research

IVF procedures and live birth rates p 133

Effect of thymosin α1 in patients with sepsis p 136

Framework for AI deployment p 138

IVF in women with low ovarian reserve or response

ORIGINAL RESEARCH Pragmatic, multicentre, randomised controlled trial

Frozen versus fresh embryo transfer in women with low prognosis for in vitro fertilisation treatment

Wei D, Sun Y, Zhao H, et al

Cite this as: *BMJ* 2025;388:e081474
Find this at doi: 10.1136/bmj-2024-081474

Study question Does the freeze-all strategy (elective freezing of all embryos followed by a planned frozen embryo transfer) increase the chance of live birth compared with fresh embryo transfer in patients with low prognosis for in vitro fertilisation success?

Methods In this pragmatic randomised controlled trial conducted at nine academic fertility centres in China, 838 patients with low prognosis (defined by few oocytes retrieved or poor ovarian reserve) were randomly assigned to either frozen embryo transfer or fresh embryo transfer on the day of oocyte retrieval. Patients in the group for fresh embryo transfer underwent a transfer after oocyte retrieval. Patients in the group for frozen embryo transfer had all of their embryos cryopreserved and underwent a transfer later. The primary outcome was live birth after embryo transfer and analysed in the intention-to-treat population.

Study answer and limitations In women with low prognosis for in vitro fertilisation success, frozen embryo transfer resulted in a lower rate of live birth compared with fresh embryo transfer (32% (132 of 419) v 40% (168 of 419); relative ratio 0.79 (95% confidence interval 0.65 to 0.94); P=0.009). The study was limited by no standardisation of the stimulation protocol, the number or stage of embryos for transfer, or the regimen for endometrial preparation before frozen embryo transfer.

Primary and secondary outcomes				
Outcomes	Frozen embryo group (n=419)	Fresh embryo group (n=419)	Relative ratio (95% CI)	P value
Primary outcome				
Live birth among all women	132 (32)	168 (40)	0.79 (0.65 to 0.94)	0.009
Secondary outcomes				
Singleton live birth among all women	112 (27)	130 (31)	0.86 (0.70 to 1.07)	0.17
Twin live birth among all women	20 (5)	38 (9)	0.53 (0.31 to 0.89)	0.01
Birth weight, g, mean (SD):				
Singleton	3331 (452)	3294 (494)	NA	0.55
Twin	2482 (330)	2390 (586)	NA	0.54
Clinical pregnancy among all women	164 (39)	197 (47)	0.83 (0.71 to 0.97)	0.02
Singleton pregnancy	135 (32)	152 (36)	0.89 (0.74 to 1.07)	0.22
Twin pregnancy	29 (7)	45 (11)	0.64 (0.41 to 1.01)	0.05
Pregnancy loss:				
Total pregnancy loss among biochemical pregnancies	61/196 (31)	50/221 (23)	1.38 (1.00 to 1.90)	0.05
Biochemical pregnancy loss among biochemical pregnancies	30/196 (15)	22/221 (10)	1.54 (0.92 to 2.57)	0.10
Clinical pregnancy loss among clinical pregnancies	31/164 (19)	28/197 (14)	1.33 (0.83 to 2.12)	0.23
First trimester pregnancy loss	29/164 (18)	24/197 (12)	1.45 (0.88 to 2.39)	0.14
Second trimester pregnancy loss	2/164 (1)	4/197 (2)	0.60 (0.11 to 3.24)	0.69
Healthy singleton live birth among all women	99 (24)	105 (25)	0.94 (0.74 to 1.20)	0.63
Cumulative live birth among all women	185 (44)	215 (51)	0.86 (0.75 to 0.99)	0.04

Data are number (%), unless otherwise specified. CI=confidence interval; NA=not available; SD=standard deviation.

COMMENTARY Transfer of fresh embryos may be a better option than use of frozen embryos

ROBERT NG/SOUTH CHINA MORNING POST/GETTY IMAGES

Simone Cornelisse
Simone.Cornelisse@radboudumc.nl
Sebastiaan Mastenbroek
See bmj.com for author details

In vitro fertilisation (IVF) has revolutionised infertility treatment and offers each year hope to millions of couples worldwide. Embryo freezing has become an increasingly prominent part of the treatment. While initially used for the storing of excess embryos after fresh embryo transfers, the so called freeze-all strategy has now become part of IVF, where no fresh transfer and all suitable embryos are frozen for transfer in subsequent menstrual cycles. In a linked research paper, Wei and colleagues studied the effectiveness of this strategy in women who have a poor prognosis of IVF treatment success (defined as nine or fewer oocytes retrieved or a poor ovarian reserve).[1]

One benefit of the freeze-all method is the ability to reduce the risk of ovarian hyperstimulation syndrome; although this risk is generally low in women with poor prognosis owing to a low ovarian response. Another rationale for the freeze-all strategy is to improve pregnancy outcomes by avoiding potential negative effects of ovarian stimulation on endometrial receptivity. By postponing embryo transfer to a subsequent cycle without ovarian stimulation, the endometrium is suggested to be more receptive. However, the process of freezing and thawing embryos is not without risks. Damage may occur during cryopreservation, storage, or thawing, and these steps can add substantial financial costs. Furthermore, treatment delays associated with elective freezing may be undesirable for some.

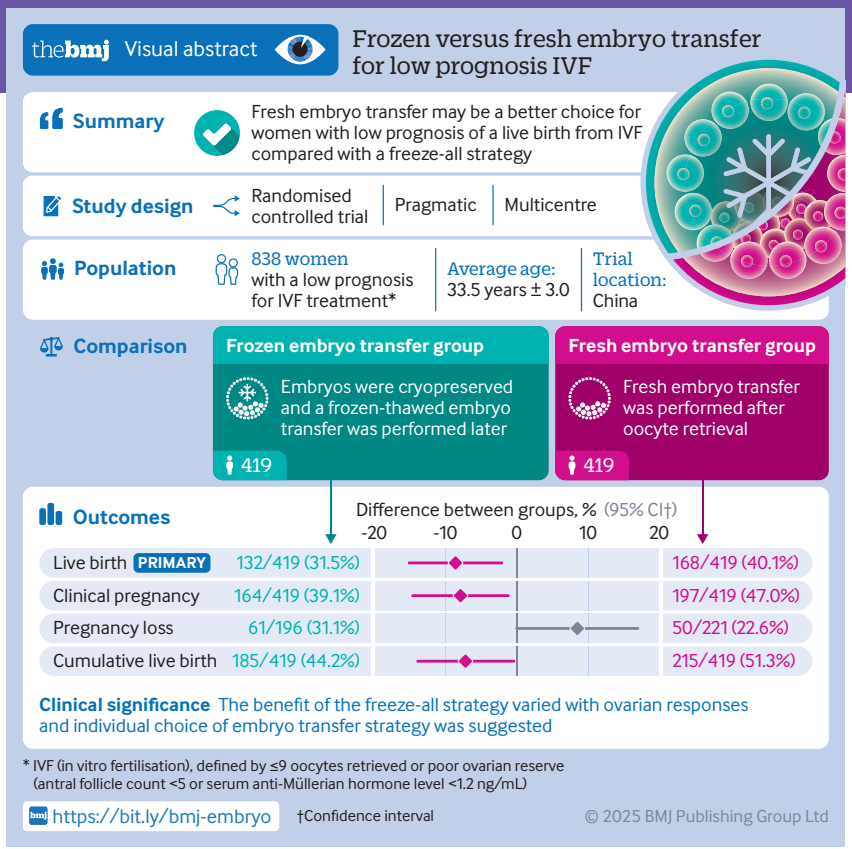
Most studies examining the freeze-all strategy have focused on women with a good prognosis of IVF treatment success, which showed similar cumulative live birth rates between fresh and frozen embryo transfer.[2] However, limited evidence exists regarding its benefits for women with a poor prognosis.[3 4] The study by Wei and colleagues addresses this critical gap. Their multicentre, randomised controlled trial involving 838 participants at

What this study adds The results suggest that fresh embryo transfer may be a better choice in terms of live birth rate for patients with low prognosis of in vitro fertilisation compared with frozen embryo transfer.

Funding, competing interests, and data sharing The funders of the study had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication. For details on competing interests and data sharing, please see bmj.com.

Study registration Chinese Clinical Trial Registry (ChiCTR2100050168).

ANDY WONG/AP/ALAMY



nine fertility centres in China compared a fresh versus frozen embryo transfer strategy in women with an antral follicle count lower than five or serum anti-Müllerian hormone <1.2 ng/mL or fewer than nine oocytes in their IVF treatment. The primary outcome was live birth rate per first transfer, while secondary outcomes included cumulative live birth rates within one year of randomisation. The findings showed a lower live birth rate (risk ratio 0.79 (95% confidence interval 0.65 to 0.94)) and a lower cumulative live birth rate in the frozen embryo transfer group (0.86 (0.75 to 0.99)) than in the fresh embryo transfer group.

Choosing live birth rate

Concerns remain about potential biases that could have influenced outcomes, such as variations in the day of embryo transfer, differences in the number of double embryo transfers, and a small number of women who underwent another oocyte retrieval to obtain more embryos. The choice of live birth rate per first transfer as the primary outcome is more often seen in trials in reproductive medicine, but the cumulative live birth rate—which considers all embryo transfers from a single oocyte retrieval cycle—is arguably more relevant from a patient centred perspective.[5-7] Although the study was not powered to assess secondary outcomes, reporting them adds valuable information to the available literature. No differences were observed in obstetrical-neonatal outcomes between the groups, where other studies suggested risks may be increased in the freeze-all group.[8-10]

The mandatory freeze-all component in these contexts may not provide benefit to these women

Previous evidence on the effectiveness of fresh versus frozen embryo transfer in IVF, including a Cochrane review, suggested no clear superiority of one strategy over the other in terms of cumulative live birth rates.[2] Among the studies included in the review, only one of eight reported lower cumulative ongoing pregnancy or live birth rates for the freeze-all strategy compared with fresh transfer.[11] Unlike the other studies, which predominantly included women with a good prognosis of IVF success, that study also included women with a poor prognosis; although, defined as no pregnancy after a period of expectant management. These studies emphasise the importance of providing high level evidence to tailor IVF strategies to meet individual patient characteristics.

Clinical implications

Wei and colleagues' trial offers valuable insights for women with a poor prognosis in IVF.[1] The study reported lower live birth rates in the freeze-all group, with no differences in neonatal outcomes, suggesting that fresh embryo transfer may be a better strategy for these patients. These results have broader implications, particularly for centres offering advanced IVF treatments such as pre-implantation genetic testing for aneuploidy or embryo banking, which involves freezing embryos from multiple IVF cycles before a first transfer. These strategies include freezing of all embryos and are often offered to women with a poor prognosis, such as those of advanced maternal age, to address declining oocyte quality and numbers. However, both approaches remain controversial, with little evidence supporting their efficacy.[12] The findings from the current study suggest that the mandatory freeze-all component in these contexts may not provide benefit to these women.

Rigorous evaluation of these strategies is needed. Any potential advantages must outweigh drawbacks, such as the lower cumulative live birth rates linked to skipping a fresh embryo transfer. Properly assessing the effectiveness of these techniques is essential for improving outcomes in this challenging patient population.

Cite this as: *BMJ* 2025;388:r49

Find the full version with references at http://dx.doi.org/10.1136/bmj.r49

Immunotherapies for sepsis and the impact of study design

ORIGINAL RESEARCH Multicentre randomised controlled trial

The efficacy and safety of thymosin α1 for sepsis (TESTS)

Wu J, Pei F, Zhou L, et al

Cite this as: *BMJ* 2025;388:e082583
Find this at doi: 10.1136/bmj-2024-082583

Study question Can the immunomodulatory drug thymosin α1 reduce mortality in adults with sepsis?

Methods This multicentre, double blinded, placebo controlled phase 3 trial was conducted across 22 centres in China from September 2016 to December 2020 in adults (18-85 years) with sepsis, diagnosed according to Sepsis-3 criteria. Participants were randomly assigned to receive either subcutaneous thymosin α1 or placebo every 12 hours for seven days. Randomisation was stratified by age (<60 and ≥60 years) and centre using a stratified block method. The primary outcome was all cause mortality 28 days after randomisation.

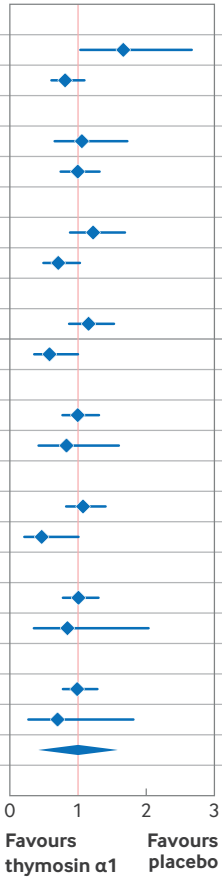
Study answer and limitations Of 1106 adults with sepsis enrolled in the study, 1089 were included in the modified intention-to-treat analyses (thymosin α1 group n=542, placebo group n=547). 28 day all cause mortality occurred in 127 participants (23.4%) in the thymosin α1 group and 132 (24.1%) in the placebo group (hazard ratio 0.99, 95% confidence interval 0.77 to 1.27; P=0.93 with log rank test). The prespecified subgroup analysis showed a potential differential effect of thymosin α1 on the primary outcome based on age (<60 years: hazard ratio 1.67, 1.04 to 2.67; ≥60 years: 0.81, 0.61 to 1.09; P for interaction=0.01) and diabetes (diabetes: 0.58, 0.35 to 0.99; no diabetes: 1.16, 0.87 to 1.53; P for interaction=0.04). The study was not able to determine the precise times from onset of sepsis to diagnosis because of individual variability and disease progression.

What this study adds This study found no conclusive evidence that thymosin α1 reduces 28 day mortality in adults with sepsis. Future research on thymosin α1 in sepsis should consider the heterogeneity of the disease, particularly focusing on participants aged 60 and older and on those with chronic conditions.

Funding, competing interests, and data sharing
Funded by Sun Yat-sen University Clinical Research Program, Guangdong Clinical Research Center for Critical Care Medicine, and SciClone Pharmaceuticals. Some authors reported grants from SciClone Pharmaceuticals and consultancy fees from various pharmaceutical companies. For details on data sharing, please see bmj.com.

Trial registration ClinicalTrials.gov NCT02867267.

Subgroup	Thymosin α1 group: No of patients	Thymosin α1 group: 28 day mortality	Placebo group: No of patients	Placebo group: 28 day mortality	Hazard ratio (95% CI)	P for interaction	Hazard ratio (95% CI) for interaction
Age (years)						0.01	0.49 (0.29 to 0.86)
<60	212	20.7 (15.3 to 26.2)	215	13.5 (8.9 to 18.1)	1.67 (1.04 to 2.67)		
≥60	330	25.2 (20.5 to 29.8)	332	31.0 (26.1 to 36.0)	0.81 (0.61 to 1.09)		
Sex						0.97	1.01 (0.58 to 1.75)
Female	182	20.9 (15.0 to 26.8)	157	21.7 (15.2 to 28.1)	1.06 (0.65 to 1.73)		
Male	360	24.7 (20.3 to 29.2)	390	25.1 (20.8 to 29.4)	0.99 (0.74 to 1.32)		
Hypertension						0.06	0.62 (0.37 to 1.01)
No	321	23.7 (19.0 to 28.3)	347	20.5 (16.2 to 24.7)	1.22 (0.88 to 1.69)		
Yes	221	23.1 (17.5 to 28.6)	200	30.5 (24.1 to 36.9)	0.71 (0.49 to 1.04)		
Diabetes mellitus						0.04	0.54 (0.30 to 0.96)
No	390	25.6 (21.3 to 30.0)	422	23.0 (19.0 to 27.0)	1.16 (0.87 to 1.53)		
Yes	152	17.8 (11.7 to 23.8)	125	28.0 (20.1 to 35.9)	0.58 (0.35 to 0.99)		
Solid malignant tumours						0.60	0.83 (0.42 to 1.64)
No	483	22.6 (18.8 to 26.3)	485	22.5 (18.8 to 26.2)	1.00 (0.77 to 1.31)		
Yes	59	30.5 (18.8 to 42.3)	62	37.1 (25.1 to 49.1)	0.83 (0.42 to 1.60)		
Coronary heart disease						0.06	0.49 (0.24 to 1.03)
No	477	23.7 (19.9 to 27.5)	484	22.7 (19.0 to 26.5)	1.08 (0.83 to 1.41)		
Yes	65	21.5 (11.5 to 31.5)	63	34.9 (23.2 to 46.7)	0.47 (0.21 to 1.01)		
COPD						0.60	0.80 (0.37 to 1.81)
No	505	23.0 (19.3 to 26.6)	496	23.2 (19.5 to 26.9)	1.00 (0.77 to 1.30)		
Yes	37	29.7 (15.0 to 44.5)	51	33.3 (20.4 to 46.3)	0.84 (0.35 to 2.04)		
Chronic kidney disease						0.76	0.87 (0.36 to 2.12)
No	514	23.0 (19.3 to 26.6)	511	23.1 (19.4 to 26.8)	0.99 (0.77 to 1.29)		
Yes	28	32.1 (14.8 to 49.4)	36	38.9 (23.0 to 54.8)	0.70 (0.27 to 1.81)		
Total	542	23.4 (19.9 to 27.2)	547	24.1 (20.6 to 27.9)	0.97 (0.76 to 1.24)		



Subgroup analyses. A Bonferroni threshold for significance for overall type I error of 0.05 was P=0.006. The HRs for relative risk of primary outcome of the two groups and associated 95% CIs were calculated with a Cox regression model adjusting for centre and age. An interaction term was added between treatment and subgroup in the Cox regression model. CI=confidence interval; COPD=chronic obstructive pulmonary disease; HR=hazard ratio

Sepsis is a life threatening syndrome initiated by micro-organisms.[1] Severe and often lethal injuries (eg, shock, multiorgan failure, and metabolic derangements) are the manifestation of the syndrome, with systemic inflammatory cascade activation and circulating mediators thought to cause host injury.

Scores of trials evaluating host-immune modifiers, many of which were the initial ventures of a new biotech industry 30 years ago, did not improve patient outcomes in sepsis.[2] Undaunted by this history, investigators in the linked trial evaluated another immune therapy, thymosin α1, in patients with sepsis, enrolling 1106 participants at 22 sites in China.[3] The trial compared thymosin α1 with placebo plus usual care in a blinded randomised trial evaluating all cause mortality at 28 days. The results showed 24% mortality with thymosin versus 23% in the control group (hazard ratio 0.99, 95% confidence interval 0.77 to 1.77), which indicate no improvement with thymosin. Several positive aspects of the trial's design include a superiority hypothesis with direct patient outcomes of all cause mortality using inferential statistics, rather than non-inferiority hypotheses on surrogate endpoints using descriptive statistics commonly used in infection trials.[4-9]

Yet, questions remain regarding the trial's methods. Were the types and numbers of patients enrolled requisite for studying this syndrome? In sepsis, clinical and preclinical evidence show that potential benefits of immunotherapies may be limited to patients with a higher baseline risk of sepsis mortality—ie, people with immune dysregulation that could cause death.[10] In patients with lower baseline mortality risk, suppressing appropriately functioning immune responses may be harmful and worsen outcomes.[11] Study investigators hypothesised 35% mortality in the control group, yet results showed 23% mortality, indicating a less sick population with most participants having appropriate immune responses. A systematic review of mortality rates in randomised trials of sepsis between 1991 and 2013 showed that while mortality has continuously declined, when controlling for baseline severity of illness, mortality has not changed over time. These data suggest that more recent sepsis trials enrolled patients who may not benefit.[12]

Surrogate outcomes instead of survival are not informative in sepsis trials

Trial design considerations

Because of this potential paradoxical situation of immunotherapy, even in higher baseline risk populations, to detect statistically significant effects greater than chance may require enrolling larger numbers of participants than included in the linked study. Numbers of participants required is based partly on the amount of benefit of interventions. Small effects often are not worth detecting clinically with an exception for all cause mortality in life threatening syndromes such as sepsis.

Surrogate outcomes instead of survival are not informative in sepsis trials given the present understanding of the syndrome. Sequential organ failure assessment (SOFA) scores are composites of biomarkers not on the causal pathway of disease.[13] Changes in these scores have previously been shown not to reflect most treatment effects on patient survival.[14] Despite patient level correlations with mortality, these sequential organ failure assessment scores are not valid trial level surrogates and are not useful endpoints for studies in serious lethal syndromes.

Despite the lack of success with host directed treatments in sepsis, the effort is worth pursuing. Paul Ehrlich studied several hundred compounds before finding salvarsan for syphilis.[15] Host directed therapies, such as steroids, have shown consistent patient benefits in improving septic shock and potentially for survival in people with severe pneumonia and covid-19.[16 17]

The focus of resources and policy in infectious disease on small molecule drugs with in vitro biological activity against antimicrobial resistant pathogens in pathogen focused development[18 19] misses the point. Seventeen of 18 deaths (94%) occurred in patients with organisms susceptible in vitro to current drugs.[20] Poor outcomes may be due to an inaccurate early diagnosis but also due to severe dysregulated immune responses, not an inability to inhibit bacterial growth. Therefore, drugs with improved in vitro potency at growth inhibition are worth developing but will not benefit all types of patients who are seriously ill with an infection if immune dysregulation is the problem. Host directed therapies may benefit the greater number of patients with susceptible as well as antimicrobial resistant disease.

New methods needed

This study emphasises again that research methods matter when studying medical interventions. Quoting Einstein, the definition of insanity is doing the same thing over and over and expecting different results. Beneficial interventions can be discarded if not studied appropriately. Improving patient outcomes in sepsis may require studying patients with specific highly lethal infections by site of infection (pneumonia *v* urinary tract infections) instead of a broad array of populations pooled together into a poorly understood syndrome. These methods may allow future studies to expand the types of interventions evaluated, including host directed therapies, enrolling patients sick enough to benefit in site specific infections. Hypotheses about immune therapies have not been adequately evaluated given the methods used to date. Broadly anti-inflammatory interventions with promising data in preclinical and phase 2 studies need testing in patients at high risk of sepsis attributable mortality. Enough participants should be enrolled in randomised confirmatory trials using superiority hypotheses about direct patient outcomes analysed by inferential statistics. The results will be worth the effort both for patients with sepsis and for justifying the costs of treatments.

John H Powers
john.powers@nih.gov
Charles Natanson
See bmj.com for author details

Cite this as: *BMJ* 2025;388:r48
Find the full version with references at http://dx.doi.org/10.1136/bmj.r48

FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare

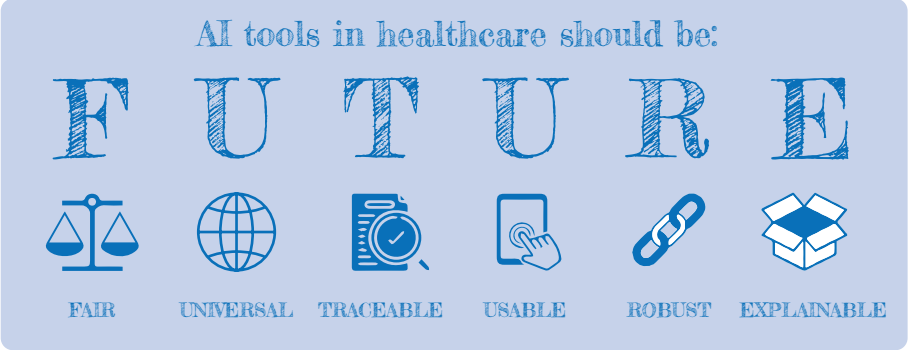
Lekadir K et al; on behalf of the FUTURE-AI Consortium

Cite this as: *BMJ* 2025;388:e081554
Find this at doi: 10.1136/bmj-2024-081554

Artificial intelligence (AI) is transforming healthcare, promising breakthroughs in disease diagnosis, treatment planning, and patient care. However, adopting AI tools in clinical practice has been challenging owing to concerns about their trustworthiness, safety, and reliability.

The FUTURE-AI Consortium, comprising 117 experts from 50 countries and a wide range of disciplines, has developed an international guideline to promote the creation and deployment of AI tools that are trustworthy by design. This global framework is built around six core principles: fairness, ensuring equitable performance across all groups; universality, enabling AI tools to work in diverse healthcare settings; traceability, providing mechanisms to monitor the AI tools over time; usability, prioritising AI tools that are centred on the users; robustness, ensuring reliability despite variations in real world data; and explainability, making AI decisions understandable for users.

The guideline includes 30 best practices covering the AI lifecycle, from initial design to real world deployment. For example, they emphasise the importance of training AI systems on representative datasets to reduce biases and conducting rigorous evaluations across varied environments to ensure broad applicability. Continuous monitoring and human oversight are also recommended to maintain AI performance and reliability. To ensure practical implementation, the framework provides step-by-step guidance tailored for diverse stakeholders, including developers, evaluators, deployers, and end users, ensuring its recommendations can be operationalised effectively. One of the framework's highlights is its emphasis on collaboration. By involving healthcare professionals, patients, ethicists, and other stakeholders, FUTURE-AI promotes the co-creation of AI tools that align with real world needs and ethical standards.



Organisation of the FUTURE-AI framework for trustworthy artificial intelligence (AI) according to six guiding principles—fairness, universality, traceability, usability, robustness, and explainability

The FUTURE-AI framework is an important step towards fostering trust in medical AI, paving the way for its safe and effective integration into healthcare systems worldwide. For more information on the FUTURE-AI recommendations, see the full publication on bmj.com.

List of FUTURE-AI recommendations, together with the expected compliance for both research and deployable artificial intelligence (AI) tools (+: recommended, ++: highly recommended)

Recommendations	Research	Deployable
Fairness		
1. Define any potential sources of bias from an early stage	++	++
2. Collect information on individuals' and data attributes	+	+
3. Evaluate potential biases and, when needed, bias correction measures	+	++
Universality		
1. Define intended clinical settings and cross setting variations	++	++
2. Use community defined standards (eg, clinical definitions, technical standards)	+	+
3. Evaluate using external datasets and/or multiple sites	++	++
4. Evaluate and demonstrate local clinical validity	+	++
Traceability		
1. Implement a risk management process throughout the AI lifecycle	+	++
2. Provide documentation (eg, technical, clinical)	++	++
3. Define mechanisms for quality control of the AI inputs and outputs	+	++
4. Implement a system for periodic auditing and updating	+	++
5. Implement a logging system for usage recording	+	++
6. Establish mechanisms for AI governance	+	++
Usability		
1. Define intended use and user requirements from an early stage	++	++
2. Establish mechanisms for human-AI interactions and oversight	+	++
3. Provide training materials and activities (eg, tutorials, hands-on sessions)	+	++
4. Evaluate user experience and acceptance with independent end users	+	++
5. Evaluate clinical utility and safety (eg, effectiveness, harm, cost-benefit)	+	++
Robustness		
1. Define sources of data variation from an early stage	++	++
2. Train with representative real world data	++	++
3. Evaluate and optimise robustness against real world variations	++	++
Explainability		
1. Define the need and requirements for explainability with end users	++	++
2. Evaluate explainability with end users (eg, correctness, impact on users)	+	+
General		
1. Engage interdisciplinary stakeholders throughout the AI lifecycle	++	++
2. Implement measures for data privacy and security	++	++
3. Implement measures to address identified AI risks	++	++
4. Define adequate evaluation plan (eg, datasets, metrics, reference methods)	++	++
5. Identify and comply with applicable AI regulatory requirements	+	++
6. Investigate and address application specific ethical issues	+	++
7. Investigate and address social and societal issues	+	+