UNKNOWN LAB / 論文とETIC #7

27項目を報告できても、臨床AIが安全だとは言えない

151人で作ったDECIDE-AIを、患者代表5人・費用除外・品質保証の境界まで読む

Daiki Morimoto / Unknown Lab / ORCID 0009-0009-6444-750X
批判的論評 Version 1.2 2026年9月3日
DOI: 10.5281/zenodo.22242002

要旨

DECIDE-AIは、AIによる意思決定支援を小規模な実臨床で評価する研究の最低報告項目を、2ラウンドmodified Delphiで作成した。18か国・20 stakeholder groupsの151人が参加し、AI固有17項目・28下位項目と一般10項目を提示した。version、入力、利用者training、workflow、上書き、エラー、間接的害、learning curve、変更履歴を可視化する点は重要である。

ただし報告指針は、研究の方法論的品質、安全性、有効性を保証しない。参加者は英国79人、患者代表5人、allied health 10人で、合意会議のallied healthは0人だった。health economicsの2項目、trust、interpretabilityも最終項目に入らなかった。本稿は、ETIC 2.0の考え方と重なる点、答えるべき問い、報告・測定・判断・停止を分ける具体的な見直し案を示す。

臨床AIの研究報告に、27のチェック項目がすべて書かれていたとする。

対象者、AIの版、入力データ、画面、利用者のtraining、ワークフロー、エラー、上書き、学習曲線、システム変更まで記録されている。従来より、はるかに透明な報告である。

それでも、そのAIが安全だとはまだ言えない。

チェックリストが示すのは、何を読者に見えるようにするかである。研究の方法が妥当だったか、害を見落としていないか、患者の生活が良くなったか、費用と負担が受け入れられるかを、自動的に判定するものではない。

今回読むDECIDE-AIは、AIによる意思決定支援を小規模な実臨床で評価するときの報告指針である。2ラウンドのmodified Delphiには、18か国・20のstakeholder groupsから151人が参加した。最終的に17のAI固有項目、28の下位項目、10の一般項目が残った。[1]

臨床AIをアルゴリズム単体ではなく、患者、利用者、画面、ワークフロー、最終判断を含む複雑な介入として見る。この視点は、ETIC 2.0の検証に重要である。

同時に、合意の構成には偏りがある。151人のうち英国が79人、臨床家が68人、工学・計算機科学が50人だった。患者代表は5人、allied healthは10人である。費用の2項目は途中で外れ、trustとinterpretabilityも最終項目に入らなかった。

本稿は、DECIDE-AIを便利なチェックリストとして紹介するだけではない。誰が項目を作り、どう絞り、何を検証し、何を検証していないのかまで読む。

2022年の訂正では、原著で漏れていたDECIDE-AI expert groupの構成員と所属が追加された。方法、人数、チェックリスト、結果の変更は報告されていない。[4]

30秒で分かる研究の要点

項目	内容
研究	早期の実臨床AI評価を報告するためのDECIDE-AI指針開発
対象	2ラウンド全体で151人、18か国、事前定義した20stakeholder groups
期間	2020年8月から2021年9月。合意会議は2021年6月14-16日
方法	2ラウンドmodified Delphi、16人の合意会議、別の16人による質的評価
定量結果	第1分析120人、第2分析136人、43,986語、13,520得点、1,151コメント
成果物	AI固有17項目・28下位項目＋一般10項目。合計27報告項目
直接言えること	専門家合意に基づく最低報告項目と、その作成過程が示された
言えないこと	DECIDE-AI準拠研究の方法が良い、AIが安全・有効・公平である
重要な限界	英国52%、患者代表3%、allied health 7%。費用項目は最終版から除外

13,520得点は、第1ラウンド6,419件と第2ラウンド7,101件の合計である。1,151コメントも228件と923件を合計した。本稿での再集計であり、独立した評価者数や臨床事例数を意味しない。

English Summary

DECIDE-AI used a two-round modified Delphi process involving 151 experts from 18 countries and 20 stakeholder groups to develop minimum reporting recommendations for early-stage live clinical evaluation of AI decision-support systems. The final guideline contains 17 AI-specific items with 28 subitems and 10 generic items. It makes system version, inputs, user training, workflow integration, human-AI agreement, errors, indirect harm, learning curves, and modifications visible.

Reporting completeness is not proof of methodological quality, safety, effectiveness, or equity. UK participants comprised 52% of the panel, only five of 151 participants were patient representatives, and no allied health professional was in the 16-member consensus group. Two health-economic items were removed, while trust and interpretability were excluded because accepted measures were lacking. This commentary proposes separating reporting, measurement, decision thresholds, and stopping or remedy procedures in ETIC 2.0.

Keywords: DECIDE-AI, clinical artificial intelligence, reporting guideline, Delphi consensus, human factors, patient involvement, AI governance, ETIC 2.0

本稿の位置づけ

本稿は公開済み査読論文、補足資料、checklist、Explanation & Elaboration、訂正をUnknown Labが批判的に読み解いた未査読の論評である。著者の結論、観察結果、因果効果として言えないこと、Unknown Labの論評、ETIC 2.0の考え方と重なる点、答えるべき問い、具体的な見直し案を分ける。個別の診断、治療、AI製品の安全認証や導入を指示しない。

補強記事08から、何を一歩深めるのか

補強記事08「ETIC 2.0には何が足りないのか」では、ETIC 2.0を検証済み尺度や臨床ガイドラインとして扱わず、公開された設計仮説として7つのゲートを通す必要があると整理した。[5]

そのゲートには、内容妥当性、使用可能性、既存手法との比較、前向き実装、外部検証、公平性、変更管理が含まれる。

DECIDE-AIは、前向き実装の手前にある重要な問いへ答える。

小規模な実臨床評価で、何を報告すれば、第三者がAIの使われ方を理解できるのか。

AIの名称と精度だけでは足りない。誰が使ったか。どの版か。何を入力したか。欠損をどう扱ったか。出力をどの画面で見せたか。最終判断は誰が行ったか。利用者はAIへ従ったか、上書きしたか。エラーと間接的な害は何か。途中でシステムを変えたか。

DECIDE-AIは、これらを一つの報告構造にする。

しかし記事08の検証ロードマップを完成させるには、もう一歩必要である。

報告項目が書かれたかだけでなく、その測定は妥当か、誰が進行を判断するか、何が起きたら止めるかを決めなければならない。

どの段階の研究を対象にするのか

DECIDE-AIの対象は、preclinical performance testと大規模な比較試験の間にある。

開発者が過去データだけで精度を測る段階ではない。実際の患者ケアへ影響する環境で、医療者や他の利用者がAIと接し、意思決定の一部へ組み込む段階である。

この段階を著者らはearly-stage live clinical evaluationと呼ぶ。

目的は、小規模に臨床的な有用性を探索し、安全性、人間要因、ワークフローを確認し、その後の大規模試験へ進む準備をすることである。

CONSORT-AIやSPIRIT-AIは、主としてAI介入の無作為化比較試験と、そのプロトコルの報告を扱う。DECIDE-AIはその前にある、探索的で変更を含み得る段階を埋める。[2]

ここでは比較群を必須にしない。研究デザインも一つに固定しない。初期研究は、臨床的有効性を検出する十分な人数を持たないことがあると認める。

これは弱点だけではない。

AIをいきなり大規模試験へ持ち込む前に、画面、入力、training、役割分担、エラー、学習曲線を小さく調べる必要がある。実装しながら明らかになる問題もある。

ただし、単群の小規模研究を丁寧に報告しても、比較効果は生まれない。探索研究を正確に探索研究として読めるようにすることが、DECIDE-AIの役割である。

誰が参加し、どう選ばれたのか

研究チームは、参加者を5つの経路から集めた。

- Steering Groupが推薦した専門家。
- 初期の文献検索で見つかった論文の著者。
- 医学誌のcommentaryに掲載した公開募集。
- 自発的に研究チームへ連絡した専門家。
- 参加者からのsnowball推薦。

募集前に20のstakeholder groupsを定義した。臨床家や工学者だけではない。患者代表、allied health、管理職、実装科学、倫理、規制、政策、支払者、資金提供者、民間部門等を含む。

第1ラウンドは138人が参加に同意し、123人が完了した。分析したのは120人である。1人は評価尺度を逆に使った疑い、2人は期限後回答のため除外された。

第2ラウンドは162人を招待し、138人が完了、136人を分析した。第1ラウンドから継続した人は110人、新規参加は28人だった。2ラウンド全体のunique expertsは151人である。

「151人、18か国、20群」という要約は正しい。だが、人数の配分も見る必要がある。

構成	全151人	合意会議16人
英国	79（52%）	9（56%）
米国	24（16%）	3（19%）
臨床家	68（45%）	8（50%）
工学・計算機科学	50（33%）	4（25%）
方法論	30（20%）	6（38%）
患者代表	5（3%）	1（6%）
allied health	10（7%）	0
支払者	1（1%未満）	0

職種・役割は複数回答であり、合計は151人または16人にならない。

英国の参加者だけで半数を超えた。患者代表は5人、合意会議では1人だった。allied healthは全体で10人いたが、合意会議にはいなかった。

少人数だから発言が無意味だったとは言えない。患者代表1人が重要な項目を動かすこともある。しかし、人数、発言、決定権は別である。最終項目が、患者、家族、生活支援職の優先順位を十分代表するとまでは言えない。

著者らも、欧州、とくに英国への偏り、臨床家と工学者の過剰代表を限界として明記した。自発的に連絡した専門家30人のうち25人、83%が欧州・英国からだったことも説明している。

いつ、どのように項目を絞ったのか

開発は2020年8月に始まり、2021年9月まで続いた。

初期リストは54項目だった。関連する報告指針や文献を調べ、Steering Groupが候補を作った。

第1ラウンドでは、4つの自由記述質問と、各項目の9点評価を使った。1-3は重要でない、4-6は重要だがcriticalではない、7-9はcriticalと分類した。

自由記述は2人が独立にテーマ分析し、不一致を合意で解決した。

段階	量と変化
第1ラウンド	43,986語、6,419得点、228コメント、64の新提案、109テーマ
第1後	9項目維持、22修正、21を13へ再編、2削除、9追加、最終53項目
第2ラウンド	7,101得点、923コメント
合意会議	2021年6月14-16日の3回、16人、匿名投票、採用閾値80%
最終構成	AI固有17項目・28下位項目、一般10項目

合意会議では、第2ラウンドの全項目を議論した。Vevoxを使って匿名投票し、blankとabstentionを除いた80%以上を採用条件とした。

80%は、事前に決めた意思決定規則である。

80%を超えたから、その項目が臨床的に妥当だと統計的に証明されたわけではない。参加者構成が変われば、投票結果も変わる可能性がある。合意と妥当性は重なるが、同じではない。

16人の質的評価は、何を確かめたのか

合意会議の後、別の16人がAI固有項目を評価した。この16人は合意会議の構成員ではなく、AIシステムの実装または関連研究の査読経験を持っていた。

評価したのは、各項目のclarityとapplicabilityである。custom formへの回答を2人が独立に分析し、不一致を合意で解決した。その結果をもとに、項目文言とExplanation & Elaborationを修正し、Consensus Groupが最終承認した。

この工程には価値がある。

専門家会議で意味が通じても、文章だけを読んだ第三者には曖昧なことがある。E&Eには各項目の理由と例が加わり、単なる短いチェック欄より理解しやすくなった。

一方、質的結果の報告には限界がある。

本文は、16人の評価を受けたこと、2人が分析したこと、文言を変えたことを示す。しかし、最終的なテーマ名、各テーマを挙げた人数、採用しなかった意見、反対意見の分布は詳しく読めない。

また、実際の研究者がチェックリストを使い、何分かかったか、どこで迷ったか、評価者間で同じ判定になったか、重要な欠落を見つけられたかを試した使用可能性研究ではない。

27項目は、何を見えるようにするのか

DECIDE-AIの最終版は、17のAI固有報告項目と10の一般項目からなる。AI固有17項目は28の下位項目を含む。[3]

研究段階	主なAI固有項目	見えるようにすること
Title/Introduction	1 Title、2 Intended use	問題、対象患者、利用者、ケア経路、意図する影響
Methods	3 Participants、4 AI system、5 Implementation	患者・データ・利用者、training、version、入力、欠損、出力画面、最終判断
Methods	6 Safety/errors、7 Human factors、8 Ethics	エラー定義、害の探索、人間要因手法、公平性等の方法
Results	9 Participants、10 Implementation、11 Modifications	missingness、曝露、adherence、workflow変化、システム変更
Results	12 Human-computer agreement、13 Safety/errors、14 Human factors	従う・上書きする理由、直接・間接的害、usability、learning curve
Discussion	15 Intended-use support、16 Safety/errors	使用目的を支持する範囲、安全profileと緩和策
Statements	17 Data availability	dataとcodeの可用性

一般10項目は、structured abstract、目的、protocol・登録・倫理、outcomes、統計解析、patient involvement、main results、subgroup、strengths and limitations、conflicts of interestを扱う。

重要なのは、AIの出力だけでなく、人がどう扱ったかを記録する点である。

利用者はAIの推奨へ同意したのか。違う判断をしたのか。AIを見て考えを変えたのか。どのようなエラーが起き、修正できたのか。間接的な害はあったか。使うほど慣れたのか。途中でversionやhardwareを変えたのか。

これらがなければ、同じAIを別の現場へ移したとき、何を再現すべきか分からない。

報告指針と、研究の質を分ける

ここが本論文の最も重要な読みどころである。

DECIDE-AIはreporting guidelineである。

研究者へ、最低限何を報告するかを示す。特定の研究デザインを命じず、比較群を常に必須にせず、各測定法の良否を採点しない。

E&Eも、チェックリストへのadherenceを研究のmethodological qualityと同一視すべきではないと注意する。

たとえば、ある研究が次のように報告したとする。

- 患者は20人だった。
- 比較群はなかった。
- AIのversionを記録した。
- 利用者は2時間trainingを受けた。
- エラーを研究者の目視で探した。
- 有害事象は報告されなかった。

この報告は、人数や方法を隠す研究より透明である。しかし、20人で十分か、目視だけで害を見つけられるか、観察期間が適切か、比較群なしで有効性を言えるかは別に評価しなければならない。

報告されていることと、良い方法で実施されていることは違う。

さらに、報告されていないことと、実施されなかったことも違う。論文に書かれていないだけかもしれない。

DECIDE-AIは、この不透明さを減らす。しかし、透明になった結果をどう判定するかまでは、別の方法論が必要である。

費用の2項目は、なぜ外れたのか

第1ラウンドで削除された2項目はhealth economic assessmentに関するものだった。

2項目ともmedian 6、IQR 2-9で、初期項目中、中央値が7未満だった唯一の項目だった。多くのコメントが、経済評価はまったく別の評価領域だと述べたため、最終候補から外れた。

この経緯を「専門家は費用を重要でないと判断した」と要約してはいけない。

中央値6は「重要だがcriticalではない」に当たる。IQR 2-9は意見の広がりも大きい。外れた理由は、早期臨床AIの最低報告項目へ入れるより、別の経済評価として扱うべきだというscopeの判断である。

しかし、現場でAIを続けられるかを考えると、費用は後回しにできない。

誰が入力を支えるのか。専門職は何分確認するのか。誤りを誰が調べるのか。version変更後の再評価を誰が担うのか。障害時に非AI経路へ戻す人員はいるか。患者や家族の無償作業は増えないか。

経済評価を別にすることと、費用・人員を見えなくすることは違う。

ETIC 2.0では、精緻な費用効果分析を初期研究へ常に求めなくても、時間、人員、無償負担、保守、監視、事故対応を最低限記録する必要がある。

trustとinterpretabilityが外れた意味

interpretabilityも最終項目に入らなかった。

一部の専門家は、AI出力の臨床価値は解釈可能性と独立し得ると論じた。また、interpretabilityを定量化・評価する一般に受け入れられた方法がないことも理由になった。

trustも議論された。

利用者はAIを使い、実世界のfeedbackを受けるほど、信頼の水準を変える。その信頼が適切でも不適切でも、AIの推奨が最終判断へ与える影響を変える。しかし臨床AIでtrustを測る共通の方法がなかったため、最終項目へ入らなかった。

著者ら自身、これは合意形成が保守的になりやすいことを示すと書く。広く同意できる概念だけが閾値を超えるため、重要だが測定が未成熟な論点は落ちやすい。

測り方が決まっていないことは、現象が存在しないことではない。

ETIC 2.0では、trustを一つの点数だけで捉えない方がよい。AIに従った割合、上書きした割合、理由、確信度、誤りを経験した後の変化、使わない選択、専門職へ確認した行動を組み合わせる。interpretabilityも、すべてのモデルに一つの説明形式を求めるのではなく、誰が、どの判断で、何を理解する必要があるかから設計する。

著者の結論、観察結果、因果効果を分ける

著者の結論

著者は、151人・18か国・20群のconsultation and consensusを通じ、早期臨床AI評価で最低限報告すべき項目を作成したと結論する。DECIDE-AIが研究のappraisalとreplicabilityを促し、大規模試験と導入への基盤になることを期待している。

研究から直接確認できる観察結果

- 2ラウンド全体で151人が参加した。
- 第1ラウンドは120人、第2ラウンドは136人の回答を分析した。
- 54の初期項目が、AI固有17項目・28下位項目と一般10項目へ整理された。
- health economic assessmentの2項目は第1ラウンド後に削除された。
- trustとinterpretabilityは測定法の合意不足等から最終項目に入らなかった。
- 英国、臨床家、工学者が多く、患者代表、allied health、支払者は少なかった。

因果効果として言えないこと

- DECIDE-AIを使うと、研究の方法論的品質が上がる。
- DECIDE-AIを使うと、AIによる害が減る。
- 27項目を報告したAIは、安全、有効、公平である。
- 151人の合意が、患者、家族、日本の医療・介護の価値を代表する。
- 報告項目から外れた費用、trust、interpretabilityは重要でない。

この研究には患者の臨床アウトカムも、介入群も、比較群もない。合意形成研究から、患者安全への因果効果を推定することはできない。

Unknown Labの批判的論評

1. AIを「モデル」から「使われ方」へ広げた点は強い

AI研究は、accuracy、AUC、F1等の性能へ集中しやすい。

DECIDE-AIは、利用者、training、入力、画面、最終判断、上書き、workflow、learning curve、version変更を同じ報告対象にする。

臨床で働くのはモデル単体ではない。AI、人、組織、機器、時間、責任の組合せである。この点は、ケア生態系全体を評価単位とするETIC 2.0の考え方と重なる。ただし、ETIC 2.0自体を直接検証した結果ではない。

2. 多様な群を列挙するだけでは、代表性は保証されない

20群を事前定義したことは丁寧である。しかし、20群が一人以上いたことと、均衡ある決定権を持ったことは違う。

患者代表は5人、合意会議では1人。allied healthは全体10人、合意会議0人だった。英国は79人である。

参加型研究を評価するときは、群の数だけでなく、各群の人数、募集経路、発言機会、投票権、採用された提案、少数意見を残す必要がある。

3. 合意は必要だが、正しさの代わりではない

80%閾値は透明で、項目採否を恣意的にしにくい。

それでも、合意は参加者と時点に依存する。早期臨床AI研究がまだ少なかった2020-2021年の判断であり、生成AI、継続更新、患者向け汎用チャットボットが広がった現在には追加項目が必要かもしれない。

合意した項目を固定した正解ではなく、改訂可能な仮説として扱うべきである。

4. 報告完全性を安全認証に見せてはいけない

チェック欄がすべて埋まると、人は安心しやすい。

だが、エラーの探し方が弱ければ「エラー0件」と報告できる。対象が狭ければ安全そうに見える。短期観察なら遅れて出る害は見えない。利用者が上書きした回数を書いても、その判断が正しかったかは別である。

DECIDE-AIは透明性を高める。安全を認証しない。この区別をサイト、PDF、チェックリストの見た目でも保つ必要がある。

5. 費用・人員・責任を別評価へ送った後が空いている

health economicsを別領域にする判断には合理性がある。しかし、別の評価が実際に行われなければ、最も現場的な負担が消える。

誰の時間が増えたか。誰が監視するか。誰が事故を説明するか。更新費用を誰が払うか。非AI代替を維持できるか。

ETIC 2.0は、この空白を埋めなければならない。

この論文からETIC 2.0が学べること

以下は、DECIDE-AIがETIC 2.0を直接検証した結果ではない。原論文の知見を、ETIC 2.0の考え方と重なる点、答えるべき問い、Unknown Labによる具体的な見直し案に分けて整理する。

区分	DECIDE-AIから受け取ること	ETIC 2.0で明確にすること
考え方と重なる点	AIを患者・利用者・入力・画面・workflowと一体で報告	ケア生態系全体を評価単位にする
考え方と重なる点	version、変更、上書き、エラー、間接的害を追う	AI、prompt、運用、責任者の変更履歴を統合する

区分	DECIDE-AIから受け取ること	ETIC 2.0で明確にすること
考え方と重なる点	human factorsをlive settingで再評価	usabilityだけでなく認知負荷、学習、過信、不信を追う
答えるべき問い	報告指針は質・安全を採点しない	報告完全性と方法論的品質を別列で判定する
答えるべき問い	患者代表3%、allied health 7%	本人・家族・生活支援職の決定権と少数意見を記録する
答えるべき問い	費用項目、trust、interpretabilityが最終項目外	測定未成熟な論点も検証課題として残す
具体的な見直し案	合意項目を最小報告集合として利用	報告、測定、判断、停止・救済の4列へ変換する

具体的な見直し案:「報告・測定・判断・停止」を分ける

DECIDE-AIをETIC 2.0へそのまま貼り付けるのではなく、各論点を4列へ変換する。

列	問い	例
報告	何を透明に書くか	version、入力、画面、利用者、training、変更、害
測定	どの方法で確かめるか	観察、system log、面接、尺度、比較、独立監査
判断	誰が何を基準に進めるか	本人、家族、専門職、組織、規制者の閾値と異論
停止・救済	何が起きたら止め、戻すか	有害事象、格差、過負担、異議、訂正、非AI代替

たとえば「AIへのagreementを報告する」だけでは足りない。

agreementをsystem logで測るのか、利用者の理由も面接するのか。過度の追従を誰が判断するのか。危険な追従が何件あれば停止するのか。誤った記録をどう訂正し、患者へ説明するのか。

この4列がつながって初めて、チェック項目が統治手順になる。

費用・人員・責任・継続可能性を横断列にする

費用を最後の一項目に置くと、実装後に忘れられやすい。

そこで各行に、時間、人員、責任、継続可能性を横断列として加える。

- 患者の入力・確認時間。
- 家族や支援者の無償作業。
- 専門職のtraining、照合、訂正時間。
- 組織の監視、保守、事故対応。
- 開発者のversion管理と再検証。
- 障害時の非AI代替と復旧。

費用効果分析を行えない探索段階でも、誰の資源を使ったかは記録できる。

次に読む論文

次に読むのは、MartindaleらのCONSORT-AI報告監査である。[6]

2020年9月のCONSORT-AI公表後に発表されたAI RCTを系統的に調べ、65報を評価した。

全CONSORT-AI項目へのconcordance中央値は90%、IQR 77-94%だった。一見すると高い。

しかし、CONSORT-AIを使ったと明示したRCTは10報だけだった。65報が掲載された52誌のうち、CONSORT-AIを明示的にmandateしたのは2誌、recommendしたのは1誌だった。algorithm version、AI interventionやcodeへのaccess、study protocolへの参照等は継続して報告不足だった。

この論文は、「指針を作る」から「現実に使われたか」へ検証を進める。

ただし完全に独立した外部評価ではない。複数の著者がCONSORT-AI開発にも関与している。65報のRCT効果やrisk of biasを評価する研究でもなく、主に報告完全性を調べた。

DECIDE-AIの次に読むことで、合意、普及、遵守、方法論的品質、患者成果を別の段階として整理できる。

結論

DECIDE-AIは、早期の臨床AI評価を透明にする重要な報告指針である。

AIのversionや入力だけでなく、患者、利用者、training、画面、workflow、最終判断、上書き、エラー、間接的害、learning curve、途中変更を記録する。臨床AIを、人と組織に組み込まれた複雑な介入として読むための具体的な足場になる。

一方、27項目は安全証明ではない。

開発には151人が参加したが、英国が52%、患者代表は3%、allied healthは7%だった。費用2項目は別領域として外れ、trustとinterpretabilityも測定法の合意不足から残らなかった。合意の規模は、代表性、方法論的品質、臨床効果を代替しない。

ETIC 2.0が受け取るべきなのは、チェック欄そのものより、その次の設計である。

何を報告するか。どう測るか。誰が進行を判断するか。何が起きたら停止し、訂正し、非AI経路へ戻すか。

この4つを分け、費用、人員、責任、継続可能性を全体へ通す。

チェックが全部ついたときこそ、次に問うべきである。

その情報から、誰が、どの基準で、安全に進めると判断したのか。

参考文献

1 Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine. 2022;28:924-933. doi:10.1038/s41591-022-01772-9

2 Vasey B, Nagendran M, Campbell B, et al. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. Nature Medicine. 2021;27:186-187. doi:10.1038/s41591-021-01229-5

3 Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. doi:10.1136/bmj-2022-070904

4 Springer Nature. Author Correction: Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine. 2022;28:1493-1494. doi:10.1038/s41591-022-01951-8

5 Morimoto D. ETIC 2.0には何が足りないのか：妥当性・安全性・外部検証を7つのゲートで設計する. Unknown Lab Research Report 08. 2026;Version 1.0. doi:10.5281/zenodo.22168116

6 Martindale APL, Llewellyn CD, de Visser RO, et al. Concordance of randomised controlled trials for artificial intelligence interventions with the CONSORT-AI reporting guidelines. Nature Communications. 2024;15:1619. doi:10.1038/s41467-024-45355-3

本稿は公開済み査読論文等を批判的に読み解いた未査読の論評であり、DECIDE-AIやETIC 2.0の有効性を検証した研究ではない。個別の診断、治療、AI製品の安全認証や導入を指示しない。
執筆: Unknown Lab編集部（AI支援）/ 専門職・研究方法確認済み