835) on 22 May 2025 before the study's initiation. While initially designed as a systematic review, in view of the small number of highly heterogeneous studies identified, the study was redesigned as a scoping review after article identification to provide a broader map of existing evidence. With a narrow evidence base, assessment on efficacy would be limited, and hence research questions were refined to define the applications and limitations of the selected studies.

Search strategy

The search included the databases of MEDLINE (through PubMed), Embase, Web of Science, SCOPUS, and PsycINFO, with the search strategies described in Supplementary Table S1. All studies published from the inception of the respective database till 1 July 2025 were included. Based on titles and abstracts, duplicate studies were first excluded. The remaining studies were then evaluated based on titles and abstracts for their relevance to the use of artificial intelligence in dermatological education.

Selection of articles

Studies published in any language, from all countries, were considered for review. Randomised controlled trials, non-randomised controlled studies, pre-post intervention studies, and observational studies were included if they referred to direct application of AI in dermatological education. Efforts were made to

obtain full-length studies for findings highlighted in conference abstracts. Studies were considered at various levels of healthcare education, including that of medical students, residents (especially dermatology and general medical trainees), and non-dermatology clinicians involved in dermatology learning. Tools used solely for clinical diagnosis without an educational aim, editorials, opinion pieces, protocols without results, and studies focused solely on AI tool development without educational evaluation were excluded. As title and abstract screening was done jointly, formal recording of reviewer disagreements and calculation of inter-rater reliability were not applicable.

Data collection

Data were extracted via a standardised collection form, with authors' names, publication date, country of publication (as defined by the setting of data collection), type of study, and level of healthcare education of the intended learner audience. Following this, categorisation was performed of the types of artificial intelligence educational tools assessed, any specific learning outcomes examined and overall conclusions.

Data analysis

In this review, the articles identified displayed significant heterogeneity, preventing meaningful pooling and synthesis of educational outcomes. As such, a scoping review was conducted according to the methodological framework proposed by Arksey and O'Malley¹⁴ and refined by Levac et al.,¹⁵ and was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist (Supplementary Table S2).¹⁶ With the extracted data, a narrative account of findings was produced, considering the types of AI interventions and the educational outcomes assessed. The limitations and gaps in these studies were also identified.

The studies were categorised based on various aspects of AI technologies specifically measured by each study. Utilising an approach developed by Quttainah et al.¹⁷ examining enablers for the use of LLMs in medical education, an assessment was made on whether the parameters of Cost, Usability, Credibility, Fairness, Accountability, Transparency, and Explainability (CUC-FATE) were considered by each included study.

Quality assessment

Only non-randomised studies were obtained, and their methodological quality was assessed. Study quality of quasi-experimental studies was assessed using the Cochrane Collaboration's Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I)

tool,¹⁸ modified for usage in before-after studies. As no established appraisal tool exists for validation of AI-based educational content, content validation studies were assessed based on a framework devised by the authors, informed by the CONsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist for content validity.¹⁹ This framework examined core tenets of the COSMIN checklist, including relevance, comprehensiveness and clarity, as well as recommendations regarding expert involvement and qualitative evaluation during instrument development.

There are also currently no suitable frameworks for assessing the quality of AI-powered interventions in medical education. The commonly used Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) criteria, while modified recently to account for machine learning methods,²⁰ work to evaluate predictive models by their performance, and cannot be directly applied to effectiveness in medical education. As such, a framework previously utilised by Bates et al.,²¹ also referenced in other reviews on artificial intelligence in healthcare education,²² was utilised.

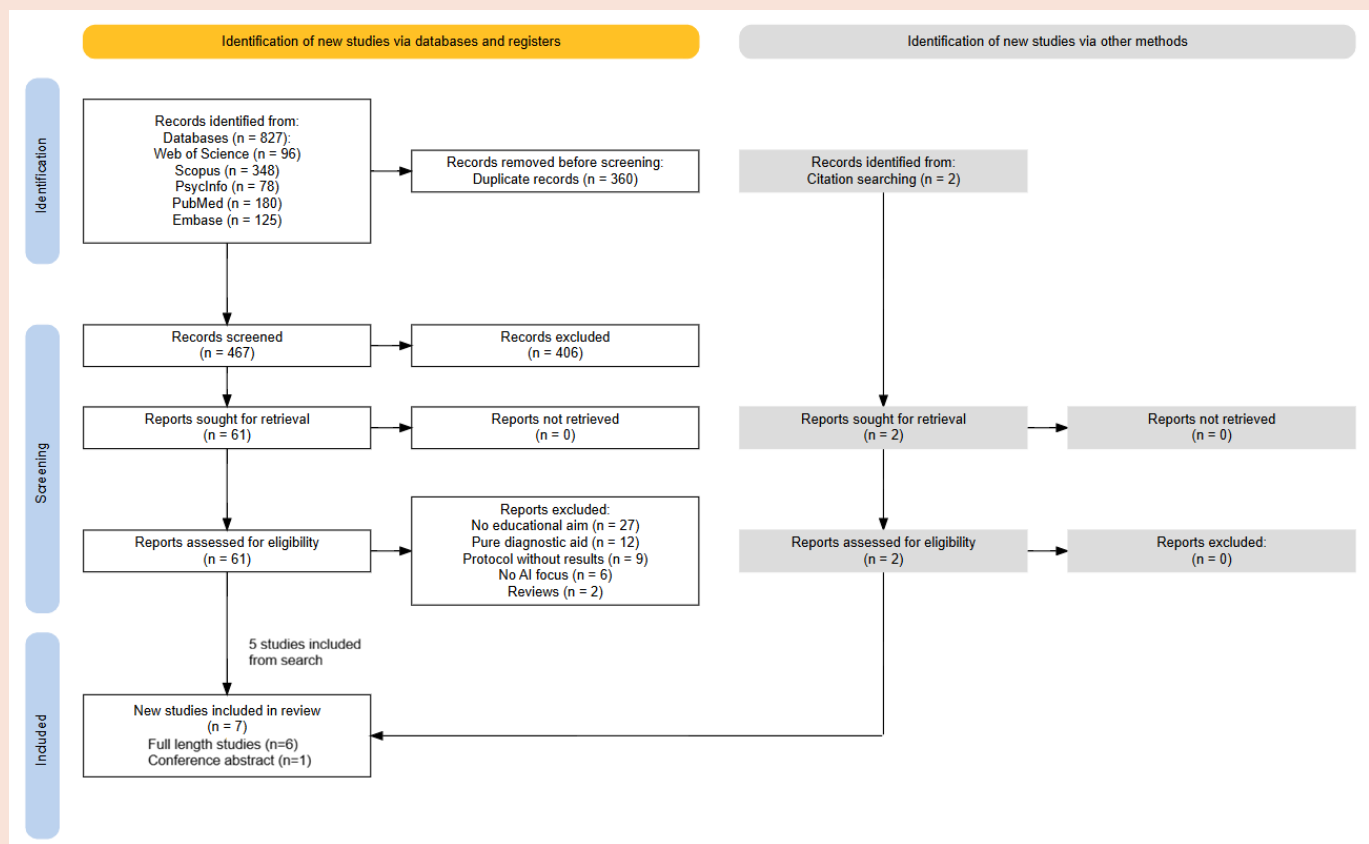
RESULTS

Search results

The comprehensive search of PubMed, Embase, Web of Science, Scopus, and PsycINFO yielded 827 studies, of which 360 duplicates were excluded (Fig. 1). From the remaining 467 studies, 406 were excluded based on titles and abstracts. In total, 61 studies were retrieved and assessed for eligibility, of which four full-length studies and one conference abstract were identified. Further search of citations yielded two other relevant articles,²³⁻²⁴ with a total of 6 full-length studies and 1 conference abstract²³⁻²⁹ selected for review.

A decision was made to include a conference abstract²⁷ to allow comprehensive mapping of the dermatological educational landscape, in accordance with a proposal by Scherer et al.³⁰ An attempt was made to contact the authors and examine study registers and published protocols to obtain further information on methods and results, but without success. However, due to its brevity, there was limited ability to appraise the abstract for its design, risk of bias, and assessed educational outcomes, which is noted in the subse-

Fig. 1. Database search.



quent discussion. The PRISMA flowchart for this study was created with the aid of an app designed by Haddaway et al.³¹

Description of included studies

The general characteristics of studies are summarised in Table 1, with specific outcomes examined summarised in Table 2 and further laid out in a frequency table for content validation studies (Supplementary Table S3). Out of the 6 identified studies and 1 conference abstract (henceforth considered with the rest of the studies), 6 of these were conducted over the past 3 years (2023–2025)^{23–28}—with 1 study focusing on an intelligent tutoring system in 2008.²⁹ Four of the studies were conducted in the US,^{24,27–29} and 1 each in Italy, Australia, and Germany.^{23,25,26}

Three of the studies (including the conference abstract) were assessed as quasi-experimental, defined here as studies analysing the effect of interventions without the assignment of experimental and control groups^{26–27,29}; while the remaining 4 studies^{23–25,28} were designed as cross-sectional studies for content validation, defined here as studies which focused on rating the quality and accuracy of AI output. Three studies had a stated targeted audience who were residents,^{25,28,29} three studies had a stated targeted audience who were medical students,^{24,26,27} and one study general practitioners.²³

AI technologies examined

Many studies focused on the utilisation of generative AI, including 2 for image generation^{25,26} and 3 for text generation.^{23,24,28} All studies examining text generation utilised ChatGPT in its various models, with 2 studies^{24,28} attempting to generate questions for use in clinical examinations. The last study²³ attempted to assess the LLM's ability to answer questions about acne vulgaris for education and was unique in assessing the LLM's ability to identify appropriate reference literature for learners.

Out of the 2 remaining studies, the first evaluated an intelligent tutoring system in dermatopathology serving as a precursor to modern AI today.²⁹ The second, a conference abstract, provided a different approach, utilising a clinical decision support system for medical students in an outpatient setting, creating a list of differential diagnoses during patient review.²⁷

Educational outcomes examined

Both full-length quasi-experimental studies^{26,29} attempted to examine the improvement of clinical competence, with 1 image-generation study²⁶ looking at the improvement in learning accuracy (albeit in a small sample of students, which cannot be interpreted as evidence of educational benefit), while another

study on intelligent tutoring²⁹ looked at the frequency of errors. The conference abstract²⁷ evaluating a clinical decision support system appeared to only look at patient perception and utilisation time of the AI tool, although the absence of a full-text limits assessment. The significant heterogeneity of these educational outcomes prevented further synthesis and comparison.

All studies designed for content validation attempted to break down the characteristics of AI-generated output important for educational use. The 3 studies focused on LLMs^{23,24,28} looked at accuracy and clarity of material generated, though they diverged significantly in aspects they found important for education. The content validation study for AI-generated images²⁵ looked similarly at accuracy and utility for clinical education.

Quality assessment

The 3 quasi-experimental trials (including the conference abstract) were analysed with the ROBINS-I tool with all studies assessed having a serious overall risk of bias (Table 3). None of the trials provided data on the participant recruitment or measures to control for confounders. In the conference abstract,²⁷ while there was limited information to assess for the risk of bias, the authors only recruited two medical students to test the intervention.

In assessing content validation studies on AI-generated materials,^{23–25,28} a content validity checklist informed by the COSMIN approach was adopted¹⁹ (Table 4). All studies had clear descriptions of measured constructs and assessed relevant (but different) characteristics of AI-generated materials. They also all used clear Likert scales for assessment, involving experts in its development and reporting on inter-rater agreement. Three out of 4 studies were significantly lacking in having a small number of individuals assessing the suitability of teaching material, and only 1 out of 4 studies reported qualitative insights.

When analysed using the approach by Bates et al.²¹ (Supplementary Table S4), the accuracy of the AI-based educational technologies was often unvalidated, and information with regard to uncertainty and data-handling was not provided in all studies.

Enablers of AI in medical education

Studies were assessed for examination of the key features of the CUC-FATE framework¹⁷ (Table 5) for effective and safe usage of AI in medical education, but many of these elements were not assessed in identified studies. While all studies examined the usability of AI technology, and 6 out of 7 studies examined their credibility, none of the studies assessed issues regarding cost.

Table 1. Description of included studies.

Reference/ country of study	Study design/ intended learner audience	AI technology assessed	Sample size/number of assessors	Study overview	Outcomes assessed
Cevik et al. (2025) ²⁵ Australia	Content validation study Residents	AI image generation	Images assessed by 3 doctors, namely, 2 plastic surgery residents and 1 specialist plastic surgeon.	Three prominent AI models were assessed on their ability to generate authentic looking images of skin cancers (squamous cell carcinoma, basal cell carcinoma and melanoma) as educational resources. Images underwent an accuracy check using AI Dermatologist's machine learning algorithm.	Adequacy and accuracy in representing real-world pathology Utility in clinical education
Naldi et al. (2024) ²³ Italy	Content validation study General practitioners	LLM	The answers and references were independently evaluated by a panel of 5 doctors, including 4 residents and 1 dermatologist	Open AI GPT-4 was instructed to generate information for a CME on the treatment of acne vulgaris. A 23-item questionnaire was developed by the authors from National Institute of Clinical Excellence (NICE) guidelines and prompted into the LLM. These were evaluated via a Likert scale, with room for qualitative judgement. The LLM was also instructed to quote 3 references from biomedical literature to support the answers provided. After these were validated, the references were evaluated by the same 5 doctors via a Likert scale, with room for qualitative judgement.	Questions were assessed on 5 domains: quality, readability, accuracy, thoroughness, and consistency with guidelines References were assessed on 3 domains: pertinence, relevance, and whether they were updated
Pielage et al. (2025) ²⁶ Germany	Quasi-experimental study Medical students	AI image generation	93 students were involved in the 3D simulation. The small user study involved 7 students for comparative analysis	AI Diffusion models were used to generate images of malignant melanoma and benign nevi. These lesions were embedded into a 3D model of a virtual patient and integrated into a fully interactive virtual 3D simulation for a group of students. A small user study was also conducted with seven students, with a learning group of four students provided with AI-generated images of nevi and melanoma, and a control group of three students not provided with these images. They were then tasked to classify 20 selected real images of nevi and melanoma.	Perceived competence of students after training Realism of lesions Improvement in learning accuracy
Rao et al. (2025) ²⁴ US	Content validation study Medical students	LLM	Assessed by 3 practising attending physicians	The Open AI GPT 4.0 model was tasked to generate 20 clinical vignettes in dermatology for medical student education, and these were assessed on a Likert scale	Alignment to scientific consensus Comprehensiveness Risk of clinical harm Demographic bias Clinical quality

Table 1. Description of included studies. (Cont'd)

Reference/ country of study	Study design/ intended learner audience	AI technology assessed	Sample size/number of assessors	Study overview	Outcomes assessed
Lee et al. (2024) ²⁷ US	Quasi-experimental study (Conference Abstract) Medical students	Clinical decision support system	Two medical studies were asked to evaluate the decision support system	The medical students evaluated 80 patients using the machine learning clinical decision support application in a dermatology clinic without an attending, with the utilisation time of the application to generate a list of differential diagnoses being assessed. Patients were also surveyed regarding their perceptions of technology during this visit.	Time of utilisation of application Perception of utilisation of system in medical education
Ayub et al. (2023) ²⁸ US	Content validation study with qualitative elements Residents (preparation for board examinations)	LLM	The questions were independently assessed by two board-certified dermatologists	CME articles from the Journal of the American Academy of Dermatology (JAAD) were uploaded into Open AI GPT 3.5, and the LLM was instructed to generate multiple choice questions.	Accuracy Complexity of questions Clarity of wording and structure
Payne et al. (2009) ²⁹ US	Quasi-experimental study Residents	Intelligent tutoring	44 pathology residents were involved in the intelligent tutoring system	Pathology residents used Slide Tutor—a Medical Intelligent Tutoring System in dermatopathology using commonly used diagnostic algorithms. This prototype provided feedback including error analysis and confirmation of correct actions, providing students with context-specific hints based on the current problem state.	Frequency of errors Analysis of goals differences and solution path

CME: continuing medical education; LLM: large language model

Table 2. Results of included studies.

Content validation studies		
Reference	Descriptive statistics	Conclusions
Naldi et al. (2024) ²³	The 23 answers were evaluated on: - Quality (101/115 or 87.8%) - Readability (109/115 or 94.8%) - Accuracy (87/115 or 75.7%) - Thoroughness (98/115 or 85.2%) - Consistency (73/95^a or 76.8%) The 69 references provided were evaluated on: - Pertinence (246/260 or 94.6%) - Relevance (237/260 or 91.2%) - Updatedness (162/260 or 62.3%)	ChatGPT demonstrated potential as a resource for dermatology continued medical education, producing appropriate responses in terms of quality and thoroughness, and high readability. However, responses were sometimes inaccurate or inconsistent with guidelines, indicating areas for improvement. The references provided by AI were positively evaluated in pertinence and relevance, but issues in currency suggest gaps in the availability of papers ChatGPT had been trained.
Rao et al. (2025) ²⁴	The questions reflected high levels of: - Alignment with scientific consensus (4.45, 95% CI 4.28–4.62) - Comprehensiveness (4.3, 95% CI: 4.11–4.89^b) - Overall quality (4.28, 95% CI: 4.10–4.47) The questions reflected low levels of: - Clinical harm (1.6, 95% CI: 1.38–1.81) - Demographic bias (1.52, 95% CI: 1.31–1.72)	Mean ratings of assessors reflected high alignment with scientific consensus, comprehensiveness and overall quality, with low suspicion of clinical harm and demographic bias. Errors relating to factual accuracy, including potential AI hallucinations, were indirectly captured in ratings for scientific alignment and clinical harm.
Cevik et al. (2025) ²⁵	Initial ratings for DALL-E had the authors rate on the following: - Resemblance to real world pathology (3/5) - Adequate representation (4/5) - Easily visible and discernible details (4/5) - Quality (4/5) - Beneficial for educational purposes (4/5)	Only DALL-E had the ability to generate distinctive features associated with skin cancer, while images created by Midjourney and BlueWillow had exaggerated fictionalised features.
Ayub et al. (2023) ²⁸	Out of the questions: 10/40 (25%) were low complexity 9/40 (22.5%) were vague or unclear 5/40 (12.5%) were inaccurate 16/40 (40%) were accurate and at an appropriate level of complexity	The model was assessed to have a limited domain-specific knowledge, with a limited ability to understand context and generate high-quality distractor options
Quasi-experimental studies		
Reference	Descriptive statistics	Conclusions
Pielage et al. (2025) ²⁶	- Skin lesions were assessed as mostly realistic (average of 6.17/10) - Subjective competence of the students increased (from 2.28 to 3.12/5) - Out of the 20 images assessed, the 4 students in the learning group (exposed to AI generated images) achieved a mean accuracy of 0.64, as compared to 0.56 for the 3 students in the control group.	The skin lesions were assessed as mostly realistic and the subjective competence of the students increased after completing the dermatology scenario in VR in the study. In the small group, a small increase in learning accuracy was observed after learning from AI-generated images, suggesting potential for medical education
Lee et al. (2024) ²⁷	89% of patients supported the use of such technology in medical student education - Average utilisation time of the application was 67.0 seconds	Such machine learning clinical decision support technology has the potential to actively engage students during their clinical rotations without being disruptive to clinic flow. Overall, patients supported the use of the technology.
Payne et al. (2009) ²⁹	- A decrease in errors over time was noted from approximately 50% to 20% - An increase in Correct-BNS (optimal solution path) actions over time was noted from approximately 20% to 50%	During the tutoring, the frequency of errors gradually declined. While student performance frequently differed from the tutor-defined optimal path, students approached the optimal solution path as they continued to be tutored.

AI: artificial intelligence; BNS: best-next-step; VR: virtual reality

^a Only 95 out of the 115 items were deemed eligible for assessment of consistency by the authors.

^b This confidence is not symmetric about the point estimate with a much higher upper bound (4.89), which may be a possible error in the original manuscript.

Table 3. Cochrane risk-of-bias assessment tool for non-randomised studies.

Bias domains	Pielage et al. (2025)²⁶	Lee et al.^a (2024)²⁷	Payne et al. (2009)²⁹
Pre-intervention			
1.1 Bias due to confounding	PY	PY	PY
1.2 Bias in selection of participants into the study	PY	PY	PY
Risk-of-bias judgment	Serious	Critical	Serious
At intervention			
2.1 Bias in classification of interventions	PY	PY	PY
Risk-of-bias judgment	Serious	Serious	Serious
Post-intervention			
3.1 Bias due to deviations from intended interventions	PN	NI	PN
3.2 Bias due to missing data	NI	NI	NI
3.3 Bias in measurement of outcomes	PY	PY	PY
3.4 Bias in selection of the reported result	PY	PY	PN
Risk-of-bias judgment	Serious	Serious	Moderate
Overall bias	Serious	Serious	Serious

NI: no information; PN: probably no; PY: probably yes

^a Assessment for this study was limited as a conference abstract

Table 4. COSMIN-inspired assessment for content validation studies.

COSMIN-inspired concept	Criterion	Cevik et al. (2025)²⁵	Naldi et al. (2024)²³	Rao et al. (2025)²⁴	Ayub et al. (2023)²⁸
Relevance/ construct definition	Clear description of construct being measured	Met	Met	Met	Met
Item relevance	Relevance of generated items to construct	Met	Met	Met	Met
Content coverage	Comprehensiveness of construct	Unmet —only a few images of skin cancers	Unmet —only used for acne vulgaris	Partially met —randomly selected vignettes under USMLE	Partially met —utilised CME articles from JAAD
Comprehensibility	Comprehensibility of rating scale	Met	Met	Met	Met
Expert review	Involvement of experts in development	Met	Met	Met	Met
Reliability evidence	Inter-rater agreement reported	Met	Met	Met	Met
Methodological quality of expert evaluation	Number of experts appropriate	Not met —only 3 individuals, with poor rigor	Met —5 individuals with 4 residents, 1 specialist	Not met —only 3 individuals	Not met —only 2 individuals
Qualitative validation evidence	Qualitative insights reported	Not reported	Reported	Not reported	Reported

CME: continuing medical education; JAAD: Journal of the American Academy of Dermatology; USMLE: United States Medical Licensing Examination

Table 5. Assessment of studies using the CUC-FATE framework.

Study	Cost	Usability	Credibility	Fairness	Accountability	Transparency	Explainability
Cevik et al. (2025) ²⁵	NR	R	R	NR	NR	NR	NR
Naldi et al. (2024) ²³	NR	R	R	NR	R	NR	R
Pielage et al. (2025) ²⁶	NR	R	R	NR	NR	R	R
Rao et al. (2025) ²⁴	NR	R	R	R	NR	NR	NR
Lee et al. ^a (2024) ²⁷	NR	R	NR	NR	NR	NR	NR
Ayub et al. (2023) ²⁸	NR	R	R	NR	NR	NR	NR
Payne et al. (2009) ²⁹	NR	R	R	NR	NR	R	R
No. of studies examining	0	7	6	1	1	2	3

NR: not reported; R: reported

^a Assessment for this study was limited as a conference abstract.

Two of the studies^{26,29} reported AI interventions created by the authors, allowing assessment of their transparency, but studies that utilised proprietary LLMs, such as the content validation studies, lacked the ability to do so. Author-created AI interventions in these studies also allowed assessment of explainability. Including a study²³ which asked the LLM to generate reference literature for provided information, a total of 3 studies specifically assessed explainability.

This same study also highlighted the presence of inaccuracies and errors in the LLM's responses, urging the necessity of expert review to prevent misinformation, thus being the only study assessing accountability.²³ Only 1 study that looked at AI-generated questions considered fairness, specifically assessing the risk of demographic bias in AI-generated questions and concluding that the risk of this is low.²⁴

DISCUSSION

This scoping review has identified a small but emerging body of literature examining the application of AI in dermatological education. Though the included studies examined a range of AI-based interventions, such as AI-generated images of dermatological conditions and LLM-generated educational content, the overall evidence base remains limited. Most studies were preliminary, small-scale and heterogeneous in design, and current evidence is insufficient in supporting conclusions regarding its effect on improving learner performance and clinical competence.

The review also highlights several questions that are still unanswered. It remains unclear which AI-based interventions are the most effective for different groups, such as medical students, residents, and general practitioners, and whether these interventions necessarily provide meaningful added value over

existing traditional educational methods. Important concerns on implementation, such as transparency and explainability, are also inconsistently addressed in studies, and represent critical concerns for educators.

Since completion of the search by this review, a broader scoping review of AI and dermatological education has been published by Lau et al.,³² also highlighting educational applications, including the use of LLMs as well as generative and adaptive imaging tools. In contrast, this present review focused on studies examining direct educational applications, deliberately excluding the substantial number of studies solely assessing AI performance in answering dermatology examination questions. By adopting this narrower educational focus, this review provides a more targeted appraisal of the current evidence for the use of AI in dermatological education, concluding that despite growing interest in the field, current evidence is sparse and limits strong conclusions.

Nevertheless, this field continues to evolve rapidly. For example, a study published after the search period examining the usage of LLMs as a dermatology case narrator for medical students³³ would have been eligible for this review. As this study was published after the prespecified search period, it was not included in the formal synthesis, but its relevance suggests that future updated reviews may identify a larger body of eligible evidence.

The findings of this review are also consistent with emerging evidence from the broader medical education literature, with a rapidly increasing number of AI-related publications over the past few years.⁶ A scoping review of AI medical image generation across health professions education (including dermatology) by Gupta et al.³⁴ also concluded that generative AI should be treated as an experimental adjunct requiring

rigorous human review. Similarly, in radiology education (a similarly visually intensive specialty), a scoping review of AI-based educational approaches by Hui et al.³⁵ identified a range of possible applications, from personalised curriculum generation and diagnostic support tools to automated evaluation systems.

While similar themes were observed in the present review, the evidence base identified for dermatology was considerably smaller, with fewer studies evaluating educational outcomes and learner performance. Together, these reviews suggest that while dermatology shares characteristics with more visual specialties—making it potentially well-suited for AI-enhanced learning, robust evidence demonstrating educational effectiveness is lacking, making further research with rigorous validation necessary to identify the most effective AI tools in education.³⁶ Other potential interventions not included in this review, such as personalised learning platforms,³⁷ integrating AI with virtual-reality based training for skin cancer screening³⁸ as well as AI-assisted learning for knowledge and milestone assessment,³⁹ have been proposed by other authors, and would benefit from implementation studies to assess educational outcomes.

Key limitations

There was considerable heterogeneity noted in study design and evaluation metrics for educational efficacy in quasi-experimental studies, with a diversity of intervention types explored. These studies were also small-scale with low sample sizes (ranging from 2 to 44), with few participants and inadequate control for confounding variables causing serious risk of bias. None of the studies employed validated instruments to assess learning outcomes, often relying on self-developed Likert scales or informal expert ratings. Similarly, none of the content validation studies used standardised measures to assess content validity of AI-generated information. This lack of standardisation prevents meaningful synthesis.

None of the studies examined compared AI-based interventions with traditional teaching methods. This absence of a control group limited conclusions about added educational value and generalisability, and such comparative analysis is essential in evaluating AI-based interventions' educational effectiveness.⁷

All studies were conducted in pilot or isolated research settings, outside the realm of formal medical school or residency training curricula, suggesting institutional adoption of AI in dermatology education is currently limited. This is consistent with existing research, which suggests that adoption of AI into medical education is still in its early developmental stages.⁴⁰ A key challenge is the alignment of existing educational frameworks and curricula with rapidly evolving AI

capacities,⁴¹ with educational centres often struggling with infrastructure and policy support required.⁴²

A further limitation relates to the review process itself, where study screening was conducted through consensus between 2 reviewers rather than independent duplicate reviews. While this collaborative approach allowed for discussion and agreement at each stage, it may have increased susceptibility to shared reviewer bias.

CUC-FATE framework

This scoping review utilised the CUC-FATE framework to examine crucial enablers for AI in medical education, demonstrating that most dimensions were unassessed by included studies. Credibility and utility were frequently assessed, and while an earlier content validation study²⁸ using an older LLM noted high levels of inaccuracy, this was not noted in newer LLM based content validation studies^{23,24} utilising newer models of the same LLM. This underscores how LLMs are rapidly evolving and improving, and older studies with less positive results may be less relevant with newer generative AI models.

Most studies lacked reporting of assessments of transparency and explainability. Content validity is limited by the inability to trace presented evidence to its origins, which is especially pertinent when proprietary technologies, such as the LLMs ChatGPT or Gemini, are used.^{43,44} These deficiencies may undermine clinicians' trust in AI-assisted dermatological education, particularly when verification of accuracy in clinical contexts is essential.⁴⁵

Encouragingly, an included image-generation study²⁶ described the creation of a custom diffusion model, providing detailed methodological transparency. While other studies focusing on deep learning algorithms and clinical decision support systems for diagnosis of melanoma^{46,47} and other skin diseases⁴⁸ were excluded from review because of the absence of educational focus, their transparent designs would allow effective adaptation for educational purposes.

Ethical concepts of accountability and fairness were rarely assessed in included studies. With the use of AI in dermatological education rapidly expanding, these omissions are concerning, given that the identification of ethical challenges is paramount for responsible development and implementation.⁴⁹ The issue of accountability, establishing liability for harmful results sustained from incorrect educational content,⁵⁰ was assessed in only 1 study²³ which emphasised the need for human oversight to ensure clinical accuracy. For fairness, AI algorithmic bias can lead to error in diagnosis of under-represented populations such as individuals of colour,⁵¹ causing erroneous education of medical professionals. This was briefly assessed in 1

study on AI-generated dermatological examination questions,²⁴ but should be more widely evaluated.

Implications for practice and future research

Despite promising innovation in the application of AI in dermatological education, this review has demonstrated that significant gaps remain, highlighting key areas for future development. Current quasi-experimental studies had a high rate of bias, emphasising the need for larger, robust, multicentre studies involving diverse participant cohorts in institutional contexts for both medical student and healthcare professional training in dermatology.

Significant methodological heterogeneity across studies severely limited comparison, highlighting the need for standardised reporting frameworks and verified evaluation tools to ensure replicability. This review has demonstrated a framework for content validation studies devised by the authors and informed by the COSMIN approach, and the CUC-FATE framework, which incorporates not just practical but important ethical considerations. Adopting such standardised frameworks would improve the quality and transparency of AI-related educational research in dermatology, allowing appropriate and effective implementation. There is a need for studies to explicitly incorporate ethical principles in their design, ensuring AI-based educational tools are equitably and responsibly implemented.

Singapore has emerged as an early regional adopter of AI in healthcare training, resulting in an increased impetus to equip faculty and students with the necessary foundational knowledge in AI in a forward-looking medical education landscape.⁵² In this aspect, efforts have been made to design curricula locally to integrate AI into medical education,⁵³ with local medical educators highlighting the benefits in equipping medical learners with skills in digital technology.⁵⁴ By using dermatology as a targeted case study, this review provides timely evidence that can inform curriculum design for AI-enhanced medical education in Singapore and similar healthcare systems.

Given the diversity of AI interventions, we recommend that further research involve learners in their design, which can improve engagement and reduce barriers to adoption. Multiple studies have consistently demonstrated that dermatology specialists,⁵⁵ trainees,⁵⁶ and students,^{40,57} while agreeing that AI will play a key role in dermatological education, expressed that they lacked training to utilise it effectively. A mixed-methods approach, combining quantitative outcomes of efficacy with qualitative data of learner experience, would be useful in examining these perceptions to enhance usage of AI in dermatological education.

Beyond the design of learner-facing interventions, successful implementation of AI in dermatological education will also need investment in faculty development. Recent medical education literature has emphasised the importance of equipping educators with AI literacy through structured faculty development models, which can include competency-based certification and interdisciplinary training.^{58,59} As AI becomes increasingly integrated into educational practice, frameworks for faculty development have been devised in broader medical education contexts,⁶⁰ although these have not been explored specifically in dermatology. Future work should therefore examine not only the effectiveness of AI interventions for learners, but also the competencies and support required for educators to implement these tools safely and effectively.

CONCLUSIONS

This review suggests that AI is an evolving tool for dermatology education, with potential benefits for learners. Existing literature is characterised by methodological heterogeneity and limited validation, and there is insufficient evidence to determine whether AI-based educational tools improve learning outcomes compared with standard educational approaches. Establishing standardised frameworks for the development, validation, and evaluation of AI-based educational technologies is crucial to generate higher-quality evidence, allowing for effective and ethical implementation in dermatology training.

Supplementary materials

Supplementary Table S1. Search strategy.

Supplementary Table S2. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.

Supplementary Table S3. Frequency table of outcomes of included content validation studies.

Supplementary Table S4. Assessment of studies using the framework proposed by Bates et al.

Ethics statement

This research did not involve the collection of patient data, human participants, or biological samples. All analyses were conducted on publicly available, published literature.

Declaration

The authors declare they have no affiliations or financial involvement with any commercial organisation with a direct financial interest in the subject or materials discussed in the manuscript.

REFERENCES

- Nazi ZA, Peng W. Large language models in healthcare and medical domain: a review. *Informatics* 2024;11:57.
- Mir MM, Mir GM, Raina NT, et al. Application of artificial intelligence in medical education: current scenario and future perspectives. *J Adv Med Educ Prof* 2023;11:133-40.
- Winkler-Schwartz A, Bissonnette V, Mirchi N, et al. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *J Surg Educ* 2019;76:1681-90.
- Fazlollahi AM, Bakhaidar M, Alsayegh A, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. *JAMA Netw Open* 2022;5:e2149008.
- Wartman SA, Combs CD. Reimagining medical education in the age of AI. *AMA J Ethics* 2019;21:E146-52.
- Zarei M, Mamaghani HE, Abbasi A, et al. Application of artificial intelligence in medical education: a review of benefits, challenges, and solutions. *Med Clin Pract* 2024;7:100422.
- Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. *JMIR Med Educ* 2019;5:e13930.
- Chen L, Chen P, Lin Z. Artificial intelligence in education: a review. *IEEE Access* 2020;8:75264-78.
- Hogarty DT, Su JC, Phan K, et al. Artificial intelligence in dermatology: where we are and the way to the future: a review. *Am J Clin Dermatol* 2020;21:41-7.
- Li R, Liu G, Zhang M, et al. Integration of artificial intelligence in radiology education: a requirements survey and recommendations from faculty radiologists, residents, and medical students. *BMC Med Educ* 2025;25:380.
- Liopyris K, Gregoriou S, Dias J, et al. Artificial intelligence in dermatology: challenges and perspectives. *Dermatol Ther* 2022;12:2637-51.
- Perrin J, Petronic-Rosic V. The potential role and restrictions of artificial intelligence in medical school dermatology education. *Clin Dermatol* 2024;42:477-9.
- Nahm WJ, Sohail N, Burshtein J, et al. Artificial intelligence in dermatology: a comprehensive review of approved applications, clinical implementation, and future directions. *Int J Dermatol* 2025;64:1568-83.
- Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19-32.
- Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010;5:69.
- Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews: checklist and explanation. *Ann Intern Med* 2018;169:467-73.
- Quttainah M, Mishra V, Madakam S, et al. Cost, usability, credibility, fairness, accountability, transparency, and explainability framework for safe and effective large language models in medical education. *JMIR AI* 2024;3:e51834.
- Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159-70.
- Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378.
- Bates DW, Auerbach A, Schulam P, et al. Reporting and implementing interventions involving machine learning and artificial intelligence. *Ann Intern Med* 2020;172:S137-44.
- Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Med Educ* 2025;25:129.
- Naldi L, Bettoli V, Santoro E, et al. Application of ChatGPT as a content generation tool in continuing medical education: acne as a test topic. *Dermatol Rep* 2024;17:10138.
- Rao AS, Kim J, Mu A, et al. Synthetic medical education in dermatology leveraging generative artificial intelligence. *NPJ Digit Med* 2025;8:247.
- Cevik J, Seth I, Rozen WM. Artificial intelligence-generated images in skin cancer education: an experimental study. *Australas J Plast Surg* 2025;8:118181.
- Pielage L, Schmidle P, Marschall B, et al. Interactive high-quality skin lesion generation using diffusion models for VR-based dermatological education. In: *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 2025:878-97.
- Lee G, Lopez D, Singh P, et al. Machine learning technology in medical student education. Conference abstract, *J Invest Dermatol* 2024;144:S59.
- Ayub I, Hamann D, Hamann CR, et al. Exploring the potential and limitations of ChatGPT in generating board-style dermatology questions: a qualitative analysis. *Cureus* 2023;15:e45934.
- Payne VL, Medvedeva O, Legowski E, et al. Effect of a limited-enforcement intelligent tutoring system in dermatopathology on student errors, goals, and solution paths. *Artif Intell Med* 2009;47:175-97.
- Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? *Syst Rev* 2019;8:264.
- Haddaway NR, Page MJ, Pritchard CC, et al. PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams. *Campbell Syst Rev* 2022;18:e1230.
- Lau LD, Tran V, Chapman W, et al. Applications of machine learning, natural language processing, and generative artificial intelligence in dermatology education and research: a scoping review. *Int J Dermatol* 2026;65:1600-10.
- Karampinis E, Bozi Tzetzis DA, Pappa G, et al. Use of a large language model as a dermatology case narrator: exploring the dynamics of a chatbot as an educational tool in dermatology. *JMIR Dermatol* 2025;8:e72058.
- Gupta K, Latinovich MF, Latinovich MF, et al. Generative artificial intelligence for medical image creation in health professions education: a scoping review. *J Med Syst* 2026;50:19.
- Hui ML, Sacoransky E, Chung A, et al. Exploring the integration of artificial intelligence in radiology education: a scoping review. *Curr Probl Diagn Radiol* 2025;54:332-8.
- Tozsin A, Ucmak H, Soyuturk S, et al. The role of artificial intelligence in medical education: a systematic review. *Surg Innov* 2024; 31:415-23.
- Gordon M, Daniel M, Ajiboye A, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Med Teach* 2024;46:446-70.
- Mergen M, Will L, Graf N, et al. Feasibility study on virtual reality-based training for skin cancer screening: bridging the gap in dermatological education. *Educ Inf Technol* 2025; 30:5251-82.
- Ramachandran V, Jairath N, Cheraghloou S, et al. Revolutionizing dermatology residency: artificial intelligence for knowledge and clinical milestones assessment. *Clin Exp Dermatol* 2024; 49:732-3.

40. Lee J, Wu AS, Li D, et al. Artificial intelligence in undergraduate medical education: a scoping review. *Acad Med* 2021;96:S62-70.
41. Rampton V, Mittelman M, Goldhahn J. Implications of artificial intelligence for medical education. *Lancet Digit Health* 2020;2:e111-2.
42. Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med* 2018;1:54.
43. Boscardin CK, Gin B, Golde PB, et al. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. *Acad Med* 2024;99:22-7.
44. Goktas P, Grzybowski A. Assessing the impact of ChatGPT in dermatology: a comprehensive rapid review. *J Clin Med* 2024;13:5909.
45. Zbrzezny AM, Krzywicki T. Artificial intelligence in dermatology: methods, clinical applications, and perspectives. *Appl Sci* 2025;15:7856.
46. Maron RC, Utikal JS, Hekler A, et al. Artificial intelligence and its effect on dermatologists' accuracy in dermoscopic melanoma classification. *J Med Internet Res* 2020;22:e18091.
47. Díaz-Ramón JL, Gardeazabal J, Izu RM, et al. Melanoma clinical decision support system: an artificial intelligence-based tool. *Cancers (Basel)* 2023;15:2174.
48. Han SS, Park I, Chang SE, et al. Augmented intelligence dermatology. *J Invest Dermatol* 2020;140:1753-61.
49. Gordon ER, Trager MH, Kontos D, et al. Ethical considerations for artificial intelligence in dermatology. *Br J Dermatol* 2024;190:789-97.
50. Ferreira AL, Lipoff JB. The complex ethics of applying ChatGPT in dermatology. *J Am Acad Dermatol* 2023;89:e157-8.
51. Rinderknecht FA, Nwandu L, Daneshjou R, et al. The impact of artificial intelligence on health equity in dermatology. *Curr Dermatol Rep* 2024;13:148-55.
52. Tan JA, Ng IKS, Hong A, et al. Empowering the next generation: integrating artificial intelligence education into medical training. *Postgrad Med J* 2026;102:185-8.
53. Cai C, Jong M, Ng YY, et al. Developing a postgraduate program for AI in medicine with Kern's six-step curriculum development approach in Singapore. *Proc AAAI Conf Artif Intell* 2025;39:28979-87.
54. Zainal H, Xin X, Thumboo J, et al. Medical school curriculum in the digital age: perspectives of clinical educators and teachers. *BMC Med Educ* 2022;22:428.
55. Patrzyk S, Bielecki W, Woźniacka A. Attitudes among Polish dermatologists regarding modern technologies. *Adv Dermatol Allergol* 2022;39:531-7.
56. Polesie S, Gillstedt M, Kittler H, et al. Attitudes towards artificial intelligence within dermatology. *Br J Dermatol* 2020;183:159-61.
57. Cho SI, Han B, Hur K, et al. Perceptions and attitudes of medical students regarding artificial intelligence in dermatology. *J Eur Acad Dermatol Venereol* 2021;35:e1-5.
58. Ahsan Z. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. *BMC Med Educ* 2025;25:1187.
59. Sánchez-Mendiola M, Anakin M, Findyartini A, et al. Applied insights for using generative artificial intelligence in faculty development in health professions education. *MedEdPublish* 2025;15:279.
60. Association of American Medical Colleges. CGEA Artificial Intelligence Framework for Educators. Washington (DC): Association of American Medical Colleges; 2024. <https://www.aamc.org/about-us/mission-areas/medical-education/advancing-ai-resource-collection/cgea-artificial-intelligence-framework-educators>. Accessed 7 June 2026.

Interpreting multivariable regression coefficients in observational clinical research: The Table 2 fallacy

John Rong Hao Tay^{1,2} FAMS, Nasir Z Bashir³ MSc, Upul Cooray^{4,5} PhD

ABSTRACT

Multivariable regression tables are common in observational clinical research, but their coefficients are often over-interpreted. A model built to estimate the effect of 1 exposure may also report coefficients for age, sex, comorbidities, behaviours, and other adjustment variables. These additional rows are frequently read as independent risk factors, even when the analysis was not designed to estimate their effects. This is the Table 2 fallacy. The problem is not the use of adjustment, but the interpretation of adjustment terms as if each were a separate causal estimate. In this commentary, we use a directed acyclic graph and a single worked example to show why the coefficient for the exposure of interest can answer the intended clinical question, while coefficients for adjustment variables may not represent clinically actionable effects. We also show how similar errors arise when interaction terms are interpreted as causal. Authors should specify the target estimand (the causal effect the analysis is designed to estimate) and exposure, choose adjustment variables from the assumed causal structure, interpret only the exposure coefficient, and fit a separate model for each further question. We set out red flags, recurring pitfalls, and questions to ask of any risk-factor table. A regression table is not a menu of modifiable risks. An association is clinically actionable only when the study was designed to support that interpretation.

Keywords: causal inference, confounding, directed acyclic graphs, effect modification, measures of effect, risk factor

INTRODUCTION

An adjusted regression table can make a modelling decision look like a clinical finding. In many observational studies, the row of interest is the exposure—the comparison the model was built to estimate. The other rows often list age, sex, comorbidities, behaviours, and other variables included to control confounding. Because all rows are displayed in the same format, they are easily read as equivalent clinical effects. That reading is often wrong.

Statistical results have long been vulnerable to over-interpretation. Over the past decade, sustained efforts have improved how clinical researchers report statistical findings. The American Statistical Association has cautioned against treating a *P* value threshold as the basis for scientific conclusions,^{1,2} and others have urged that statistical significance not be treated as a statement of truth.³ Several journals now discourage, or decline to publish, *P* values in favour of effect estimates and confidence intervals (CIs),^{4–6} while estimation graphics have been promoted to make both the estimate and its uncertainty visible.⁷ A similar case for CIs was made in Singapore in 2010.⁸

These reforms are important, but they leave a more basic problem untouched—an estimated coefficient may be accompanied by a CI and still be given the wrong causal interpretation. In clinical research, observational studies often aim to answer causal questions using multivariable regression models. An effect estimate and CI for the exposure of interest are reported, alongside coefficients for variables included for adjustment. The issue is not adjustment itself, but a habit that remains common in observational clinical research—reading every coefficient in the adjustment model as if it were a causal effect. Presenting the coefficient for every adjustment variable within a table, and interpreting each as an independent effect estimate, is referred to as the Table 2 fallacy, named for the table in which it characteristically appears.⁹

The fallacy arises because a multivariable model is usually intended to answer a specific clinical question. For example, a study may seek to estimate the effect of rheumatoid arthritis on cardiovascular risk while adjusting for age, smoking, hypertension, and diabetes. The adjustment variables are included to obtain a less biased estimate of the effect of rheumatoid arthritis; they are not necessarily included to estimate their own effects on cardiovascular risk. This can be understood using a directed acyclic graph, a diagram in which arrows represent assumed causal

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Health Services Research and Population Health Programme, Duke-NUS Medical School, Singapore

² Department of Restorative Dentistry, National Dental Centre Singapore, Singapore

³ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

⁴ University of Sydney School of Dentistry, Sydney, Australia

⁵ Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Australia

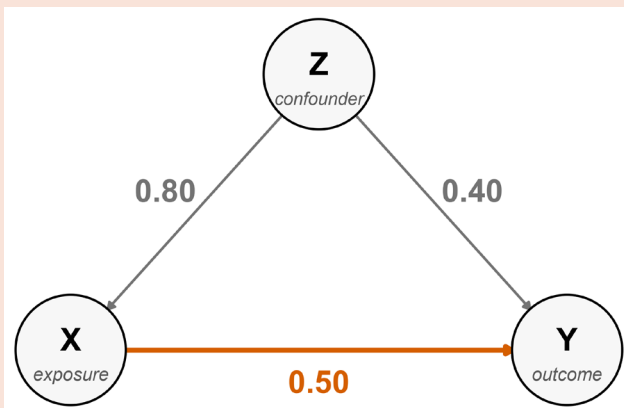
Correspondence: Dr John Rong Hao Tay, Health Services and Systems Research Programme, Duke-NUS Medical School, 8 College Road, Singapore 169857. Email: john.tay.r.h@singhealth.com.sg

Accepted: 17 July 2026

Published Online First: 23 July 2026

relationships between variables (Fig. 1).¹⁰ In such a framework, the total effect of an exposure is its overall influence on the outcome through all causal pathways. Some of these pathways may pass through intermediate variables, known as mediators (steps on the causal path from exposure to outcome). To estimate the exposure's effect without confounding, investigators adjust for variables that influence both the exposure and the outcome.¹⁰ The adjustment set is chosen to answer a single causal question: the effect of the exposure on the outcome. Consequently, although the adjustment variables' coefficients appear in the same regression table, they are adjustment terms, not separate causal estimates.

Fig. 1. Directed acyclic graph for the worked example.



The exposure X is rheumatoid arthritis, the outcome Y is cardiovascular risk, and the confounder Z is smoking. Smoking causes both rheumatoid arthritis and cardiovascular risk, while rheumatoid arthritis also causes cardiovascular risk; Z is therefore adjusted for because it confounds the X – Y relationship. Arrows are labelled with the data-generating coefficients used in the simulation: $Z \rightarrow X = 0.8$, the direct $Z \rightarrow Y$ effect = 0.4, and the target $X \rightarrow Y$ effect = 0.5 (highlighted). Data were generated from a linear model with standard normal errors (Supplementary Appendix S1).

Consider the same example, simplified to a single confounder: a model for the effect of rheumatoid arthritis on cardiovascular risk, adjusting for smoking, a cause of both (Fig. 1). In plain terms, smoking is included to help estimate the effect of rheumatoid arthritis, not to answer a separate question about smoking itself. In this model, the coefficient for rheumatoid arthritis can be read as an estimate of its total effect on cardiovascular risk in the simplified causal diagram, because there is no mediator between rheumatoid arthritis and cardiovascular risk. The coefficient for smoking is different. Although it may appear to represent smoking's effect on cardiovascular risk, it actually represents smoking's effect after accounting for rheumatoid arthritis. In other words, it estimates only the part of smoking's effect that is not explained by its influence on rheumatoid arthritis.

If smoking increases cardiovascular risk partly by increasing rheumatoid arthritis, then the adjusted smoking coefficient understates smoking's total effect.

Estimating smoking's total effect is a separate causal question. To answer it, rheumatoid arthritis should not be included in the adjustment set, because it lies on the causal pathway from smoking to cardiovascular risk. The model would instead need to adjust for confounders of smoking itself.¹¹ In this simplified example, there are none, so a model containing smoking alone would be sufficient, but only because smoking is assumed to have no confounders here, not as a general rule. A different question, such as separating smoking's direct effect from the part acting through rheumatoid arthritis, would require methods designed for mediation. The fitted model and estimates are provided in the Supplementary Appendix S1.

The same reasoning applies when a table suggests that the exposure effect differs between groups. Suppose a study asks whether rheumatoid arthritis raises the risk of an adverse cardiovascular outcome, adjusts for smoking, and includes a rheumatoid arthritis $\times$ smoking interaction term to allow the rheumatoid arthritis effect to differ between smokers and non-smokers. The results table suggests the association is stronger in smokers than in non-smokers. It is tempting to read this causally: that smoking worsens the effect of rheumatoid arthritis, and that smoking cessation would reduce the excess risk. Strictly, however, this table has only asked whether the rheumatoid arthritis effect differs across smoking groups. That is effect modification—how the effect of 1 exposure varies across levels of another variable. It is not the same as estimating the joint causal effect of both rheumatoid arthritis and smoking. To answer that second question, smoking would need to be treated as a second exposure, with its own causal question and sufficient adjustment set.¹² The difference between smoking groups is represented by an interaction term, but that term is still part of the model, and on its own, it describes effect modification, not a causal interaction. It should not be read causally unless the causal question and the adjustment set required to answer it have been specified. Interpreting it causally is therefore another manifestation of the Table 2 fallacy, even when the main estimate for rheumatoid arthritis is valid.

The problem becomes clearer with unmeasured factors. Smokers may differ from non-smokers in ways that are hard to capture completely, including diet, physical activity, and access to preventive care. These factors may influence both smoking and cardiovascular risk. Although adjusting for smoking may suit the rheumatoid arthritis question, smoker and non-smoker comparisons are still made within the structure the

adjustment model creates. Apparent differences between these groups may therefore reflect not only genuine effect modification but also residual confounding or modelling artefacts. In the extreme, the entire apparent difference between smokers and non-smokers could reflect such unmeasured factors rather than any effect smoking cessation would alter; the coefficients for smoking, age, sex, and their interactions remain quantities the analysis was never designed to interpret causally.

The error is not remedied by a change of wording. Calling these estimates “independent associations” still invites each row to be read as a distinct effect, and terms such as “potential risk factor” and “predisposing factor” further imply an actionable, quasi-causal relationship.^{13,14} The causal interpretation then often reappears in the discussion, through mechanistic explanations or recommendations for intervention. A disclaimer of “association only” relocates the fallacy rather than resolving it.

Several steps help to avoid the Table 2 fallacy. First, define the purpose of the analysis. Observational data can be used to describe patterns, build prediction

models, or answer causal questions. If the aim is description, report descriptively. If the aim is prediction, evaluate prediction performance rather than interpreting coefficients as intervention effects. If the aim is causal inference, specify 1 exposure, 1 outcome, and 1 target estimand. The exposure estimand is interpretable as causal only if the causal question is well defined and the adjustment set is sufficient; it must control confounding without adjusting for mediators or colliders of the exposure–outcome relationship. The remaining variables are used for adjustment, not to be read as a menu of risk factors. If another variable seems interesting in its own right, treat it as a distinct question, with its own causal diagram, its own sufficient adjustment set, and its own model. A well-presented table makes this explicit (Table 1): the exposure is set apart and interpreted, while the adjusted variables are grouped and flagged as included for confounding control, not as separate causal effects.

Second, pre-specify the adjustment set from the assumed causal structure, rather than selecting adjustment variables by stepwise procedures or statistical significance. These procedures may be

Table 1. Adjusted risk ratios from a hypothetical, illustrative multivariable model for incident cardiovascular events, with each coefficient grouped by its causal role. The rheumatoid arthritis exposure is represented here as graded inflammatory disease activity, to illustrate a multi-level exposure.

Variable	Adjusted RR ^d	95% CI
Exposure—interpretable as the target effect^a		
No inflammatory disease	1.00	—
Low disease activity	1.20	0.95–1.52
Moderate disease activity	1.45	1.18–1.78
High disease activity	1.80	1.40–2.31
Adjustment variable—coefficients not causally interpretable^{b,c}		
Female sex	0.93	0.82–1.05
Age (per 10 years)	1.55	1.43–1.68
Current smoking	1.40	1.22–1.61
Lower income	1.25	1.08–1.45
Diabetes mellitus	1.50	1.30–1.73
Hypertension	1.38	1.20–1.59

CI: confidence interval; RR: risk ratio

^a The exposure block is the causal contrast the model was designed to estimate, using a pre-specified adjustment set derived from the causal diagram; it is interpretable only when that adjustment set is sufficient and contains no mediator or collider of the exposure. Its category contrasts form part of the same exposure definition and should be interpreted together.

^b The adjustment variables are included for confounding control, not because their coefficients are target effects. Each variable has its own causal structure and would require its own adjustment set before its coefficient could be interpreted causally. Treating these coefficients as independent risk factors is the Table 2 fallacy.

^c Adjustment set: sex, age, smoking, lower income, diabetes mellitus, hypertension.

^d Risk ratios are collapsible: when there is no confounding and no effect modification on the risk ratio scale, the marginal risk ratio equals the stratum-specific risk ratio. For non-collapsible measures such as the odds ratio and hazard ratio, conditional and marginal estimates may differ even in the absence of confounding, mediation, or effect modification, for reasons unrelated to the Table 2 fallacy.^{15,16}

useful for prediction, but they do not define a causal adjustment set. A variable may be needed for confounding control even if it is not statistically significant, while a statistically significant variable may be inappropriate to adjust for if it is a mediator or a collider (a variable influenced by both the exposure and the outcome, which adjustment can bias).¹⁷

Third, present the primary result as an estimation graphic rather than a dense coefficient table in which every variable invites a causal reading. The graphic shows the single-adjusted exposure effect with its full uncertainty rather than a significance test; in the worked example, the smoking-adjusted outcome is plotted for 2 groups, with the adjusted effect and its bootstrap distribution shown alongside (Supplementary Fig. S1). The same principle of reporting the adjusted effect with its uncertainty applies whatever the effect measure (mean difference, risk ratio, or risk difference).

Finally, because no observational adjustment can be assumed complete, the extent of residual confounding should be quantified, for example with an E-value, which indicates how strong an unmeasured confounder would need to be to explain away the observed exposure effect.¹⁸ Box 1 lists red flags, recurring pitfalls, and questions a reviewer can ask of any risk-factor table.

Box 1. Spotting the Table 2 fallacy: Red flags, common pitfalls, and questions for reviewers.

Red flags in a manuscript

- Every row in a multivariable “risk factors” table is discussed as if it were clinically actionable.
- Adjustment variables are labelled as “independent predictors”, “risk factors”, or “predisposing factors”.
- The discussion section gives mechanisms or intervention implications for covariates that were not the main exposure of interest.
- Interaction or subgroup terms are interpreted causally, as a joint effect, without a clearly stated causal question.
- Variables were selected by stepwise selection or statistical significance, then interpreted causally.

Common pitfalls

- Reading an adjustment-variable coefficient as its total effect.
- Assuming the adjustment set for the exposure also works for every other covariate.
- Using cautious “association only” language, but later making causal or clinical claims in the discussion section.
- Treating statistical significance as evidence of causal importance.

Questions a reviewer can ask

- What causal question, or set of target estimands, was the model designed to answer?
- Was the adjustment set chosen using subject-matter knowledge (e.g. a causal diagram) or selected statistically?
- Are any non-exposure coefficients interpreted as independent risk factors and clinically actionable findings?
- For each non-exposure coefficient that is interpreted, is there a separate causal question and justified adjustment set?

Rule of thumb

- Interpret the coefficient the model was designed to estimate.
- For each further causal question, specify the question, causal assumptions, adjustment set, and model separately.

None of these steps is demanding, yet together they turn a regression table from a list of numbers into a clearly defined estimate whose causal interpretation can be assessed.

CONCLUSION

The Table 2 fallacy arises when coefficients included for adjustment are interpreted as separate causal effects. Avoiding it requires a clearly defined exposure, the effect to be estimated, the underlying assumptions, and the variables needed for adjustment. This principle applies to every table reporting estimates from a multivariable model: interpret the effect the analysis was designed to estimate, and treat the remaining coefficients as adjustment terms rather than findings in their own right. The distinction requires no new software or larger sample, only the habit of asking what causal question each coefficient answers and whether the study design can answer it.

Supplementary materials

Fig. S1. Estimation graphic for the adjusted exposure effect (illustrative simulated data).

Appendix S1. R code for the simulation.

Ethics statement

Not applicable. This article is a methodological commentary using simulated and illustrative data only.

Declaration

The authors declare that they have no affiliations or financial involvement with any commercial organisation with a direct financial interest in the subject or materials discussed in the manuscript. There is no funding or conflict of interest to declare.

Acknowledgments

John Rong Hao Tay is supported by the Singapore National Medical Research Council Research Training Fellowship (RTF24jan-0007). Nasir Z Bashir is supported by the Wellcome Trust (322777/Z/24/Z).

REFERENCES

1. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat* 2016;70:129-33.
2. Putt ME. Assessing risk factors with information beyond *P* value thresholds: Statistical significance does not equal clinical importance. *Cancer* 2021;127:1180-5.
3. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019;567:305-7.
4. Trafimow D, Marks M. Editorial [Banning null hypothesis significance testing procedures]. *Basic Appl Soc Psychol* 2015;37:1-2.
5. Lang JM, Rothman KJ, Cann CI. That confounded *P*-value. *Epidemiology* 1998;9:7-8.
6. Hayat MJ, Staggs VS, Schwartz TA, et al. Moving nursing beyond *p* < 0.05. *Res Nurs Health* 2019;42:244-5.

7. Ho J, Tumkaya T, Aryal S, et al. Moving beyond P values: Data analysis with estimation graphics. *Nat Methods* 2019; 16:565-6.
8. Tan SH, Tan SB. The correct interpretation of confidence intervals. *Proc Singap Healthc* 2010;19:276-8.
9. Westreich D, Greenland S. The Table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol* 2013;177:292-8.
10. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37-48.
11. Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: Review and recommendations. *Int J Epidemiol* 2021;50:620-32.
12. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology* 2009;20:863-71. Erratum in: *Epidemiology* 2011;22:752.
13. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009;20:488-95.
14. Tay JRH, Cooray U. Letter to Editor for “Potential predisposing features of external cervical resorption: An observational study (Patel et al., 2025)”. *Int Endod J* 2025;58:797-8.
15. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999;14:29-46.
16. Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010; 21:13-5.
17. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. Cham: Springer; 2015.
18. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: Introducing the E-value. *Ann Intern Med* 2017; 167:268-74.

Transcatheter aortic valve implantation for severe aortic regurgitation with Trilogy system: Initial Southeast Asia experience

Jonathan Yap^{*1,2} MRCP, Samuel Ji Quan Koh^{*1} MRCP, Zameer Abdul Aziz³ MBBS, Sivaraj Govindasamy³ MBBS, Victor Tar Toong Chao³ FRCS, Mohammed Rizwan Amanullah¹ MRCP, See Hooi Ewe^{1,2} PhD, Kay Woon Ho^{1,2} MRCP

Dear Editor,

Aortic regurgitation (AR) is a common valvular heart disease, with a cohort study reporting an approximate prevalence of 13% in men and 8.5% in women.¹ Severe AR portends significant mortality and morbidity if untreated.² Aortic valve surgery remains the gold standard for management.³ However, a significant proportion are treated conservatively because of high surgical risks, significant comorbidities, or the patient's refusal of surgery.⁴ Despite the success of transcatheter aortic valve implantation (TAVI) for the management of aortic stenosis (AS), existing devices have not been designed for the treatment of AR. In patients with limited surgical options, these devices have been used off-label with higher associated risks.⁵ The recent development of dedicated devices has changed the landscape.^{6,7} In particular, the ALIGN-AR study demonstrated the safety and efficacy of the use of the Trilogy transcatheter heart valve (THV) (JenaValve Technology, Irvine, CA, US) in treating patients with severe tricuspid AR.⁷ The Trilogy valve is a porcine pericardial THV composed of 3 leaflets within a self-expanding nitinol stent scaffold. It has 3 locators which are centred on the native aortic cusps, limiting implant depth and allowing clipping to the native leaflets for secure anchoring.⁷ To the authors' knowledge, this is the first case series in Southeast Asia (SEA) to describe the initial experience with the Trilogy THV for the treatment of severe tricuspid AR.

Consecutive patients who underwent TAVI for severe tricuspid AR with the Trilogy THV in a tertiary cardiac centre were included. Severe AR was diagnosed according to the European Society of Cardiology 2025 guidelines,³ and patients were evaluated by a multidisciplinary heart team. A computed tomography (CT) scan was performed to assess the feasibility of the transfemoral route and the suitability/sizing for implantation. All procedures were performed via the transfemoral route under general anaesthesia and transoesophageal echocardiographic guidance. In

summary, a dedicated 18 French sheath (measuring 85cm) was inserted via the femoral route to the sinotubular junction. The valve was then inserted via the sheath to the aortic root, after which the sheath was retracted to the descending aorta. The locators were adjusted, and after all 3 locators were confirmed in the native cusps and appropriate depth was achieved, the valve was deployed under rapid pacing (Fig. 1). Technical success and all clinical outcomes were defined according to the Valve Academic Research Consortium-3 criteria⁸ both in-hospital and at 3 months.

A total of 5 patients were treated (median age 75 years, range: 49–83 years). Two patients had preexisting left or right bundle branch blocks (LBBB/RBBB). Three patients were New York Heart Association (NYHA) Class II, while 2 were NYHA I with left ventricular (LV) dysfunction. Four patients were deemed high risk for surgery, and 1 patient repeatedly declined surgery in view of age (75 years old). Echocardiographic and CT measurements are also outlined in Supplementary Table S1.

Overall, procedural success was achieved for all patients, with a median procedure time of 85 minutes. There were no major in-hospital complications. Post-procedural echocardiogram showed mild paravalvular AR in 1 patient, while the rest had no residual AR. However, there was a significant incidence of conduction system disease post-implant, with 1 patient developing new LBBB, and 3 patients developing complete heart block (CHB) requiring device implantation (of which 2 had underlying LBBB and RBBB, respectively). Of the 3 patients, the degree of oversizing ranged from 6.06–18%, with 2 requiring permanent pacemaker (PPM) implantation and 1 requiring cardiac resynchronisation therapy with a defibrillator (because of concomitant indication for primary prevention with an LVEF <35%). Notably, all 3 patients had the smaller of the 2 possible valve sizes implanted. Median length of stay was 6 (range: 3–13) days, largely due to monitoring and management of conduction issues. Two patients

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Department of Cardiology, National Heart Centre Singapore

² Duke-NUS Medical School, Singapore

³ Department of Cardiothoracic Surgery, National Heart Centre Singapore, Singapore

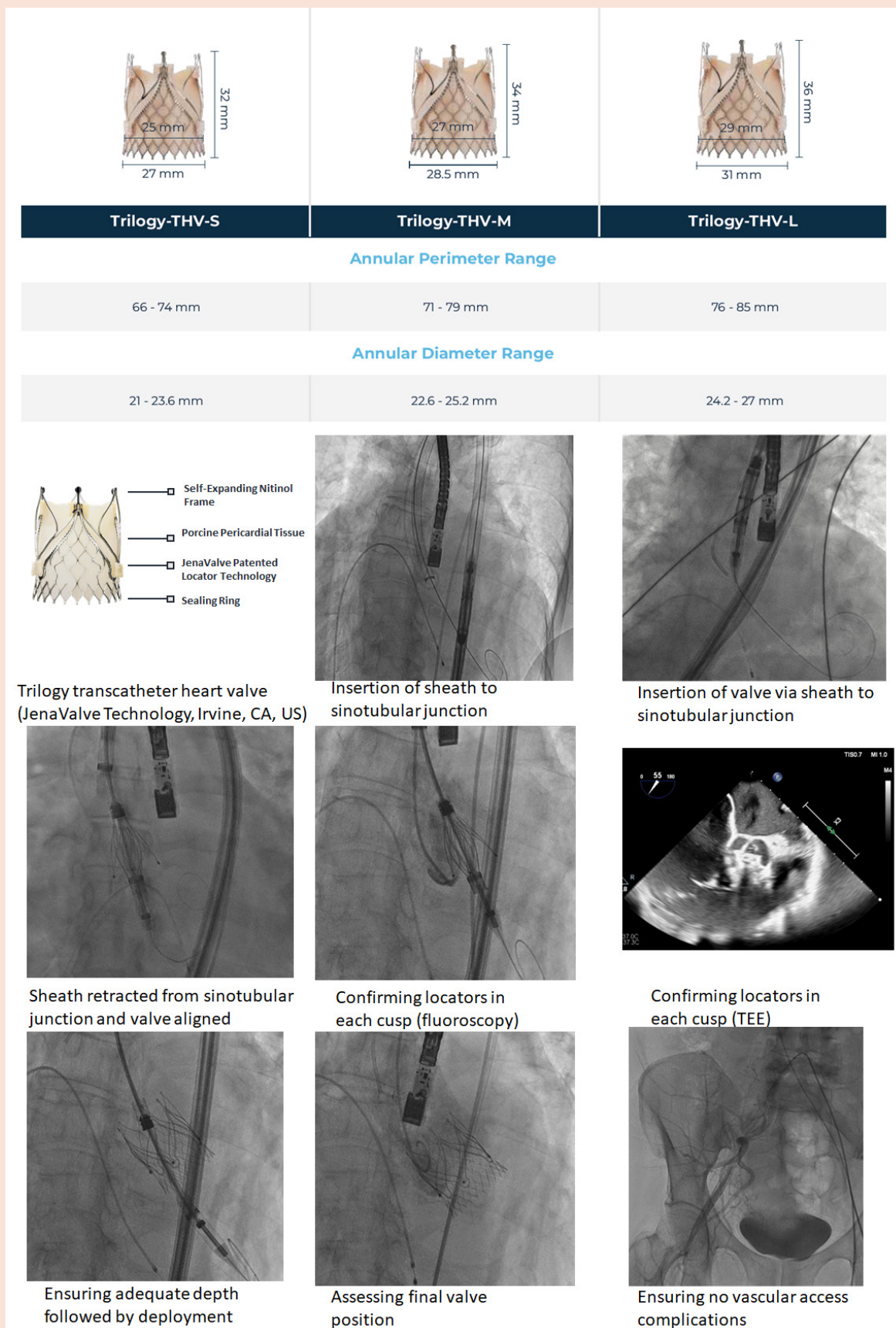
* Joint first authors

Correspondence: A/Prof Jonathan Yap, National Heart Centre Singapore, 5 Hospital Dr, Singapore 169609. Email: jonyap@yahoo.com

Accepted: 20 May 2026

Published Online First: 2 June 2026

Fig. 1. Trilogy transcatheter heart valve (JenaValve Technology, Irvine, CA, US): sizing chart, design,⁹ and deployment.



S: small; M: medium; L: large; THV: Trilogy transcatheter heart valve; TEE: transesophageal echocardiography
Image of sizing chart and valve courtesy of JenaValve Technology, Inc.

who did not require PPM had hospital stays of 3 and 4 days, respectively.

At 3 months, all patients were well, with 4 of the 5 patients at NYHA I and 1 patient stable at NYHA II. Valve performance remained excellent, with no bioprosthetic dysfunction and minimal or no residual AR. Patients with LV dilation demonstrated reverse remodelling.

This first case series in SEA demonstrated that the Trilogy THV represents a safe and feasible option for patients with severe tricuspid AR.

Prior to the advent of AR-specific THVs, there was off-label use of pre-existing TAVI devices for those who were unable to undergo surgery. PANTHEON, the largest investigator-initiated international registry evaluating the off-label use of such TAVI devices for AR, demonstrated an increased incidence of residual paravalvular leak (9.5%), valve embolisation (12.4%), need for a second valve (10.5%), and surgical conversion (2%).⁵ In contrast to AS, there is often greater annular dilation, and a lack of leaflet and annular calcification, which makes anchoring of the THV challenging.

The advent of AR-specific TAVI devices has changed this. These devices have a dedicated mechanism to anchor on the native valve leaflets, independent of the degree of calcification.¹⁰ These include the J-Valve (JC Medical Inc, Burlingame, CA, US) and the Trilogy THV. In the North American experience with the latest transfemoral iteration of the J-Valve,¹¹ procedural success was 81% (22/27) overall and 100% in the last 15 cases. At 30 days, 1 death and 1 stroke occurred. In ALIGN-AR, using the Trilogy valve,⁷ procedural success was 95% (171/180). At 30 days, there was a 2% mortality rate and a 2% stroke rate. In this case series, there was 100% procedural success with no complications. At 3 months, there was improvement in symptoms with no significant AR.

Several limitations should be noted with this treatment strategy. First, the development of conduction system abnormalities necessitating device implantation remains a concern. In ALIGN-AR, the rate of PPM implantation was 24.0%.⁷ While the rate of device implantation is higher in this small case series (3/5), 2 of the patients had pre-existing conduction system abnormalities, which put them at a higher risk for the development of CHB. Notably, ALIGN-AR showed a decrease in this complication with less aggressive oversizing and slight modification of implantation technique with operator experience. Further studies are required to assess predictors of the development of conduction system abnormalities, as well as potential mitigators. Second, as with all novel devices, long-term durability will require extended studies.

In this first case series in SEA, the Trilogy THV was safe and efficacious in the treatment of severe tricuspid AR. There was 100% procedural success, and no mortality or significant AR at short-term follow-up. However, a significant proportion of patients developed CHB necessitating device implantation. Current gaps include identifying measures to mitigate conduction system abnormalities and assessing long-term durability.

Supplementary material

Table S1. Baseline characteristics, procedural details, and clinical outcomes (n=5).

Ethics statement

Ethics approval for the use of the patient registry was granted by the SingHealth Centralised Institutional Review Board (2014/2165).

Declaration

Jonathan Yap received speaker's honorarium from Abbott, Biosensors, Boston Scientific, Edwards, Johnson & Johnson, and Medtronic and is a proctor for Abbott, Edwards and Medtronic. Mohammed Rizwan Amanullah received speaker's honorarium from GE HealthCare and Philips. The other co-authors declare there are no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Keywords: aortic regurgitation, transcatheter aortic valve implantation, JenaValve, TAVI, Trilogy System, JVTs

REFERENCES

1. Singh JP, Evans JC, Levy D, et al. Prevalence and clinical determinants of mitral, tricuspid, and aortic regurgitation (the Framingham Heart Study). *Am J Cardiol* 1999;83:897-902.
2. Dujardin KS, Enriquez-Sarano M, Schaff HV, et al. Mortality and morbidity of aortic regurgitation in clinical practice. A long-term follow-up study. *Circ* 1999;99:1851-57.
3. Praz F, Borger MA, Lanz J M-CM, et al. 2025 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J* 2025;46:4635-736.
4. Lung B, Baron G, Butchart EG, et al. A prospective survey of patients with valvular heart disease in Europe: The Euro Heart Survey on Valvular Heart Disease. *Eur Heart J* 2003;24:1231-43.
5. Poletti E, De Backer O, Scotti A, et al. Transcatheter Aortic Valve Replacement for Pure Native Aortic Valve Regurgitation: The PANTHEON International Project. *JACC Cardiovasc Interv* 2023;16:1974-85.
6. Gao L, Mao R, Zeng J, et al. Single arm meta-analysis of the J-Valve system for aortic regurgitation in Chinese populations. *Front Cardiovasc Med* 2025;12:1436789.
7. Vahl TP, Thourani VH, Makkar RR, et al. Transcatheter aortic valve implantation in patients with high-risk symptomatic native aortic regurgitation (ALIGN-AR): a prospective, multicentre, single-arm study. *Lancet* 2024;403:1451-59.

8. Généreux P, Piazza N, Alu MC, et al. Valve Academic Research Consortium 3: updated endpoint definitions for aortic valve clinical research. *Eur Heart J* 2021;42:1825-57.
9. JenaValve Technology. JenaValve Trilogy Heart Valve System Procedural Manual, PE-0114 Rev B. Irvine, California: JenaValve Technology; 2024.
10. Costanzo P, Bamborough P, Peterson M, et al. Transcatheter Aortic Valve Implantation for Severe Pure Aortic Regurgitation With Dedicated Devices. *Interv Cardiol* 2022;17:e11.
11. Garcia S, Ye J, Webb J, et al. Transcatheter Treatment of Native Aortic Valve Regurgitation: The North American Experience With a Novel Device. *JACC Cardiovasc Interv* 2023;16:1953-60.

Real-world drug utilisation of nintedanib and pirfenidone for interstitial lung disease in Singapore

Tingfeng Lee¹ MSc, Yi Jing Lim¹ PharmD, Andrew Yunkai Li² MBBS, Ziqin Ng² MBBS, Zhikui Chrystal Huang³ BBiomedSc, Gin Tsen Chai² MBBS

Dear Editor,

Antifibrotic therapies such as nintedanib and pirfenidone have transformed interstitial lung disease (ILD) treatment, representing a significant advancement in managing progressive fibrosing conditions. While these agents are used across Asia, their use in Southeast Asia, for example in Singapore, Malaysia or Indonesia, remains significantly under-represented in literature. To address these knowledge gaps in real-world use of antifibrotics among Southeast Asian population, the authors conducted a retrospective drug use evaluation at the ILD subspecialty service in Tan Tock Seng Hospital in Singapore from January 2016 to October 2024, with the aim to evaluate the usage patterns and clinical outcomes of nintedanib and pirfenidone in patients with ILD.

A total of 44 patients were included in this analysis, 26 (59%) patients received nintedanib and 18 (41%) received pirfenidone. Among the nintedanib group, 5 patients were subsequently switched to pirfenidone, with 1 patient later returning to nintedanib therapy. For analysis purposes, patients were grouped according to their initial antifibrotic medication. All patients screened were included and no patients were excluded for data analysis.

Baseline characteristics were comparable between the groups, with no differences in age, race, body mass index, and smoking status (Supplementary Table S1). There were more males on pirfenidone than nintedanib (100% vs 80.8%, $P=0.048$). Baseline forced vital capacity (FVC) (percentage predicted) and diffusing capacity of the lungs for carbon monoxide (DLCO) (percentage predicted) were comparable between the groups. The median number of comorbidities was 3 (interquartile range [IQR] 2–4) in the pirfenidone group and 2 (IQR 1–4) in the nintedanib group. Patients with ≥ 3 comorbidities were more common in the pirfenidone group (61.1% versus [vs] 38.5%). The most prevalent comorbidities were diabetes mellitus, pulmonary hypertension, and liver disease,

with no significant differences in individual comorbidity patterns between the groups. Idiopathic pulmonary fibrosis (IPF) was the predominant diagnosis, though its prevalence differed significantly between the groups (nintedanib: 79.5% vs pirfenidone: 100%, $P=0.006$). The remaining nintedanib patients (20.5%) had progressive pulmonary fibrosis (PPF), comprising equal numbers of PPF, fibrotic hypersensitivity pneumonitis, and connective tissue disease-associated ILD.

Drug interruption/discontinuation was significantly more common in the pirfenidone group (77.8%) compared to the nintedanib group (53.8%, $P=0.049$) (Table 1). Dose reduction was observed in 26.9% of nintedanib patients but none in the pirfenidone group. The main reasons for treatment modifications were adverse events (nintedanib: 42.3% vs pirfenidone: 27.8%) and deaths (nintedanib: 15.4% vs pirfenidone: 27.8%), with no significant difference in overall modification reasons between groups ($P=0.804$). Other reasons included precautionary measures for cardiovascular concerns or worsening organ function, and miscellaneous factors such as financial concerns or pill burden.

Target dose achievement, defined as total daily dose of nintedanib 300 mg daily and pirfenidone 2403 mg daily, was higher in the nintedanib group (50.0% vs 33.3%, $P=0.272$). The mean daily dose achieved was 256 ± 50.8 mg for nintedanib and 1691 ± 607.5 mg for pirfenidone. Among patients not achieving target doses, adverse events were the primary reason in both groups, though more prominent in the pirfenidone group (50.0% vs 34.6%). Other reasons included advanced disease concerns, and precautionary dose limitations due to comorbidities such as chronic liver disease or progressive chronic kidney disease.

The safety analysis revealed distinct adverse event profiles between the 2 medications. Liver enzyme elevations were significantly more frequent with nintedanib (38.5% vs 11.1%, $P=0.046$) of which

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Division of Pharmacy, Tan Tock Seng Hospital, Singapore

² Department of Respiratory and Critical Care Medicine, Tan Tock Seng Hospital, Singapore

³ Clinical Research & Innovation Office, Tan Tock Seng Hospital, Singapore

Correspondence: Mr Tingfeng Lee, Division of Pharmacy, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433.

Email: tingfeng.lee@nhghealth.com.sg

Accepted: 15 April 2026

Published Online First: 22 April 2026

Table 1. Usage pattern, adverse event profile, and clinical outcomes of antifibrotic therapy in patients with interstitial lung disease.

	Nintedanib (n=26)	Pirfenidone (n=18)	P value
Drug interruption, no. (%) ^a	14 (53.8)	14 (77.8)	0.049
Dose reduction, no. (%)	7 (26.9)	0	
Reason for drug interruption or dose reduction, no. (%)			0.804
Side effect	11 (42.3)	5 (27.8)	
Deceased	4 (15.4)	5 (27.8)	
Precaution ^b	2 (7.7)	1 (5.6)	
Others ^c	4 (15.4)	3 (16.7)	
Achieved target doses of antifibrotic, no. (%)	13 (50.0)	6 (33.3)	0.272
Average total daily dose of antifibrotic, mean SD, mg	256.0±50.8	1691.0±607.5	
Reasons for not achieving target dose, no. (%)			0.100
Side effect(s)	9 (34.6)	9 (50.0)	
Precautions ^d	0	2 (11.1)	
Advanced disease	1 (3.8)	0	
Others ^e	0	1 (5.6)	
Not specified	3 (11.5)	0	
Adverse events, no. (%)			
Elevated liver enzymes	10 (38.5)	2 (11.1)	0.046
Pattern of liver injury			0.394
Hepatocellular injury	4 (15.4)	0	
Cholestatic injury	3 (11.5)	0	
Mixed injury	2 (7.7)	2 (11.1)	
Unable to be determined	1 (3.8)	0	
Gastrointestinal side effect	9 (34.6)	8 (44.4)	0.545
Dermatological side effect	0	1 (5.6)	0.409
General disorder ^f	0	3 (16.7)	0.062
No documented side effect experienced	10 (38.5)	8 (44.4)	0.761
Time from start of antifibrotic to follow-up pulmonary function test, mean SD, months	9.9±3.4	10.0±4.3	0.918
Change in FVC, mean SD, percentage predicted	4.0±14.1	-1.5±14.9	0.335
Change in DLCO, mean SD, percentage predicted	-1.76±9.7	-12.5±11.7	0.069
Mortality, no. (%)	13 (50.0)	12 (66.7)	
Time from start of antifibrotic to death, median (95% CI), weeks	85 (0–182)	110 (71–149)	0.542

CI: confidence interval; DLCO: diffusing capacity of the lungs for carbon monoxide; FVC: forced vital capacity; SD: standard deviation

^aDrug interruptions include patients who are deceased.

^bPrecautions comprise cardiovascular concerns (initiation of antiplatelet, myocardial infarction), new diagnosis of liver cirrhosis, worsening renal function.

^cOthers include advanced disease and uncertain benefits, revision of diagnosis of interstitial lung disease, financial concerns, patient defaulted as overseas, pill burden.

^dPrecautions include concerns about hepatotoxicity in view of chronic liver disease with hepatic steatosis, renal function in view of progressive chronic kidney disease.

^ePatient declined dose escalation and had wanted to try Ayurvedic medications.

^fGeneral disorders (e.g. fatigue, non-cardiac chest pain, general feeling of discomfort).

4 patients showed hepatocellular injury patterns, 3 showed cholestatic injury, and 2 showed mixed injury patterns while 2 on pirfenidone showed mixed injury patterns.

Gastrointestinal side effects including nausea, vomiting, diarrhoea, anorexia, and poor appetite were comparable between groups (nintedanib: 34.6% vs pirfenidone: 44.4%, $P=0.545$). General disorders such as fatigue, non-cardiac chest pain, and general feeling of discomfort were observed exclusively with pirfenidone (16.7%, $P=0.062$). Dermatological side effects were rare, with only 1 patient in the pirfenidone group experiencing an eczema flare. Notably, a substantial proportion of patients did not experience any side effects during antifibrotic therapy—encompassing 38.5% in the nintedanib group and 44.4% in the pirfenidone group ($P=0.761$).

Overall medication adherence was good in both treatment groups, as measured by medication possession ratio (MPR). The median MPR was 98.6% for nintedanib and 94.4% for pirfenidone ($P=0.394$). Good adherence, defined as $MPR \geq 80\%$, was achieved in 78.3% of the nintedanib group and 71.4% of pirfenidone group ($P=0.732$).¹

After a mean follow-up of 9.9 ± 3.4 months for nintedanib and 10.0 ± 4.3 months for pirfenidone, both groups showed no statistically different patterns in lung function changes (Table 1). Mean FVC change was comparable between the groups (nintedanib: $4.0 \pm 14.1\%$ predicted vs pirfenidone: $-1.5 \pm 14.9\%$ predicted, $P=0.335$). Similarly, mean DLCO change showed no significant difference between the groups (nintedanib: $-1.76 \pm 9.7\%$ predicted vs pirfenidone: $-12.5 \pm 11.7\%$ predicted, $P=0.069$). Survival analysis revealed no statistically significant difference between the treatment groups. Mortality rates over the study period were 50% in the nintedanib group and 67% in the pirfenidone group. Median survival was 85 weeks (95% CI 0–182) for nintedanib and 110 weeks (95% CI 71–149) for pirfenidone (log-rank $P=0.542$).

The study's retrospective design is limited by potential biases including incomplete documentation and under-reporting of mild adverse events. Treatment selection was based on physician discretion considering multiple factors including approved indications (pirfenidone restricted to IPF only), patient comorbidities, potential drug interactions, contraindications to specific antifibrotic agents, pill burden preferences, and cost considerations. Notably, nintedanib is cheaper than pirfenidone in this study's healthcare setting (about SGD700 or approximately USD550; 20% cheaper per month at full doses). While MPR is widely utilised for medication adherence

assessment, it has limitations including inability to confirm actual medication ingestion, and potential imprecision over short intervals.⁷

This study provided important real-world evidence on antifibrotic therapy use, safety, and effectiveness in patients with ILD in Singapore. Despite frequent dose modifications and substantial comorbidity burden, most patients maintained high medication adherence, and experienced lung function preservation comparable to international studies. However, adverse event profiles, particularly hepatic enzyme elevations, were more pronounced than in clinical trials, highlighting the need for enhanced monitoring and individualised management strategies.

Supplementary material

Supplementary Table S1. Baseline demographic and clinical characteristics of patients with interstitial lung disease treated with nintedanib or pirfenidone.

Ethics statement

Ethics approval was obtained from the National Healthcare Group Domain Specific Review Board (2024-4281). Waiver of informed consent was granted as this was a retrospective study involving minimal risk to participants and using only existing medical records data.

Declaration

No generative artificial intelligence (AI) or AI-assisted technologies were used in the preparation of this manuscript. Gin Tsen Chai reports honoraria for lectures and advisory board fees from Boehringer Ingelheim Singapore Pte Ltd paid via his institution. Ziqin Ng reports honoraria for lectures from Boehringer Ingelheim Singapore Pte Ltd paid via her institution. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. The authors have no relevant financial or non-financial interests to disclose.

Keywords: antifibrotic, interstitial lung disease, nintedanib, pirfenidone, respiratory medicine, Singapore

REFERENCES

1. Karve S, Cleves MA, Helm M, et al. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. *Curr Med Res Opin* 2009;25:2303-10.
2. Chang CY, Wei YF, Chen CY, et al. Real world experience on the effectiveness and safety of pirfenidone in patients with idiopathic pulmonary fibrosis in Taiwan. *Front Med (Lausanne)* 2023;10:1242260.
3. Lee HY, Jung SY, Jang JH, et al. Efficacy of Pirfenidone According to Dose in Patients with Idiopathic Pulmonary Fibrosis: A

- Prospective, Observational, Single-Center Cohort Study. *Life (Basel)* 2023;13:2118.
4. Tran T, Šterclová M, Mogulkoc N, et al. The European MultiPartner IPF registry (EMPIRE): validating long-term prognostic factors in idiopathic pulmonary fibrosis. *Respir Res* 2020;21:11.
5. Moodley Y, Goh N, Glaspole I, et al. Australian Idiopathic Pulmonary Fibrosis Registry: Vital lessons from a national prospective collaborative project. *Respirology* 2014;19:1088-91.
6. Agency for Care Effectiveness, Ministry of Health, Singapore. Nintedanib and pirfenidone for treating idiopathic pulmonary fibrosis. 2 January 2019. [https://isomer-user-content.by.gov.sg/68/281338d5-1484-4ad0-b7ff-c2c809291177/nintedanib-and-pirfenidone-for-ipf-\(2-jan-2019\).pdf](https://isomer-user-content.by.gov.sg/68/281338d5-1484-4ad0-b7ff-c2c809291177/nintedanib-and-pirfenidone-for-ipf-(2-jan-2019).pdf). Accessed 22 April 2026.
7. Shah KK, Touchette DR, Marrs JC. Research and scholarly methods: Measuring medication adherence. *J Am Coll Clin Pharm* 2023;6:416-26.

A primary care perspective on gender-focused mental health services in Singapore: Correspondence

Jeremiah Pereira¹ MMed (Int Med)

Dear Editor,

This author read with interest the recent letter by Puthran et al. on gender-focused mental health services in Singapore,¹ as well as the subsequent correspondence by Fung et al., which contextualises the integration of such services within general psychiatric care.²

This author welcomes several key directions highlighted across both pieces, particularly the recognition of evolving models of care, and the need for system-wide capability building. This letter response offers a primary care perspective to further contextualise these discussions.

The informed consent model. First, it is encouraging that Puthran et al. acknowledge the current predominance of a mental health assessment model, noting that “most providers seem more comfortable with a prior psychiatric assessment... rather than an informed consent approach.” This distinction is important. While psychiatric involvement remains essential in selected cases—particularly those with complex presentations, significant psychiatric instability, or diagnostic uncertainty—over-reliance on mandatory assessment frameworks risks creating unnecessary barriers to care.³ The American College of Obstetricians and Gynecologists states that “current consensus is that an informed consent process without a separate letter from a mental health care professional is more than adequate for initiating therapy for those patients who wish to medically transition”.⁴ The Endocrine Society’s 2017 guidelines similarly moved away from requiring obligatory psychosocial evaluation, a position affirmed in the World Professional Association for Transgender Health Standards of Care Version 8.⁵

Research demonstrates that the informed consent model is associated with high patient satisfaction, and significantly shorter time to treatment initiation.⁶ For individuals with clear, persistent gender incongruence who are otherwise medically suitable for treatment, an informed consent model allows for more timely care without compromising safety.

Delineating case complexity. Second, this author supports the framing of case complexity, particularly the statement that “more complex cases requiring a second opinion or trained multidisciplinary support” should be referred to specialised services.¹

Implicit within this is the recognition that not all cases are complex. A substantial proportion of individuals presenting with gender-related concerns may be appropriately managed in primary care settings, provided clinicians are adequately trained and supported. Multiple healthcare systems have successfully implemented primary care-led gender-affirming hormone therapy programmes in cities including Toronto, Vancouver, Boston, and Sydney.³ Clear delineation between straightforward and complex presentations is therefore critical to avoid over-centralisation and bottlenecks in specialist services.

Co-occurring conditions as rationale for integration. Third, both letters highlight the high prevalence of co-occurring mental health and neurodevelopmental conditions among gender-diverse individuals.

From a primary care standpoint, this should not be viewed as a barrier to decentralised care, but rather as a rationale for it. Primary care physicians routinely manage mild-to-moderate mood disorders, anxiety conditions, and neurodevelopmental presentations within a biopsychosocial framework.⁵ Integrating gender care into this existing scope aligns with established primary care competencies and supports continuity of care.

Clinical guidelines for co-occurring autism and gender dysphoria emphasise the importance of assessment but do not preclude gender-affirming care. Rather, they recommend an extended diagnostic period with overlap between assessment and treatment.⁶ The presence of co-occurring conditions should trigger appropriate assessment and support, not automatic exclusion from primary care management.

Building system-wide competence. Finally, this author strongly agrees with Fung et al. in that “multiplying subspecialty silos” is unlikely to be

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Pulse Clinic, Singapore

Correspondence: Dr Jeremiah Pereira, Pulse Clinic, 118A Neil Road, 2nd Floor, Singapore 088854. Email: jeremiahjamespereira@gmail.com

Accepted: 12 May 2026

Published Online First: 19 May 2026

sustainable in a resource-constrained system, and that “building system-wide competence” should be prioritised.²

In this regard, primary care represents a natural and scalable platform for capability building. The medications used for gender transition are common and can be safely prescribed by primary care physicians with appropriate training. Primary care physicians already have experience managing hormone-based therapies in contexts such as menopause, andropause and contraception. With appropriate training—including understanding the diagnostic criteria for gender dysphoria, prescribing and monitoring hormone therapy, managing common side effects, and recognising when specialist referral is needed—similar principles can be extended to gender-affirming hormone therapy.

Multiple studies emphasise that creating welcoming environments, training all staff in culturally competent care, and integrating gender-affirming services into existing care structures are more sustainable than creating isolated specialty clinics. The AFFIRM framework developed from patient narratives emphasises that care should be flexible, accessible, multidisciplinary, and community-centered.⁹

Singapore’s own mental health landscape demonstrates the value of primary care integration. Studies show that patients often consult multiple providers before reaching psychiatric services, with primary care serving as a key pathway.¹⁰

A pragmatic model for Singapore. Taken together, a pragmatic model for Singapore may involve:

- (1) Primary care-led management of straightforward cases using an informed consent framework.
- (2) Clear referral pathways for complex or uncertain presentations requiring specialist mental health assessment.
- (3) Shared care models linking primary care with multidisciplinary specialist services, where primary care providers maintain ongoing management with access to specialist consultation as needed.
- (4) System-wide training to improve baseline competence across providers, with emphasis on cultural humility, clinical skills, and recognition of when specialist referral is appropriate.

Such an approach would align with the goals of accessibility, equity, and sustainability, while avoiding unnecessary delays in care for patients who could otherwise be managed safely in community settings. The continuum of care framework recommended by

recent guidelines— emphasising creating an affirming environment, establishing diagnosis and goals, selecting appropriate pharmacotherapy, and monitoring both hormone levels and clinical outcomes—can be delivered effectively in primary care settings with appropriate training and specialist backup.¹¹

Ethics statement

Not applicable as no participants were recruited.

Declaration

The authors declare there are no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Keywords: endocrinology, family medicine, internal medicine, mental health, psychiatry

REFERENCES

1. Puthran R, Tan HT, Rawtaer I, et al. Gender-focused mental health services in Singapore: A brief history and future directions. *Ann Acad Med Singap* 2026;55:167-8.
2. Fung DSS, Verma S, Ong SH, et al. Gender-focused mental health services in Singapore: A brief history and future direction: Correspondence. *Ann Acad Med Singap* 2026;55:228.
3. Wylie K, Knudson G, Khan SI, et al. Serving Transgender People: Clinical Care Considerations and Service Delivery Models in Transgender Health. *Lancet* 2016;388:401-11.
4. Health Care for Transgender and Gender Diverse Individuals. ACOG Committee Opinion, Number 823. *Obstet Gynecol* 2021;137:e75-88.
5. Bisno DI, Lubitz S, Marshall I, et al. A National Survey of United States-Based Endocrinologists Who Prescribe Gender-Affirming Hormone Therapy. *Endocr Pract* 2023;29:465-70.
6. Spanos C, Grace JA, Leemaqz SY, et al. The Informed Consent Model of Care for Accessing Gender-Affirming Hormone Therapy Is Associated With High Patient Satisfaction. *J Sex Med* 2021;18:201-8.
7. Peh ALH, Tan GCS, Soon WSW, et al. Psychiatry in Primary Care and Training: A Singapore Perspective. *Singapore Med J* 2021;62:210-2.
8. Strang JF, Meagher H, Kenworthy L, et al. Initial Clinical Guidelines for Co-Occurring Autism Spectrum Disorder and Gender Dysphoria or Incongruence in Adolescents. *J Clin Child Adolesc Psychol* 2018;47:105-15.
9. Quint M, Bailar S, Miranda A, et al. The AFFIRM Framework for Gender-Affirming Care: Qualitative Findings From the Transgender and Gender Diverse Health Equity Study. *BMC Public Health* 2025;25:491.
10. Chang S, Jeyagurunathan A, Abdin E, et al. Mapping the Steps to Reach Psychiatric Care in Singapore: An Examination of Services Utilized and Reasons for Seeking Help. *Gen Hosp Psychiatry* 2021;73:38-45.
11. Goldstein Z, Krasowski MD, Greene DN. Gender-Affirming Feminizing Hormone Therapy. *JAMA Intern Med* 2025;185:1027-8.

"Gazing into the crystal ball of mortality prediction in conservative kidney care": Correspondence

Norhasyimah Tamin¹ MMed (Brunei), Fazean Idris² MRCGP (UK), Shin Bey Ho³ MRCP (UK),
Vui Heng Chong⁴ FRCP (UK), Jackson Tan³ FRCP (UK)

Dear Editor,

We refer to the *Annals* article, "Gazing into the crystal ball of mortality prediction in conservative kidney care", which highlights the paucity of data contextualising mortality predictors within Southeast Asian populations and the call to embrace adoption of conservative kidney management (CKM).¹ In this letter, we would like to present our perspective on CKM, as a fourth modality of kidney replacement therapy (KRT), set against the backdrop of the current economic crisis of dialysis care.

We conducted a retrospective cohort study of all patients aged over 70 years with end-stage kidney disease (ESKD) in Brunei who received either KRT or CKM between 2018 and 2023. Data from the Brunei Dialysis and Transplant Registry included sex, race, comorbidities, age, KRT initiation dates, baseline biochemical markers, and outcomes such as survival, hospitalisations, place of death, cardiac arrests, and intensive care admissions. The study aimed to report survival rates and compare demographic and clinical differences between groups. The KRT group included patients starting dialysis electively or acutely, while the CKM group included patients enrolled in the CKM clinic. Both groups of patients were selected from the national pre-dialysis clinic to either start dialysis or be referred for CKM. Patients with metastatic cancer or severe dementia were excluded from both groups in this analysis.

Statistical analyses were conducted using R version 4.6.0 (R Foundation for Statistical Computing, Vienna, Austria). Independent t-test was used to compare means between the 2 groups while Pearson's chi-squared test was used to compare associations between categorical variables. The Kaplan-Meier method was used to estimate patient survival and the log-rank (Mantel-Cox) test to compare the curves. Results were considered statistically significant if the *P* value was less than 0.05.

Table 1 summarises the demographic and clinical characteristics of both groups. In total, 182 patients fulfilled the inclusion criteria, of whom 160 received KRT and 22 underwent CKM. Within the KRT cohort, 104 patients (65%) commenced haemodialysis acutely, of whom 41 (26%) required intensive care treatment. Only 29 patients (18%) had a functioning arteriovenous fistula at the time of first dialysis, while 2 patients were initiated directly onto peritoneal dialysis. Compared with the CKM group, the KRT group was significantly younger with higher rates of diabetes and coronary artery disease, but had higher baseline serum creatinine, phosphate, and lower albumin levels.

By the end of the follow-up period, 55% (n=88) in the KRT group and all 22 patients in the CKM group had died, corresponding to a 45% relative risk reduction in mortality in the KRT group. The mean survival time was 16.69 ± 15.05 months in the KRT group compared with 7.31 ± 8.09 months in the CKM group. The 6-, 12-, 18-, and 24-month survival rates were 82%, 63%, 56%, and 43%, respectively, for patients receiving KRT, compared with 55%, 27%, 23%, and 0% for those managed with CKM. Among all patients, only KRT and the age group of 70–74 years demonstrated significantly better survival. Among patients who died, the KRT group experienced a significantly higher number of hospitalisations, were less likely to die at home and more likely to experience hospital-activated cardiac arrests.

Our study demonstrated that KRT confers a significant survival benefit compared with CKM, despite the KRT group's lower baseline estimated glomerular filtration rate and higher burden of comorbidities. Not surprisingly, existing literature similarly favours a survival advantage among patients who commence dialysis. Reindl-Schwaighofer et al. compared 8622 elderly patients receiving haemodialysis with 174 patients managed with CKM, and found that haemodialysis was associated with a significantly

The *Annals* is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Department of Geriatric Medicine, RIPAS Hospital, Brunei Darussalam

² UBD PAPRSB Institute of Health Sciences, Universiti Brunei Darussalam, Brunei Darussalam

³ Department of Renal Medicine, RIPAS Hospital, Bandar Seri Begawan, Brunei Darussalam

⁴ Department of Internal Medicine, RIPAS Hospital, Bandar Seri Begawan, Brunei Darussalam

Correspondence: Dr Jackson Tan, Department of Renal Medicine, RIPAS Hospital, Jalan Putera Al-Muhtadee Billah, Bandar Seri Begawan, Brunei Darussalam BA1712; Email: drjacksontan74@gmail.com

Accepted: 8 June 2026

Published Online First: 15 June 2026

Table 1. Comparison of demographic and clinical factors between KRT and CKM groups in Brunei Darussalam (2018–2023).

	KRT group (n=160)	CKM group (n=22)	P value
Mean age (years), mean±SD	75.59±4.46	83.50±4.19	<0.05
Median age (years)	75	82	NA
Sex: male, no. (%)	85 (53)	10 (45)	0.07
Sex: female, no. (%)	75 (47)	12 (55)	
Race: Malay, no. (%)	137 (86)	19 (86)	0.85
Race: Chinese, no. (%)	23 (14)	3 (14)	
Comorbidities: DM, no. (%)	135 (84)	11 (50)	<0.05
Comorbidities: Hypertension, no. (%)	157 (98)	22 (100)	0.19
Comorbidities: CAD, no. (%)	56 (35)	3 (14)	<0.05
Creatinine (µmol/L), mean±SD	981.7±356.1	595.9±240.4	<0.05
eGFR (mL/min), mean±SD	5.30±5.87	7.91±3.25	<0.05
Haemoglobin (g/dL), mean±SD	9.21±1.62	9.38±1.75	0.65
Albumin (g/L), mean±SD	30.24±5.75	33.95±6.28	<0.05
Phosphate (mmol/L), mean±SD	2.20±0.74	1.46±0.41	<0.05
6-month survival (%)	82	55	NA
12-month survival (%)	63	27	NA
18-month survival (%)	56	23	NA
24-month survival (%)	43	0	NA
Overall survival (months), mean±SD	16.69±15.05	7.31±8.09	<0.05
Acute dialysis, no. (%)	104 (65)	0 (0)	<0.05
Elective dialysis, no. (%)	56 (35)	2 (11) ^a	<0.05
Using AVF on first dialysis, no. (%)	29 (18)	0 (0)	NA
ICU admission, no. (%)	41 (26)	0 (0)	<0.05
Number of hospitalisations, mean±SD	4.71±3.81	1.70±0.75	<0.05
Dying at home (among those who died), no. (%)	12/88 (14)	8/22 (36)	<0.05
Recorded hospital cardiac arrests (among those who died), no. (%)	15/88 (17)	0/22 (0)	<0.05

AVF: arteriovenous fistula; CAD: coronary artery disease; CKM: conservative kidney management; DM: diabetes mellitus; eGFR: estimated glomerular filtration rate; KRT: kidney replacement therapy; NA: not applicable; SD: standard deviation

^a Two patients had a few sessions of dialysis as part of short-term palliative measures.

lower risk of death (hazard ratio [HR] 0.23, 95% confidence interval [CI] 0.18–0.29; $P<0.001$) compared with CKM.² Likewise, a Southeast Asian cohort from Thailand involving 669 KRT patients and 50 CKM patients reported a mortality HR of 4.97 in the CKM group.³

Though we did not directly assess quality of life, our study uniquely evaluated surrogate indicators—demonstrating fewer hospital admissions, fewer acute

procedural interventions, fewer “unplanned” deaths, and more time spent at home in the CKM cohort. These outcomes may better reflect local Asian religious-cultural priorities surrounding dignity and the sanctity of death. Our findings are consistent with those of Hussain et al.,⁴ who likewise observed that patients receiving KRT were more likely to require acute hospital admissions and to die in hospital. Carson et al. also reported a significantly greater hospitalisation burden

in the KRT group, with patients spending an average of 6.9% of their remaining days as inpatients due to dialysis-related complications, compared with 4.3% in the CKM group.⁵

But, at what cost? This rhetorical question cannot be meaningfully examined through a purely financial lens, especially in Brunei and similar settings where healthcare is free for citizens and permanent residents. Although this removes direct expenses for patients and families, it may foster delayed decision-making about dialysis until late-stage disease. Consequently, many patients present urgently with severe ESKD complications such as fluid overload, metabolic acidosis, and hyperkalaemia, requiring immediate intervention. After stabilisation, some continue dialysis due to the lack of cost barriers, gratitude for life-saving treatment, and early improvements in well-being. However, they may not yet understand the lifelong demands and burdens of ongoing dialysis. The benefit of foresight, informed by the experience of seasoned clinicians, could ideally help guide such vulnerable patients towards a holistic CKM approach.

The 2026 World Kidney Day theme, Kidney Health for All: Caring for People, Protecting the Planet, highlighted the high financial costs, environmental impact, and reduced quality of life often linked to haemodialysis.⁶ It supported practical CKM, especially for elderly and frail patients, by promoting care aligned with personal values and wishes, and reducing unnecessary family, economic, and environmental burdens.

Longevity without dignity is an empty victory. Our study concludes that although KRT increases longevity, it often imposes a heavily encumbered lifestyle with significant psychosocial and medical costs. CKM may offer a meaningful alternative for selected elderly patients by preserving dignity rather than prolonging life. We summarise our position with the principle that medicine should seek to “add life to years rather than add years to life.”

Ethics statement

The study was approved by the Universiti Brunei Darussalam Institutional Health and Research Ethics Committee (Reference: UBD/PAPRSBIHSREC/2022/11).

Declaration

The authors declare that they have no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript. There is no conflict of interest or funding to declare.

Keywords: conservative kidney management, dialysis, geriatric medicine, nephrology, palliative

REFERENCES

1. Sia CSM, Tan JN, Wong ETY. Gazing into the crystal ball of mortality prediction in conservative kidney care. *Ann Acad Med Singap* 2025;54:518-20.
2. Reindl-Schwaighofer R, Kainz A, Kammer M, et al. Survival analysis of conservative vs. dialysis treatment of elderly patients with CKD stage 5. *PLoS One* 2017;12:e0181345.
3. Noppakun K, Tantraworasin A, Khorana J, et al. Survival rates in comprehensive conservative care compared to dialysis therapy in elderly end-stage kidney disease patients: a propensity score analysis. *Ren Fail* 2024;46:2396448.
4. Hussain J, Mooney A, Russon L. Comparison of survival analysis and palliative care involvement in patients aged over 70 years choosing conservative management or renal replacement therapy in advanced chronic kidney disease. *Palliat Med* 2013; 27:829-39.
5. Carson RC, Juszczak M, Davenport A, et al. Is maximum conservative management an equivalent treatment option to dialysis for elderly patients with significant comorbid disease? *Clin J Am Soc Nephrol* 2009 Oct;4:1611-9.
6. Vanholder R, Abdellatif D, Soares Dos Santos AC Jr, et al. Kidney health for all: caring for people, protecting the planet. *Kidney Int* 2026;109:408-17.

Authors' reply

Dear Editor,

We thank the authors for their interest in our editorial and for sharing their experience and outcomes following the implementation of a conservative kidney management (CKM) clinic. We believe it is important to continue to increase awareness and acceptance of CKM as a management strategy for kidney failure, and we commend the authors for contributing to this ongoing conversation.

The authors' observations that patients who were initiated on dialysis had a survival benefit but increased hospitalisations and were more likely to die in hospital are in keeping with the findings in a recent systematic review.¹ Given that a randomised controlled trial would be both methodologically and ethically challenging, well-conducted observational studies with defined outcomes and patient-reported outcome measures will most likely inform the future evidence base guiding shared decision-making regarding CKM versus dialysis.

The authors highlighted that the costs of dialysis and CKM cannot be examined solely from a financial perspective, given the variation in healthcare financing structures and approaches across healthcare systems. While this is certainly true regarding what the patient pays, there have been reports evaluating overall cost differences between dialysis and CKM. Considering the costs associated with outpatient clinic visits, investigations, dialysis treatments, and inpatient hospitalisation, Verberne et al. found that the annual cost for patients aged 70 years and above who opted

for CKM was approximately 20% of that for patients who chose dialysis.² Beyond cost, the same group also evaluated value, and found that CKM achieved similar survival and health-related quality of life outcomes with lower treatment burden and treatment costs, especially in the oldest patients and those with severe comorbidity.³

However, costs and value represent but 1 component of the decision-making framework. For the individual patient, the decision between dialysis and CKM is shaped not only by values but also by perceived survival benefits, symptom burden, quality of life, caregiver burden, and religious or cultural beliefs. A qualitative study conducted with patients recruited from 9 renal units in England found that patients opting for different treatments held contrasting beliefs about the potential advantages of dialysis.⁴ Some dialysis and predialysis patients believed that dialysis could extend their lives and offer them a better quality of life or help them maintain their current quality of life. The prioritisation of survival, even with substantial symptom burden and reduced independence, is an individual choice that transcends financial considerations. Nevertheless, many studies have reported that healthcare providers significantly influence the shared decision-making process, particularly in how CKM is presented and how the patient's overall prognosis is communicated. Patients often perceive a lack of choice as the reason for starting dialysis.⁵ Decision-making is often shaped by differing perspectives among patients, caregivers, and physicians, with the "right decision" seldom perfectly aligned among stakeholders.⁶ While structured frameworks, such as MIND-IT, have been proposed to guide this process,⁷ meaningful shared decision-making ultimately requires something less easily protocolised: time—not merely 15 minutes within the confines of a clinical consultation, but time afforded across multiple dedicated conversations, to unpack complex interventions, understand individual values, and arrive at decisions that are both informed and deeply human. Clinicians need to be trained and spend time presenting kidney failure treatment options accurately and in a balanced manner so that patients appreciate and understand the uncertainties and trade-offs between dialysis and CKM, as highlighted in the KDIGO 2024 Guideline for the Evaluation and Management of CKD.^{8,9} Undergirding this process are the principles of respect and autonomy, where the focus should be on facilitating informed, values-concordant decision-making rather than directing patients towards a predetermined treatment pathway.

The authors' findings are a timely addition to the emerging body of literature on CKM outcomes. We reiterate our call for better-conducted studies,

including patient-reported and caregiver-centred outcomes, to refine the evidence base underpinning shared decision-making in kidney failure care. Equally important will be continued efforts to ensure that treatment discussions are balanced, individualised, and grounded in respect for patient autonomy, recognising that the "best" treatment may differ according to each patient's goals, priorities, and lived experience.

REFERENCES

1. Buur LE, Madsen JK, Eidemak I, et al. Does conservative kidney management offer a quantity or quality of life benefit compared to dialysis? A systematic review. *BMC Nephrol* 2021;22:307.
2. Verberne WR, Ocak G, van Gils-Verrij LA, et al. Hospital Utilization and Costs in Older Patients with Advanced Chronic Kidney Disease Choosing Conservative Care or Dialysis: A Retrospective Cohort Study. *Blood Purif* 2020;49:479-89.
3. Verberne WR, Dijkers J, Kelder JC, et al. Value-based evaluation of dialysis versus conservative care in older patients with advanced chronic kidney disease: a cohort study. *BMC Nephrol* 2018;19:205.
4. Tonkin-Crine S, Okamoto I, Leydon GM, et al. Understanding by older patients of dialysis and conservative management for chronic kidney failure. *Am J Kidney Dis* 2015;65:443-50.
5. Verberne WR, Konijn WS, Prantl K, et al. Older patients' experiences with a shared decision-making process on choosing dialysis or conservative care for advanced chronic kidney disease: a survey study. *BMC Nephrol* 2019;20:264.
6. Buur LE, Bekker HL, Rodkjaer LØ, et al. Decisional needs in people with kidney failure, their relatives and health professionals about end-of-life care options: A qualitative interview study. *J Adv Nurs* 2024;80:3345-58.
7. Toft BS, Rodkjaer L, Andersen AB, et al. Measures used to assess interventions for increasing patient involvement in Danish healthcare setting: a rapid review. *BMJ Open* 2022;12:e064067.
8. Hole B, Rooshenas L, Morton R, et al. 'It's basically "have that or die"': a qualitative study of older patients' choices between dialysis and conservative kidney management. *BMJ Open* 2025;15:e095185.
9. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, Stevens PE, Ahmed SB, et al. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int* 2024;105:S117-314.

Charmaine Si Min Sia^{1,2} MRCP (UK),
Jia Neng Tan^{1,2} MRCP (UK),
Emmett Tsz Yeung Wong^{1,2,3} FRCP (Edin)

¹ Division of Nephrology, Department of Medicine, National University Hospital, Singapore

² Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

³ National University Centre for Organ Transplantation, National University Hospital, Singapore

Correspondence: Dr Emmett Tsz Yeung Wong, Division of Nephrology, Department of Medicine, National University Hospital, Singapore; Email: e.wong@nus.edu.sg

ANNALS, ACADEMY OF MEDICINE, SINGAPORE

81 Kim Keat Road, #11-00 & #12-00 NKF Centre, Singapore 328836

Email: annals@ams.edu.sg | Website: <https://www.annals.edu.sg>