eISSN: 2234-2591
Open Access

Guidelines

Ewha Med J 2025;48(3):e48
https://doi.org/10.12771/emj.2025.00668

The Ewha Medical Journal

TRIPOD+AI 지침: 회귀 또는 머신러닝 방법을 사용하는 임상 예측모델 보고를 위한 최신 지침

TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods: a Korean translation

Gary S. Collins[1*], Karel G. M. Moons[2], Paula Dhiman[1], Richard D. Riley[3,4], Andrew L. Beam[5], Ben Van Calster[6,7], Marzyeh Ghassemi[8], Xiaoxuan Liu[9,10], Johannes B. Reitsma[2], Maarten van Smeden[2], Anne-Laure Boulesteix[11], Jennifer Catherine Camaradou[12,13], Leo Anthony Celi[14,15,16], Spiros Denaxas[17,18], Alastair K. Denniston[4,9], Ben Glocker[19], Robert M. Golub[20], Hugh Harvey[21], Georg Heinze[22], Michael M. Hoffman[23,24,25,26], André Pascal Kengne[27], Emily Lam[12], Naomi Lee[28], Elizabeth W. Loder[29,30], Lena Maier-Hein[31], Bilal A. Mateen[17,32,33], Melissa D. McCradden[34,35], Lauren Oakden-Rayner[36], Johan Ordish[37], Richard Parnell[12], Sherri Rose[38], Karandeep Singh[39], Laure Wynants[40], Patricia Logullo[1]

*For further information on the authors' affiliations, see **Additional information**.*

요약

최근 인공지능(artificial intelligence, AI) 방법, 특히 머신러닝의 발전에 따라 예측모델 개발에 대한 관심과 투자규모가 크게 증가하고 있다. 예측모델 연구가 실제 사용자에게 가치 있게 활용되기 위해서는, 연구자가 왜 연구를 수행했는지, 무엇을 했는지, 그리고 어떤 결과를 얻었는지를 투명하고 완전하며 정확하게 기술해야 한다. TRIPOD 지침의 개정판은 AI 방법을 적용한 예측모델 연구 전반을 일관성 있게 안내하고, 회귀분석이든 머신러닝이든 적용방법에 관계없이 모두를 아우르는 지침을 제공한다. TRIPOD+AI 지침은 27개 항목의 체크리스트, 각 항목별 보고 권고사항을 상세히 설명하는 확장 체크리스트, 그리고 13개 항목의 초록 전용 TRIPOD+AI 체크리스트로 구성된다. TRIPOD+AI의 목표는 저자가 연구를 완전하게 보고하도록 돕고, 동료 평가자와 편집자, 정책 입안자, 최종 사용자, 그리고 환자가 AI 기반 연구의 데이터, 방법, 결과 및 결론을 명확히 이해하도록 돕는 것이다. 이 권고안을 준수하면 연구에 소요되는 시간, 노력, 비용의 활용 효율성을 높일 수 있을 것이다.

서론

2015년에 발표된 TRIPOD (transparent reporting of a multi-

*Corresponding email: gary.collins@csm.ox.ac.uk
Received: July 17, 2025 Revised: July 30, 2025 Accepted: July 30, 2025

It is a Korean translation of the Collins GS. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378. The translation was done with the permission of the TRIPOD Group. Korean medical terminology is based on the English-Korean Medical Terminology 6th edition, available at: https://term.kma.org/index.asp. Korean translation was done by Sun Huh (https://orcid.org/0000-0002-8559-8640), Hallym University. The Korean proofreading was conducted by YoonJoo Seo (https://orcid.org/0000-0002-0202-8352), InfoLumi, and the back-translation was performed by Jeong-Ju Yoo (https://orcid.org/0000-0002-7802-0381), Soonchunhyang University Bucheon Hospital. The back-translation was confirmed by Gary S. Collins (https://orcid.org/0000-0002-2772-2316), the first author of the original TRIPOD+AI statement.

variable prediction model for individual prognosis or diagnosis; 개인별 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고) 지침은 예측모델의 개발 또는 성능 평가 연구에 대한 최소 보고 권고사항을 제시하였다. 이후 예측모델 분야에서는 머신러닝에 기반한 인공지능(artificial intelligence, AI) 기법의 보편화 등 다양한 방법론적 발전이 이루어져 예측모델 개발에 활용되고 있다. 이에 따라 TRIPOD 지침의 개정이 필요하게 되었다. TRIPOD+AI는 회귀분석이든 머신러닝 방법이든 상관없이 예측모델 연구의 보고를 위한 일관된 지침을 제공한다. 이 새로운 체크리스트는 TRIPOD 2015 체크리스트를 대체하는 것으로, 기존 버전은 더 이상 사용하지 말아야 한다. 이 논문에서는 TRIPOD+AI의 개발과정과 함께, 각 보고 권고사항에 대해 더 상세히 설명하는 27개 항목의 확장 체크리스트와 초록용 TRIPOD+AI 체크리스트를 제시한다. TRIPOD+AI의 목적은 예측모델의 개발 또는 성능 평가 연구에서 완전하고 정확하며 투명한 보고를 장려하는 것이다. 완전한 보고는 연구 평가, 모델의 평가 및 실제 적용을 촉진할 것이다.

예측모델은 다양한 의료환경에서 사용되며, 특정 결과값이나 위험도를 산출하는 데 활용된다. 대부분의 모델은 특정 건강상태(진단)의 존재 가능성이나 특정 결과가 미래에 발생할지(예후)를 예측한다[1]. 주요 용도는 임상적 의사결정 지원으로, 예를 들어 추가 검사가 필요한지, 질환 악화나 치료효과 모니터링, 치료 또는 생활습관 변화 시작 여부 결정 등에 활용된다. 널리 알려진 예측모델로는 EuroSCORE II(심장 수술)[2], Gail 모델(유방암)[3], Framingham 위험 점수(심혈관질환)[4], IMPACT(외상성 뇌손상)[5], FRAX(골다공증 및 고관절 골절)[6] 등이 있다.

예측모델은 생의학 문헌에서 매우 풍부하게 보고되고 있으며, 매년 수천 개의 모델이 출판되고 그 수는 점차 증가하고 있다[7,8]. 다양한 건강상태와 임상 결과에 대한 매우 많은 모델이 개발되었다. Coronavirus disease 2019 (COVID-19) 팬데믹 첫 1년 동안만 해도, 진단 및 예후 예측모델 연구가 최소 731편 발표되었다[9]. 이렇게 예측모델 개발에 대한 관심이 높음에도 불구하고, 이 분야에서는 보고의 투명성과 완전성, 그리고 그에 따른 활용 가능성에 대한 우려가 오래전부터 지속되어 왔다[10,11]. 보고가 불완전하거나 부정확하다면, 동료 평가자, 편집자, 의료인, 규제 당국, 환자, 그리고 일반 대중을 포함한 독자 입장에서는 연구설계와 방법을 비판적으로 평가하고 결과에 신뢰를 갖거나, 추가로 모델을 평가하거나 실제 적용하기가 어렵다. 모델에 대한 불충분한 보고로 설계나 데이터 수집, 연구 수행과정의 결함이 가려질 수 있으며, 이러한 모델이 임상경로에 실제로 적용되면 위해로 이어질 수 있다. 특히 편향을 줄이기 위한 충분한 조치가 마련되지 않은 경우 위해가 발생할 수 있다. 더 나은 보고는 신뢰를 제고하고, 예측모델의 의료현장 적용에 대한 환자 및 대중의 수용도에 긍정적인 영향을 미칠 수 있다. 연구자는 자신의 연구를 완전하고 투명하며 정직하게 보고할 윤리적 · 과학적 의무가 있다. Altman 등[12]이 언급한 바와 같이 "좋은 보고는 선택이 아니라 연구의 필수 요소"이며, 그렇지 못한 보고는 결국 불필요한 연구 낭비에 불과하다[13].

불완전한 보고에 대한 우려에 대응하기 위해[10,11,14,15], TRIPOD 지침이 2015년에 발표되어(TRIPOD 2015) 최소 보고 권고사항을 제시하였다[16,17]. TRIPOD 2015는 총 37개 항목의 체크리스트로 구성되어 있으며, 이 중 25개 항목은 개발 및 검증 연구 모두에 공통으로 적용되고, 모델 개발 연구에 6개, 검증 연구에 6개의 추가 항목이 있다. 또한 각 체크리스트 항목의 근거, 좋은 보고의 사례, 예측모델 연구의 설계 · 수행 · 분석 관련 논의 등을 담은 설명 및 해설 문서도 함께 제공된다[17]. TRIPOD 2015는 당대의 주류였던 회귀분석 기반 모델에 주로 초점을 맞추었다. 이후 예측모델 연구의 초록 보고(TRIPOD for Abstracts)[18], 군집화 데이터 사용 연구(TRIPOD-Cluster)[19,20], 예측모델 연구의 체계적 문헌고찰 및 메타분석(TRIPOD-SRMA)[21], 연구 프로토콜 준비 지침(TRIPOD-P)[22] 등 추가 지침이 개발되었다. 이러한 모든 지침과 별도 작성용 체크리스트 서식은 TRIPOD 웹사이트(https://www.tripod-statement.org/)에서 확인할 수 있다.

TRIPOD 2015 발표 이후, 예측모델링 분야에는 샘플 사이즈 산정 지침[23-27], 성능 평가방법[28-32], 공정성[33], 재현성[34], 오픈 사이언스 원칙 적용[35] 등 다양한 방법론적 발전이 이루어졌다. 이 중에서도 가장 큰 변화와 진보가 나타난 영역은 AI로 분류되는 기법의 발전에 기반한 분야이다. 데이터 접근성의 향상과 기성 머신러닝 소프트웨어의 보급으로 예측모델 개발은 훨씬 빠르고 쉬워졌다. 여러 임상환경과 광범위한 결과 · 건강상태에 대한 수많은 예측모델이 문헌에 보고되고 있으며, 동일 결과나 건강상태, 대상 집단에 대해 복수의 모델이 존재하는 경우도 많다[7,8,36]. 따라서 예측모델의 품질을 비판적으로 평가하고, 특정 환경이나 사용목적에 적합한지 이해하는 능력이 더욱 중요해졌다. 이러한 능력은 완전하고 투명한 보고를 전제로 한다.

하지만 예측모델 연구를 평가한 체계적 문헌고찰에서는, 연구설계나 데이터 수집의 결함[37,38], 미흡한 방법론 사용[37,38], 핵심 세부 사항이 누락된 불완전한 보고[39-54], 그에 따른 높은 편향 위험[41,49,55-57], 오픈 사이언스 관행의 미준수[58], 과도한 해석이나 이른바 'spin'의 문제[59,60] 등이 자주 드러난다. 이러한 결함은 모델의 유용성과 안전성에 심각한 의문을 제기하며, 보건의료격차가 심화할 우려도 있다[61]. TRIPOD 2015는 모델링 방식이 중립적이며 보고 권고 대부분이 비회귀적 접근법에도 적용되지만, 머신러닝 기반 모델에는 추가적인 보고 고려사항이 필요하다. 예를 들어 회귀 기반 모델과 달리 머신러닝 모델은 구조의 유연성과 복잡성 때문에 단순 공식이 나오지 않거나, 사용된 예측변수 자체가 불명확한 경우가 많다. 이 때문에 기존 TRIPOD 2015에서는 다루지 않은 추가 보고사항이 필요하다. 방법론적 진전뿐 아니라, 공정성[62], 오픈 사이언스 관행의 확산[63], 환자 및 대중의 연구 · 실제 적용 참여 확대[64,65] 등도 함께 반영할 필요가 있다.

이 논문의 목적은 개정된 TRIPOD 지침의 개발과정을 설명하고, 새로운 TRIPOD+AI 체크리스트를 제시하며, 그 활용법을 논의하는 것이다. TRIPOD+AI는 예측모델 연구의 전반을 조화롭게 정비하고, 회귀분석과 머신러닝 방식 모두에 일관된 지침을 제공하는 것을 목표로 한다[66]. 여기서 "+"는 회귀분석 또는 머신러닝(딥러닝, 랜덤 포레스트 등) 접근법으로 개발된 예측모델 연구에 대해 통합된 보고 권고사항을 제공함을 의미한다. 또한 관련 연구가 일반적으로 AI로 분류되는 보고 지침과의 일관성을 위해 "AI"라는 용어가 추가되었으나, 이 논문에서는 이해를 돕기 위해 주로 머신러닝이라는 용어를 사용한다(Table 1) [67-73]. TRIPOD+AI 보고 지침에서 사용하는 주요 개념에 대한 용어 설명은 Box 1에서 확인할 수 있다.

TRIPOD+AI 개발과정

이 절에서는 머신러닝 또는 회귀방법을 이용하여 진단 또는 예후 예측모델을 개발하거나, 이들 모델의 성능을 평가(검증)하는 연구의 보고를 지원하기 위한 지침인 TRIPOD+AI 지침의 개발과정을 기술한다. '검증된 예측모델'이라는 개념은 존재하지 않으므로[76], 이 논문에서는 혼동을 피하고 용어를 통일하기 위하여 검증(validation) 대신 평가(evaluation)라는 용어를 사용하였다(Box 1). 머신러닝이 포함된 기타 생의학 연구 유형의 보고를 위한 기존 및 개발 중인 가이드라인은 Table 1에 상세히 정리하였다. TRIPOD+AI 체크리스트는 EQUATOR Network의 권고에 따라[77], 문헌고찰과 전문가 합의과정을 통해 개발되었다. G.S.C.와 K.G.M.M.의 주도로 다양한 전문성과 경험을 반영한 위원(G.S.C., K.G.M.M., R.D.R., A.L.B., J.B.R., B.V.C., X.L., P.D.)을 선정하여 지침 개발을 감독할 운영위원회를 구성하였다.

2019년 4월에는 TRIPOD+AI 이니셔티브를 알리는 해설 논문이 발표되었으며[78], 2019년 5월 7일에는 EQUATOR Network에 개발 중인 보고 지침으로 공식 등록되었다(https://www.equator-network.org/). 2021년 3월 25일에는 Open Science Framework (https://osf.io/zyacb/)에 개발과정 및 방법론을 담은 연구 프로토콜이 공개되었다. 이 프로토콜은 머신러닝 기반 예측모델의 품질 평가 및 편향 위험도구(PROBAST+AI)의 개발과정도 포함하고 있으며, 2021년에 출판되었다[79]. TRIPOD+AI 개발과정에서 활용한 합의 기반 방법의 보고는 ACCORD (Accurate Consensus Reporting Document) 권고를 준수하였다[80].

윤리적 고려

본 연구는 2020년 12월 10일 옥스퍼드대학교 중앙 연구윤리위원회(Central University Research Ethics Committee, University of Oxford)의 승인을 받았다(R73034/RE001). 델파이(Delphi) 설문 참여자에게는 설문 시작 전에, 그리고 합의 회의 참여자에게는 회의 시작 전에 참여자 안내문을 전자적으로 제공하였다. 델파이 설문 참여자는 설문 응답 전 전자 동의서를 제출하였다.

후보 항목 목록 도출

G.S.C.와 K.G.M.M.이 TRIPOD 2015 [16,17]를 바탕으로 초기 항목 목록의 초안을 작성하였다. 이후, TRIPOD-Cluster [19,20], TRIPOD for Abstracts [18], CAIR [81], MI-CLAIM [82], CLAIM [68], MINIMAR [83], SPIRIT-AI [71], CONSORT-AI [72] 및 운영위원회가 추가로 확인한 문헌[34,84-89]을 참고하여 추가 항목을 도출하였다. 머신러닝 기반 예측모델 연구의 보고, 방법론, 과도한 해석을 평가한 체계적 문헌고찰 결과[37-39,48,51,54,59,60]도 항목 목록 구성에 반영하였다. 운영위원회

Table 1. 머신러닝을 활용한 보건의료 연구의 보고 지침

보고 지침(reporting guideline)	적용 범위(scope)
STARD-AI	인공지능 기반 진단 정확도 평가 연구(작성 중)[67]
TRIPOD+AI	인공지능(머신러닝 방법 포함)을 이용한 예측 모델 개발 또는 성능 평가 연구
CLAIM	인공지능을 활용한 의료영상 연구[68]
DECIDE-AI	인공지능 기반 의사결정 지원시스템의 초기 임상 평가(안전성, 인간 요인 평가 포함)[69]
CHEERS-AI	인공지능 중재의 비용 효과성 등 건강경제학적 평가 연구[70]
SPIRIT-AI	인공지능 요소가 포함된 중재의 임상시험 연구 프로토콜[71]
CONSORT-AI	인공지능 요소가 포함된 중재의 임상시험 보고서[72]
PRISMA-AI	인공지능 중재에 관한 체계적 문헌고찰 및 메타분석(작성 중)[73]

STARD, 진단 정확도 보고 기준(Standards for Reporting of Diagnostic Accuracy); TRIPOD, 개인 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고(Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis); AI, 인공지능(artificial intelligence); CLAIM, 의료영상 인공지능 연구 체크리스트(Checklist for Artificial Intelligence in Medical Imaging); DECIDE, 근거 기반 혁신의 도입 및 확산을 위한 보건의료 의사결정(Decisions in health Care to Introduce or Diffuse innovations using Evidence); CHEERS, 건강경제학적 평가 통합 보고 기준(Consolidated Health Economic Evaluation Reporting Standards); SPIRIT, 중재 임상시험 프로토콜 권고(Standard Protocol Items: Recommendations for Interventional Trials); CONSORT, 임상시험 보고 통합 기준(Consolidated Standards of Reporting Trials); PRISMA, 체계적 문헌고찰 및 메타분석 보고 권고(Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

Box 1. TRIPOD+AI에서 사용된 용어 해설

아래 정의 및 설명은 TRIPOD+AI* 가이드라인의 맥락에 한정된 것이며, 다른 연구 분야에는 반드시 적용되지 않을 수 있다.
인공지능(artificial intelligence): 통상적으로 인간의 지능이 필요한 과업을 수행할 수 있는 모델 및 알고리즘을 개발하는 컴퓨터 과학 분야.
보정(calibration): 관찰된 결과와 모델에서 추정된 값 간의 일치 정도. 보정은 일반적으로 추정값(x축)과 관찰값(y축)을 그래프로 나타내고, 개별 데이터의 유연한 보정 곡선을 함께 제시하여 평가하는 것이 가장 바람직하다.
진료 경로(care pathway): 특정 건강 문제 관리 또는 환자의 진료 전 과정을 포괄하는 구조적 · 조정된 진료계획.
클래스 불균형(class imbalance): 결과 사건이 발생한 집단과 발생하지 않은 집단의 빈도가 불균등한 현상.
변별력(discrimination): 모델의 예측이 결과 발생 집단과 미발생 집단을 얼마나 잘 구분하는지의 정도. 변별력은 이항 결과의 경우 c-통계량(또는 곡선하면적[area under the curve], 수신자조작특성곡선하영역[area under the receiver operating characteristic curve])으로, 시점-사건(time-to-event) 결과는 c-지수로 정량화된다.
평가 또는 테스트 데이터(evaluation or test data): 예측모델의 성능을 추정하는 데 사용되는 데이터. '테스트 데이터' 또는 '검증 데이터'로도 불린다.[a] 평가 데이터는 모델 훈련, 하이퍼파라미터 튜닝, 모델 선택 등에 사용된 데이터와 구분되어야 하며, 두 데이터 세트 간 참가자의 중복이 없어야 한다. 평가 데이터는 모델이 실제로 사용될 대상 인구를 대표해야 한다.
공정성(fairness): 예측모델이 연령, 인종/민족, 성별/젠더, 사회경제적 지위 등과 같은 특성을 바탕으로 개인 또는 집단을 차별하지 않는 특성.
하이퍼파라미터(hyperparameters): 모델 개발 또는 학습과정을 제어하는 값.
하이퍼파라미터 튜닝(hyperparameter tuning): 특정 모델 구축 전략에 가장 적합한 (하이퍼)파라미터 설정을 찾는 과정.
내부 검증(internal validation): 모델이 개발된 동일한 집단을 대상으로 예측모델의 성능을 평가하는 것(예: 훈련-테스트 분할, 교차검증, 부트스트래핑[bottstrapping] 등).
머신러닝(machine learning): 데이터로부터 명시적으로 프로그래밍하지 않고 학습하고 예측이나 의사결정을 내릴 수 있는 모델을 개발하는 인공지능의 한 분야.
모델 평가(model evaluation): c-통계량 등으로 모델의 변별력, 보정도(보정도 그래프, 보정 기울기 등), 임상적 유용성(의사결정 곡선 분석 등)을 추정하여 모델의 예측 정확도를 평가하는 과정. 이 과정을 예측모델의 평가라 부른다[74,75].
결과(outcome): 예측하고자 하는 진단 또는 예후 사건. 머신러닝에서는 이를 목표값(target value), 반응변수(response variable), 또는 레이블(label)이라고 지칭하기도 한다.
예측 변수(predictor): 개인 수준(예: 나이, 수축기 혈압, 성별, 질병 단계, 라디오믹스 특성) 또는 집단 수준(예: 국가)에서 측정되거나 할당될 수 있는 특성. 입력값, 특성(feature), 독립변수, 공변량 등으로도 불린다.
훈련 또는 개발 데이터(training or development data): 예측모델의 훈련 또는 개발에 사용되는 데이터. 이상적으로는, 훈련 데이터가 모델 실제 사용 인구를 대표해야 한다.
TRIPOD, 개인별 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고(transparent reporting of a multivariable prediction model for individual prognosis or diagnosis); AI, 인공지능(artificial intelligence).

[a]검증 데이터(validation data)는 연구마다 의미가 다를 수 있다. 예를 들어, 머신러닝 연구에서 검증 데이터는 파라미터 튜닝에 사용되는 데이터 또는 모델 성능 평가(대개 외부 검증이라고도 함)에 사용되는 데이터를 의미할 수 있다. 이 가이드라인에서는 혼동을 방지하기 위해 모델 성능 평가에 사용된 데이터를 평가 데이터(evaluation data)라 명명하였다.

는 이러한 자료를 바탕으로, 제목(1항목), 초록(1항목), 서론(3항목), 방법(37항목), 결과(15항목), 논의(5항목), 기타(3항목)를 포함하는 65개의 고유 후보 항목으로 최종 목록을 정비하였다. 이 목록은 이후 설명하는 수정 델파이 합의과정에서 활용되었다.

델파이 패널 선정

델파이 조사 참여자는 운영위원회가 선정하였으며, 관련 논문 저자, 소셜미디어(예: 트위터) 모집공고, 그리고 개인 추천을 통해 모집하였다. 이에는 다른 델파이 참여자가 추천한 전문가도 포함된다. 운영위원회는 지리적 및 학문적 다양성을 확보하고, 주요 이해관계자 집단—예를 들어 연구자(통계학자, 데이터 과학자, 역학자, 머신러닝 연구자, 임상의, 영상의학과 전문의, 윤리학자 등), 의료전문가, 학술지 편집자, 연구비 지원기관, 정책 입안자, 보건의료 규제기관, 예측모델의 실제 사용자(환자 및 일반 대중 등)—을 포괄하도록 참여자를 선정하였다. 참여자는 대학, 병원, 1차 의료기관, 생의학 학술지, 비영리기관, 영리기관 등 다양한 환경에서 모집하였다.

델파이 참여자의 최소 표본 수에는 제한을 두지 않았다. 선정된 모든 참여자에 대해 운영위원회 구성원이 전문성이나 관련 경험을 확인하였다. 이후 각 참여자에게 이메일로 연구설명, 목표, 연락처 등이 포함된 안내자료와 함께 참여 초대장을 발송하였다. 참여자가 동의하면 델파이 패널로 등록되어 설문 링크를 받았다. 델파이 패널은 설문 참여에 대한 금전적 보상이나 선물을 제공받지 않았다.

델파이 과정

델파이 설문조사는 Welphi 온라인 플랫폼(www.welphi.com)을 통해, 각 참여자가 개별적으로 온라인(영어)으로 응답할 수 있도록 설계되어 배포되었다. 이 플랫폼은 각 참여자에게 개별 링크를 발

송하고 응답자에게 코드를 부여하여 익명성을 보장한다. 패널에게는 연구의 목적과 범위, 참여방법, 플랫폼 사용법, 문의처 등을 포함한 안내자료를 제공하였다. 참여자들은 각 항목에 대해 '제외 가능,' '포함 가능,' '포함 권장,' '포함 필수' 중 하나로 평가하도록 요청받았다. 또한 각 항목에 대해 자유롭게 의견을 남기거나 신규 항목을 제안할 수 있었다. 자유 서술식 응답은 P.L.이 취합 및 분석하였으며, 이를 바탕으로 G.S.C.와 K.G.M.M.이 항목의 재서술, 통합, 신규 항목 제안을 논의하였다. 운영위원회 구성원 전원에게 델파이 설문 참여 기회가 주어졌다.

1차 라운드 참여자

292명에게 초대장과 설문 참여 링크가 발송되었으며, 1차 라운드는 2021년 4월 19일부터 5월 13일까지 진행되었다. 2021년 5월 5일에 확인 메시지가 발송되었다. 초대된 292명 중 170명(부분 응답자 8명 포함)이 설문을 완료하였다. 참여자는 총 22개국에서 모집되었으며, 주요 국가는 영국(n=52), 미국(n=31), 네덜란드(n=23), 캐나다(n=20)였다. 5개 대륙(유럽 100명, 남미 2명, 북미 51명, 오세아니아 4명, 아시아 13명)에서 응답하였고, 7명은 국가를 밝히지 않았다.

참여자들은 자신의 주요 연구/업무 분야를 복수 선택할 수 있었다. 통계 · 데이터 과학(n=70), AI 또는 머신러닝(n=69), 임상(n=50), 역학(n=40), 예측(n=18), 영상의학(n=18), 보건 정책/규제(n=10), 생의학 연구(n=7), 학술지 편집자(n=6), 메타연구/보고(n=6), 병리학(n=2), 연구비 지원 기관(n=2), 윤리(n=2), 기술 개발/실행(n=2), 유전학/유전체(n=2), 의생명공학(n=2), 보건 경제(n=2) 등이 보고되었다.

2차 라운드 참여자

2차 델파이 라운드는 2021년 12월 16일부터 2022년 1월 17일까지 진행되었다. 1차 라운드 설문을 완료한 모든 참여자가 2차 라운드에 초대되었으며, 1차 미응답자와 1차 이후 추가 추천된 참여자도 재초대되었다. 2차 라운드 초대장은 총 395명에게 발송되었고, 200명(부분 응답자 15명 포함)이 설문을 완료하였다. 응답자는 27개국에서 모집되었으며, 역시 영국(n=70), 미국(n=37), 네덜란드(n=19), 캐나다(n=19)가 다수를 차지하였다. 6개 대륙(유럽 123명, 남미 3명, 북미 56명, 오세아니아 7명, 아시아 10명, 아프리카 1명)에서 응답이 있었다. 주요 분야는 통계 · 데이터 과학(n=78), AI 또는 머신러닝(n=72), 임상(n=49), 역학(n=51), 예측(n=19), 영상의학(n=26), 보건정책(n=12), 생의학 연구(n=14), 학술지 편집자(n=13), 메타연구/보고(n=6), 의생명공학(n=5), 연구비 지원기관(n=2), 유전학/유전체(n=4), 환자 대표/참여(n=3), 보건 경제(n=2), 윤리(n=1) 등이었다.

체크리스트 항목의 진화(1차→2차 라운드)

수정된 델파이 1차 라운드에서는, 참여자들이 문헌고찰 및 기존 보고 체크리스트에서 도출한 65개 후보 항목을 평가하였다. 항목 포함에 대해 '포함 권장' 또는 '포함 필수'로 응답한 경우 합의한 것으로 간주하였다. 프로토콜에서 정의한 대로[79], 70% 이상 합의에 도달한 항목만 2차 라운드로 이월되었다. 70% 미만의 항목은 제외하거나 통합 또는 재서술되어 재평가 대상으로 제시되었다. 이러한 수정은 수백 건의 패널 의견을 반영하여 이루어졌다.

2차 라운드에서는 1차 라운드의 집계결과(https://osf.io/zyacb/)를 참고하도록 안내하고, 59개 후보 항목(제목 1, 초록 1, 서론 4, 방법 32, 결과 11, 논의 8, 기타 2)에 대해 평가하도록 하였다. 환자 및 공공 참여 관련 항목은 포함 합의율이 69%로 70% 기준에 약간 못 미쳤으나, 운영위원회는 합의 회의에서 이 항목을 논의 항목으로 유지하기로 결정하였다.

환자 및 대중 참여 회의

2022년 4월 8일, Health Data Research UK의 환자 및 대중 참여 그룹(Patient and Public Involvement and Engagement, PPIE; https://www.hdruk.ac.uk/about-us/involving-and-engaging-patients-and-the-public/) 소속 9명을 대상으로 온라인 회의가 진행되었다. 이 회의는 University of Warwick의 Sophie Staniszewska가 주재하였다. 이 회의는 연구 프로토콜에 계획되어 있지 않았으며, 출판된 프로토콜[79]과의 유일한 차이점이었다. PPIE 그룹은 회의에 앞서, TRIPOD+AI 프로젝트의 요약(https://osf.io/zyacb/ 참조), PPIE 그룹원 한 명이 작성한 요약문, 그리고 체크리스트 초안을 전달받았다. 회의에서 GSC는 TRIPOD+AI 이니셔티브의 세부 내용, 프로젝트 현황, 2차 델파이 설문결과를 바탕으로 한 초안 지침을 발표하였다. 이후 참여자들은 질의응답을 진행하며 프로젝트의 목표와 범위에 대해 논의하였다. 명확성을 높이기 위해, 회의 중 제기된 의견과 회의 이후 받은 서면 피드백을 바탕으로 체크리스트 초안이 수정되었다. PPI 그룹의 세 명이 다양한 이해관계자가 참여한 2022년 7월 5일의 온라인 합의 회의에 초대되었으며, 이 중 두 명이 실제로 참석하였다. 원고는 세 명의 PPI 회원에게 전달되어 의견을 받고 승인절차를 거쳤다.

합의 회의(consensus meeting)

2022년 7월 5일, G.S.C.와 K.G.M.M.의 사회로 온라인 합의 회의가 개최되었다. 주요 이해관계자 그룹과 다양한 학문 분야, 지리적 다양성이 균형 있게 반영되도록 참가자를 선정하였다. 총 28명이 회의의 전부 또는 일부에 참석하였으며, 이 중 1명(P.L.)은 투표권이 없는 참관자였다. 초청받은 참가자들에게는 TRIPOD+AI 개요, 합의 회의 진행방식 및 안내, 2차 델파이 설문 종합결과 요약, 그리고 TRIPOD+AI 체크리스트 초안이 포함된 문서(https://osf.io/zyacb/)를 회의 전에 이메일로 발송하였다. 체크리스트 초

안은 제목(1항목), 초록(1항목), 서론(4항목), 방법(32항목), 결과(11항목), 논의(8항목), 기타(2항목) 등 총 59개 항목을 포함하였다.

2차 라운드에서 다수 항목에 대한 강한 지지가 확인되었기 때문에, 이 중 17개 항목이 전체 회의에서 토론 및 표결 대상이 되었다. 논의 후 각 항목에 대해 TRIPOD+AI 체크리스트 포함 여부를 1분간 투표하도록 하였으며, 온라인 회의 플랫폼의 투표 기능이 사용되었다. 이 17개 항목에는 2차 라운드에서 합의에 도달하지 못한 1개 항목과, 2차 라운드 이후 재서술되었거나 TRIPOD 2015에 포함되지 않았던 신규 항목 16개가 포함되었다. 이들 17개 항목에 대한 논의와 표결을 거쳐 최종 TRIPOD+AI 체크리스트를 확정하였다.

TRIPOD+AI 지침

TRIPOD+AI는 통계적 또는 머신러닝 방법을 이용해 예측모델을 개발하거나 평가(검증)하는 연구의 적절한 보고에 필수적인 항목들로 구성된 체크리스트로(Table 2), TRIPOD 2015의 주요 변화와 추가 사항은 Box 2에 요약되어 있다. TRIPOD+AI 체크리스트는 제목(항목 1), 초록(항목 2), 서론(항목 3, 4), 방법(항목

Table 2. 예측모델 연구 보고를 위한 TRIPOD+AI 체크리스트

섹션/주제	하부 주제	항목	개발/평가[a]	체크리스트 항목
제목	제목	1	D;E	연구가 다변량 예측모델의 개발 또는 성능 평가임을, 대상 집단 및 예측할 결과와 함께 명시한다.
초록	초록	2	D;E	TRIPOD+AI 초록 체크리스트 참조
서론	배경	3a	D;E	보건의료 맥락(진단 또는 예후 등) 및 예측모델 개발/평가의 근거를 설명하고, 기존 모델에 대한 참고문헌을 포함한다.
		3b	D;E	대상 집단과 예측모델의 진료 경로 내 의도된 목적 및 사용자를 기술한다(예: 의료인, 환자, 일반인 등).
		3c	D;E	사회인구학적 집단 간 알려진 건강불평등을 기술한다.
	목적	4	D;E	연구의 목적을 구체적으로 명시하며, 예측모델의 개발 또는 검증 중 어떤 연구인지(또는 둘 다인지) 기술한다.
방법	데이터	5a	D;E	개발 및 평가 데이터의 출처를 각각 기술하고(예: 무작위 임상시험, 코호트, 진료정보, 레지스트리 등), 데이터 활용의 근거와 대표성을 설명한다.
		5b	D;E	참가자 데이터의 수집 기간(시작 및 종료), 그리고 해당 시기 종료 여부(추적 종료 등)를 명확히 한다.
	참가자	6a	D;E	연구환경의 주요 요소(예: 1차 진료, 2차 진료, 일반 인구), 기관 수와 위치를 명시한다.
		6b	D;E	연구 참가자의 선정기준을 기술한다.
		6c	D;E	적용된 치료(있는 경우)와 개발/평가과정에서의 처리방법을 설명한다.
	데이터 준비	7	D;E	데이터 전처리 및 품질 관리방법, 그리고 이 과정이 사회인구학적 집단 간 유사했는지 여부를 설명한다.
	결과	8a	D;E	예측하는 결과 및 평가 시점, 결과 선정의 근거, 결과 평가방법이 사회인구학적 집단에서 일관되게 적용됐는지 명확히 기술한다.
		8b	D;E	결과 평가에 주관적 해석이 필요한 경우, 평가자의 자격 및 인구통계적 특성을 설명한다.
		8c	D;E	예측결과 평가의 눈가림 수행 여부 및 방법을 보고한다.
	예측변수	9a	D	초기 예측변수의 선정 근거(문헌, 기존 모델, 가용 변수 등) 및 모델 구축 전 사전 선정과정을 설명한다.
		9b	D;E	모든 예측변수를 명확히 정의하고, 측정 시점과 방법(및 결과/다른 예측변수의 눈가림 여부 포함)을 기술한다.
		9c	D;E	예측변수의 측정에 주관적 해석이 필요한 경우, 평가자의 자격 및 인구통계적 특성을 설명한다.
	표본크기	10	D;E	연구 규모 산출근거를(개발/평가별로) 설명하고, 연구질문에 충분한 규모였음을 정당화하며, 표본크기 산출 세부 내용을 포함한다.
	결측 데이터	11	D;E	결측 데이터 처리 방법 및 누락 사유를 기술한다.
	분석방법	12a	D	데이터 사용(개발/성능 평가 목적 등) 및 분석방법, 데이터 분할 여부와 표본크기 요건 고려사항을 명시한다.
		12b	D	모델 유형에 따라 예측변수의 분석 처리(함수형, 재조정, 변환, 표준화 등)를 설명한다.
		12c	D	모델 유형, 근거[b], 모든 모델 구축 단계(하이퍼파라미터 튜닝 등), 내부 검증방법을 명시한다.
		12d	D;E	집단 간(병원, 국가 등) 모델 파라미터 및 성능 추정치의 이질성 처리 및 정량화 방법을 기술한다. 추가사항은 TRIPOD-Cluster 참조.[c]

(Continued on the next page)

Table 2. Continued

섹션/주제	하부 주제	항목	개발/평가[a]	체크리스트 항목
		12e	D;E	모델 성능 평가에 사용된 모든 지표 및 그래프(근거 포함)를 명시하고, 필요한 경우 여러 모델 간 비교방법도 기술한다.
		12f	E	모델 평가에서 파생된 모델 수정(재보정 등)을 전체 또는 특정 집단/환경별로 기술한다.
		12g	E	모델 평가 시, 예측값 산출방식(수식, 코드, 오브젝트, API 등)을 설명한다.
	클래스 불균형	13	D;E	클래스 불균형 처리방법, 적용 이유, 사후 재보정 방법을 기술한다.
	공정성	14	D;E	모델 공정성 향상을 위한 방법 및 근거를 설명한다.
	모델 산출값	15	D	예측모델의 산출값(확률, 분류 등)을 명확히 하고, 분류기준 및 임계값 선정방법을 상세히 설명한다.
	개발-평가 차이	16	D;E	개발 데이터와 평가 데이터 간 환경, 선정기준, 결과, 예측변수의 차이를 기술한다.
	윤리 승인	17	D;E	연구를 승인한 기관윤리위원회 또는 윤리위원회의 명칭과, 연구 참가자의 동의(또는 윤리위원회의 동의 면제) 절차를 명시한다.
오픈 사이언스	연구비	18a	D;E	본 연구의 연구비 출처 및 후원자 역할을 기술한다.
	이해관계	18b	D;E	모든 저자의 이해관계 및 재정적 공시를 명시한다.
	프로토콜	18c	D;E	연구 프로토콜의 접근 가능 위치를 알리고, 프로토콜 미작성 시에는 해당 사실을 명시한다.
	등록	18d	D;E	연구 등록정보(등록기관, 등록번호 포함)를 제공하고, 미등록 시에는 해당 사실을 명시한다.
	데이터 공유	18e	D;E	연구 데이터의 접근 가능성 및 공유방식을 기술한다.
	코드 공유	18f	D;E	분석코드의 접근 가능성 및 공유방식을 기술한다.[d]
환자 및 공공 참여	환자 및 공공 참여	19	D;E	연구설계, 수행, 보고, 해석, 확산 중 어느 단계에서든 환자/공공 참여 내역을 상세히 기술하거나, 참여가 없음을 명시한다.
결과	참가자	20a	D;E	연구 내 참가자 흐름(결과 발생 유무별 참가자 수, 추적관찰 요약 포함)을 기술하고, 필요 시 도식화한다.
		20b	D;E	전체 및 환경별 주요 특성(날짜, 주요 예측변수, 치료내역, 표본 수, 결과 발생 수, 추적기간, 결측 데이터 등)을 보고하고, 인구집단별 차이도 명시한다.
		20c	E	모델 평가에서 주요 예측변수(인구통계, 예측변수, 결과 등)의 개발 데이터와의 분포 비교를 제시한다.
	모델 개발	21	D;E	각 분석(모델 개발, 하이퍼파라미터 튜닝, 평가 등)별 참가자 수 및 결과 사건 수를 명시한다.
	모델 명세	22	D	예측모델(수식, 코드, 오브젝트, API 등) 상세 내역을 제공하고, 새로운 개인 예측 또는 제3자 평가 · 구현에 필요한 접근 제한 여부(무료, 독점 등)를 명확히 기술한다.[e]
	모델 성능	23a	D;E	신뢰구간을 포함한 모델 성능 추정치, 주요 하위집단(예: 사회인구학적)별 성능, 시각화 자료(그래프 등) 제시를 고려한다.
		23b	D;E	집단 간 모델 성능의 이질성이 평가된 경우 결과를 보고한다. 추가 내용은 TRIPOD-Cluster 참고[c]
	모델 수정	24	E	모델 수정(예: 업데이트, 재보정) 및 수정 후 성능 결과를 보고한다.
논의	해석	25	D;E	주요 결과에 대한 종합적 해석을 제시하고, 목적 및 기존 연구 맥락에서 공정성 문제를 논의한다.
	한계	26	D;E	비대표성 표본, 표본크기, 과적합, 결측 데이터 등 연구의 한계 및 이로 인한 편향, 통계적 불확실성, 일반화 가능성에 미치는 영향을 논의한다.
	활용성	27a	D	입력 데이터(예측변수 등) 품질이 낮거나 제공 불가할 때의 평가 및 처리방식을 설명한다.
		27b	D	모델 적용 및 입력 데이터 활용 시 사용자의 상호작용 필요성, 요구되는 전문성 수준을 명확히 한다.
		27c	D;E	모델의 적용성과 일반화 가능성에 초점을 두고, 향후 연구과제를 논의한다.

TRIPOD, 개인별 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고(Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis); AI, 인공지능(artificial intelligence).

[a]D : 예측모델 개발에만 해당, E : 예측모델 평가에만 해당, D;E : 개발과 평가 모두에 해당. [b]모든 모델 구축 접근법에 대해 별도로 기술. [c]TRIPOD-Cluster는 클러스터(예 : 병원, 센터 등)를 명시적으로 고려하거나 성능 이질성을 탐색하는 연구 보고 체크리스트. [d]데이터 정제, 특성 엔지니어링, 모델 구축 및 평가 등 분석코드에 해당. [e]신규 예측 위험 추정을 위한 모델 구현 코드에 해당.

Box 2. TRIPOD 2015의 주요 변경 및 추가 사항

- **새로운 체크리스트:** 랜덤 포레스트, 딥러닝 등 어떠한 회귀 또는 머신러닝 방법을 사용한 예측모델 연구도 포함할 수 있도록 보고 권고사항을 새롭게 마련하였고, 회귀 및 머신러닝 커뮤니티 간 용어를 통합하였음.
- **TRIPOD+AI 체크리스트 도입:** TRIPOD+AI 체크리스트가 기존 TRIPOD 2015 체크리스트를 대체하므로, 더 이상 TRIPOD 2015는 사용하지 않아야 함.
- **공정성에 대한 강조:** 공정성(Box 1)을 특별히 강조하여, 보고서에서 공정성 문제를 다루기 위해 어떤 방법이 사용되었는지 반드시 언급하도록 하였고, 체크리스트 전반에 공정성 요소를 포함함.
- **초록 보고 지침 추가:** 초록 작성 시 참고할 수 있도록 TRIPOD+AI for Abstracts를 별도 포함함.
- **모델 성능 항목 수정:** 저자가 주요 하위집단(예: 사회인구학적 집단)에서 모델 성능을 평가할 것을 권고하도록 해당 항목을 수정함.
- **환자 및 공공 참여 항목 신설:** 연구의 설계, 수행, 보고(및 해석), 확산과정에서 환자 및 공공의 참여에 대해 상세히 기술하도록 저자에게 요청하는 항목을 새롭게 추가함.
- **오픈 사이언스 섹션 신설:** 연구 프로토콜, 등록, 데이터 공유, 코드 공유 등에 관한 하위항목을 포함한 오픈 사이언스 섹션을 도입함.

TRIPOD, 개인별 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고(transparent reporting of a multivariable prediction model for individual prognosis or diagnosis); AI, 인공지능(artificial intelligence).

5-17), 오픈 사이언스 관행(항목 18), 환자 및 대중 참여(항목 19), 결과(항목 20-24), 논의(항목 25-27) 등 총 27개의 주요 항목으로 구성된다. 일부 항목은 복수의 세부 항목을 포함하고 있어, 총 52개의 체크리스트 세부 항목으로 구성된다.

TRIPOD+AI는 예측모델의 개발, 예측모델 성능 평가(검증), 또는 이 둘을 모두 다루는 연구를 포괄한다. D:E로 표기된 항목은 예측모델 개발 및 평가 연구 모두에 공통으로 적용된다(Table 2). 체크리스트 중 D로 표기된 항목은 예측모델 개발 연구에, E로 표기된 항목은 모델 성능 평가 연구에 적용된다. 예측모델의 개발과 평가를 모두 포함한 연구의 경우, 모든 체크리스트 항목이 적용된다.

TRIPOD+AI는 예측모델 연구의 학술지 또는 학회 초록을 위한 별도의 체크리스트도 포함하고 있다. 이 체크리스트는 기존 TRIPOD for Abstracts 지침을 업데이트한 것으로[18], 새로운 내용을 반영하고 TRIPOD+AI와의 일관성을 유지하도록 설계되었다(Table 3).

TRIPOD+AI의 권고사항은 예측모델 연구의 수행과정을 투명하게 보고하도록 안내하는 것이며, 예측모델 개발 또는 평가방법 자체를 규정하는 것은 아니다. 이 체크리스트는 연구의 질을 평가하는 도구가 아니다. 예측모델의 질과 편향 위험성 평가에는 PROBAST [90,91] 및 곧 공개될 PROBAST+AI [79] 사용을 권장하며, 관련 정보는 https://www.probast.org/에서 확인할 수 있다.

TRIPOD+AI 사용방법

TRIPOD+AI 체크리스트는 기존 TRIPOD 2015 체크리스트를 대체하므로, 이제 더 이상 TRIPOD 2015 체크리스트는 사용하지 않아야 한다. 만약 예측모델 연구에서 클러스터(예: 다수 병원, 다수 데이터 세트)를 고려했다면, 저자는 TRIPOD-Cluster의 추가 보고 권고사항[19,20]을 참고해야 한다. 2015년판 설명 및 해설 문서는 여전히 대부분의 TRIPOD+AI 보고 항목에 대한 배경 및 예시를 제공하는 중요한 자료로 남아 있다[17] (많은 항목이 변경되지 않았거나 최소한만 변경되었기 때문이며, TRIPOD+AI에 대한 보다 상세하고 최신의 해설 문서는 별도로 제작 중이다). TRIPOD+AI는 논문 작성 초기 단계부터 활용하여 모든 핵심 세부사항을 누락 없이 보고할 것을 권장한다. 각 항목별로 간단한 근거와 안내를 담은 목록형 확장 체크리스트(Supplement 1)를 개발하여, TRIPOD+AI의 실제 적용을 지원하고자 하였다.

TRIPOD+AI 체크리스트의 많은 항목은 논문 내에서 자연스러운 순서로 배열되지만, 일부 항목은 그렇지 않을 수 있다. 예측모델 논문이나 출판물 내에서 각 권고사항이 반드시 어디에 위치해야 하는지 구조적 형식을 별도로 규정하지 않으며, 해당 순서는 학술지의 투고양식에 따라 달라질 수 있다.

TRIPOD+AI에 담긴 권고사항은 최소한의 보고 기준이므로, 저자는 추가적인 정보를 제공할 수 있다. 논문 본문의 분량 제한이나 표 · 그림 개수 제한 등으로 인해 보고가 어려운 경우, 일부 요청된 정보나 추가 자료는 보충자료로 보고하고 그 위치를 본문에서 명시하면 된다. 필요한 정보가 이미 공개적으로 접근 가능한 연구 프로토콜에 보고되어 있다면, 해당 문서를 참조하는 것으로 충분하다. 특정 체크리스트 항목을 알 수 없거나 해당이 되지 않아 보고할 수 없다면 그 사실을 명확히 밝혀야 한다. 보충자료에 포함되지 않은 추가 파일이나 연구자료는 Open Science Framework, Dryad, figshare 등과 같은 범용 또는 소속기관의 오픈 액세스 저장소에 영구적으로 공개해야 한다. 추가 파일의 접근 정보(예: doi 번호)는 논문 본문이나 출판물에 반드시 명시 · 연결해야 한다.

저자는 각 항목이 본문 내 어디에 보고되어 있는지(페이지 또는 줄 번호) 명시한 완성된 체크리스트를 제출하도록 권장되는데, 이는 편집 및 동료 평가과정에 도움이 된다. 별도 작성용 TRIPOD+AI 체크리스트 양식은 Supplement 2 및 www.tripod-statement.org에서 다운로드할 수 있다.

TRIPOD+AI 관련 소식, 공지, 정보는 TRIPOD 웹사이트(www.tripod-statement.org)와 X (구 트위터, @TRIPODStatement) 등 소셜미디어 계정에서 확인할 수 있다. 또한 건강 연구의

Table 3. 학술지 또는 학회 초록에 포함해야 할 예측모델 연구의 필수 항목(TRIPOD+AI for Abstracts[a])

섹션 및 항목	체크리스트 항목
제목	1. 연구가 다변량 예측모델의 개발 또는 성능 평가임을, 대상 집단 및 예측할 결과와 함께 명시한다.
배경	2. 보건의료 맥락 및 모든 모델의 개발/성능 평가근거를 간략하게 설명한다.
목적	3. 연구목적을 구체적으로 명시하며, 모델 개발, 평가 또는 둘 다에 해당하는지 포함한다.
방법	4. 데이터 출처를 설명한다.
	5. 데이터 수집 시 적용된 선정기준과 환경을 설명한다.
	6. 예측모델이 예측하고자 하는 결과(예후모델의 경우 예측기간 포함)를 명시한다.
	7. 모델 유형, 모델 구축 단계 요약, 내부 검증방법[b]을 명시한다.
	8. 모델 성능 평가에 사용된 지표(예: 변별도, 보정, 임상적 유용성 등)를 명확히 기술한다.
결과	9. 참가자 수 및 결과 사건 수를 보고한다.
	10. 최종 모델의 예측변수를 요약한다†.
	11. 신뢰구간을 포함한 모델 성능 추정치를 보고한다.
고찰	12. 주요 결과에 대한 종합적 해석을 제시한다.
등록	13. 등록번호 및 등록기관(또는 저장소) 명칭을 명시한다.

TRIPOD, 개인별 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고(Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis); AI, 인공지능(artificial intelligence).
[a]이 체크리스트는 2020년에 발표된 TRIPOD for Abstracts statement [17]를 기반으로 하였으며, TRIPOD+AI statement와의 일관성을 위해 개정 · 업데이트되었음.
[b]예측모델 개발 연구에만 해당되는 항목임.

질 및 투명성 향상 네트워크(EQUATOR Network; https://www.equator-network.org/)를 통해서도 TRIPOD+AI 지침이 배포 및 홍보된다. TRIPOD+AI의 다국어 번역도 적극적으로 환영하며, 번역을 희망하는 경우 교신저자에게 연락하면 된다. 번역과정은 원저자와의 협력 및 승인을 포함한 구조적이고 사전 정의된 절차를 따르도록 하며, 번역 관련 추가 안내는 TRIPOD 웹사이트에서 확인할 수 있다(www.tripod-statement.org).

고찰

TRIPOD+AI는 국제적 다기관 다학제 합의과정을 통해 개발되었다. 이 지침은 회귀분석 또는 머신러닝 방법을 활용하여 예측모델을 개발하거나 평가(검증)하는 연구에 대해 최소한의 보고 권고사항을 제공한다. 지침 개발 당시에는 최근 급속히 발전하고 있는 파운데이션 모델 및 대형 언어 모델(예: ChatGPT 등)은 별도로 고려하지 않았으므로, TRIPOD+AI는 비생성형 모델을 주요 대상으로 한다. 그러나 이 지침의 많은 원칙은 보건 분야 생성형 AI 연구의 투명성 확보에도 적용 가능하다. 앞으로 TRIPOD+AI가 계속해서 유효성을 유지하고 AI 및 머신러닝의 발전을 반영하기 위해서는, 예를 들어 생성형 접근법에 대한 명시적 반영 등 주기적인 업데이트가 필요하다.

TRIPOD+AI는 TRIPOD 2015를 개정하여 개발되었으며, 문헌의 체계적 고찰, 델파이 설문조사, 온라인 합의 회의를 기반으로 권고사항을 정립하였다. TRIPOD+AI의 보고 항목을 충실히 기술하면, 연구방법의 질적 평가, 연구결과의 투명성 향상, 과도한 해석의 방지, 재현 및 복제 가능성 제고, 예측모델의 실제 적용 등에 모두 도움이 될 것이다. 체크리스트 항목은 최소한의 보고 기준으로, 저자는 데이터, 연구설계, 방법, 분석, 결과, 논의 등에서 추가적인 세부사항을 제공하는 것이 일반적이다.

TRIPOD+AI는 TRIPOD 2015에서는 부족하거나 명확히 언급되지 않았던 공정성(fairness) 이슈를 전반에 걸쳐 강조한다[33]. 예측모델 연구에서의 공정성은 특히 의료 분야에서 매우 중요하며, AI 및 머신러닝이 임상 의사결정 지원도구로 활용되면서 더욱 주목받고 있다. 이 맥락에서의 공정성이란, 예측모델이 특정 집단에 불리하게 작용하지 않으며, 기존의 건강불평등을 심화하지 않고 (이상적으로는 이를 완화 · 개선하는 방향으로) 설계 · 사용됨을 의미한다[92]. 공정성의 중요한 측면 중 하나는 모델 개발과 평가에 활용되는 데이터가 대표성과 다양성을 갖추고, 데이터 편향의 한계를 인지 · 관리 · 완화하는 것이다. 현재 STANDING Together 이니셔티브에서는 AI 보건 데이터 세트의 다양성, 포용성, 일반화 가능성을 높이기 위한 표준을 개발 중이다[62].

이상적으로는, 데이터에는 다양한 연령, 성별/젠더, 인종 · 민족, 건강상태 또는 동반 질환, 지역적 배경의 정보가 모두 포함되어야 하며, 이러한 다양성이 예측모델의 실제 사용 대상 인구를 대표해야 한다. 만약 모델 개발에 사용된 데이터가 의도한 전체 인구집단을 충분히 반영하지 못한다면, 데이터에 포함되지 않은 집단에서는 해당 모델이 기대한 만큼의 성능을 보이지 않을 수 있음을 명확히 밝혀야 한다. 모델 평가에 사용된 데이터가 목표 인구집단을 대표하지 못한다면, 특정 하위집단(개인적, 사회적, 임상적 속성별)의 예측 정확도 추정에 편향이 생기거나 오해를 일으킬 수 있다.

데이터 세트 내 소수 집단 또는 의료 소외 집단의 충분한 대표성 확보는 공정성을 달성하기 위한 핵심 요소이지만, 단순한 대표성만

Table 4. TRIPOD+AI 보고 지침 준수: 이해관계자별 잠재적 이익

사용자/이해관계자	권장 조치	잠재적 이익
학술기관	연구자에게 예측모델 개발, 평가, 적용 시 TRIPOD+AI 준수 권장 또는 의무화	예측모델 연구의 설계, 분석, 보고의 투명성 문화 증진
	초기 경력 연구자를 대상으로 투명하고 완전한 보고의 중요성과 이점을 교육, TRIPOD+AI 지침에 맞는 논문 · 학위 논문 작성 권장	산출 연구의 질, 책임성, 재현성, 복제 가능성, 유용성 향상
연구자	논문 작성 시 TRIPOD+AI 준수	보고의 완결성과 질 향상
		예측모델 논문에 요구되는 최소한의 세부 정보에 대한 인식 증가
		산출 연구의 질, 책임성, 재현성, 복제 가능성, 유용성 향상
		모델의 독립적 평가를 용이하게 하는 세부 정보 보고 증가
학술지 편집자	논문 제출 시 저자에게 TRIPOD+AI 및 체크리스트 작성 요구 또는 의무화	예측모델 논문에 대한 학술지 요구사항과 기대치에 대한 이해도 향상
	심사자에게 TRIPOD+AI 활용 권장	저자의 이해도 향상에 따른 심사 효율성 증가
		출판 논문의 질, 책임성, 재현성, 복제 가능성, 유용성 향상
심사위원	보고의 완결성 평가에 TRIPOD+AI 사용	심사 효율성과 질 향상
		누락된 중요 정보에 대한 구체적 피드백 제공 용이
연구비 지원기관	연구자가 연구비 신청 시 TRIPOD+AI 사용 권장 또는 의무화	연구결과의 활용성 증대, 불충분한 보고로 인한 연구 낭비 감소
	연구비 수혜 연구가 타인에게도 활용될 수 있도록 보장	
환자, 공공, 연구 참여자	저자, 심사자, 학술지, 연구비 지원기관의 TRIPOD+AI 준수 옹호	연구결과에 대한 신뢰도 향상
		예측모델 연구에 대한 이해도 증진, 연구 내 건강형평성 고려 촉진
		정밀의료 및 맞춤형 질환 관리에서 환자 보고 결과와 임상연구 결과 정렬
체계적 문헌고찰자/메타연구자	TRIPOD+AI로 보고 완결성 평가	위험도 평가도구와 병행 시 연구의 질 평가 향상(예: PROBAST)
	질 및 편향 평가 시 TRIPOD+AI 참고	메타분석에 필요한 데이터 확보 용이
정책 결정자	연구의 투명하고 완전한 보고를 위해 TRIPOD+AI 활용 권장 또는 의무화	예측모델 평가 또는 적용 결정이 완전하고 투명하게 보고된 정보에 근거하도록 보장
		근거 기반 정책 권고의 신뢰성 제고
규제 기관	임상 심사자가 의료기기 소프트웨어 등 예측모델 기반 제품 규제 심사 시 TRIPOD+AI로 임상시험 보고 완결성 평가	보고된 사용 목적과 규제상 의도 일치 확인
		의료기기 규제 심사 및 주요 임상시험 보고에서 모범사례와 일치 유도
		공통 표준 도입 유도로 제조사의 임상시험 보고 공개 장려
기술/의료기기 제조사	기술/기기 개발 · 제조에 필요한 모델 정보의 충분성 검증	공통 표준 도입 유도로 제조사의 임상시험 보고 공개 장려
의료인		
	구매 · 임상 활용 전 충분한 모델 정보 확인	모델 적용 대상군 및 지원 임상적 결정에 대한 이해도 향상
		예측결과에 대한 이해도와 한계 인식 증가
		연구결과에 대한 신뢰도 향상

TRIPOD, 개인별 예후 또는 진단을 위한 다변량 예측모델의 투명한 보고(Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis); AI, 인공지능(artificial intelligence).

으로는 완전한 공정성이 보장되지 않는다[61,93]. 이에 따라, TRIPOD+AI는 체크리스트 전반에 걸쳐 공정성 관련 항목을 포함하고 있다(배경 3c; 방법 5a, 7, 8a, 8b, 9c, 12f, 14; 결과 20b, 23a; 논의 25, 26번 항목 등).

의료 분야의 공정성이란, 예측모델의 개발, 평가, 실제 임상경로 내 적용 및 확산과정에서 환자, 대중, 임상의 등 다양한 이해관계자를 적극적으로 참여시키는 것 또한 포함한다[94]. 다양한 관점의 참여는, 예측모델이 모든 사람의 요구를 충족시키고 공정하게 사용되도록 설계 · 운영되는지 확인하는 데 기여하며, 건강형평성을 촉진한다. TRIPOD+AI는 대중 및 환자 참여에 관한 19번 항목을 통해 예측모델 연구에 환자와 대중 참여를 통합을 장려하고, 단순한 형식적 절차가 아닌 오픈 사이언스 및 참여의 원칙을 촉진하며, 더 높은 임상 및 대중 수용성을 지향한다.

TRIPOD+AI는 오픈 사이언스 관행을 강조한다[35]. 오픈 사이언스 관행은 예측모델 연구의 투명성, 재현성, 연구자 간 협력 증진에 필수적이다[95]. 연구 등록, 프로토콜 · 데이터 · 코드 · 예측모델의 공개 등은, 타 연구자가 결과를 검증하고 새로운 데이터에서 모델 성능과 안전성을 평가할 수 있도록 한다. 오픈 사이언스는 연구자들이 서로의 성과를 기반으로 추가 연구를 진행할 수 있게 하여, 보건의료 분야 발전의 효율성을 높인다. 이는 예측모델의 정확성, 신뢰성, 완전성을 높여 결과적으로 환자 치료에도 긍정적 영향을 줄 수 있다. 데이터가 개방적으로 공유될 경우 임상의와 연구자는 더 크고 다양한 환자 데이터를 바탕으로 모델을 개발 · 평가할 수 있으며[96], 이는 예측 정확도 향상 및 임상적 의사결정 개선으로 이어질 수 있다. 이에 TRIPOD+AI는 자금 출처(18a), 이해상충(18b), 프로토콜 공개(18c), 연구 등록(18d), 데이터 및 코드 공유(18e, 18f) 등 오픈 사이언스 관련 항목을 포함한다.

TRIPOD+AI의 주요 사용자 및 수혜자는 논문을 집필하는 연구자, 논문을 평가하는 학술지 편집자 및 동료 평가자, 그 밖의 이해관계자(예: 학술기관, 정책 입안자, 연구비 지원기관, 규제기관, 환자, 연구 참여자, 대중 등)로 예상된다(Table 4). 이 지침은 임상 예측모델 개발 및 검증 연구, 의학 연구 논문, 소프트웨어 · 도구 관련 보고 등 근거 기반 보고가 필요한 모든 분야에 적용될 수 있다.

학술지 편집인과 출판사는 TRIPOD+AI의 준수를 장려하기 위해 저자 안내문에 이를 명시하고, 논문 투고 및 심사 과정에서 그 사용을 의무화하며, 권고사항의 준수를 필수 요건으로 삼는 것이 바람직하다. 연구비 지원기관 역시, 예측모델 연구의 지원 신청 시 TRIPOD+AI 권고에 따른 보고 계획 제출을 요구함으로써 연구 낭비를 최소화하고 효율적 자원 활용을 도모할 것을 권고한다.

Additional information

[1]*Centre for Statistics in Medicine, UK EQUATOR Centre, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK*
[2]*Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands*
[3]*Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK*
[4]*National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, Birmingham, UK*
[5]*Department of Epidemiology, Harvard T H Chan School of Public Health, Boston, MA, USA*
[6]*Department of Development and Regeneration, KU Leuven, Leuven, Belgium*
[7]*Department of Biomedical Data Science, Leiden University Medical Centre, Leiden, Netherlands*
[8]*Department of Electrical Engineering and Computer Science, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA*
[9]*Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK*
[10]*University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK*
[11]*Department of Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-University of Munich, Munich, Germany*
[12]*Patient representative, Health Data Research UK patient and public involvement and engagement group*
[13]*Patient representative, University of East Anglia, Faculty of Health Sciences, Norwich Research Park, Norwich, UK*
[14]*Beth Israel Deaconess Medical Center, Boston, MA, USA*
[15]*Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA, USA*
[16]*Department of Biostatistics, Harvard T H Chan School of Public Health, Boston, MA, USA*
[17]*Institute of Health Informatics, University College London, London, UK*
[18]*British Heart Foundation Data Science Centre, London, UK*
[19]*Department of Computing, Imperial College London, London, UK*
[20]*Northwestern University Feinberg School of Medicine, Chicago, IL, USA*
[21]*Hardian Health, Haywards Heath, UK*
[22]*Section for Clinical Biometrics, Centre for Medical Data Science, Medical University of Vienna, Vienna, Austria*
[23]*Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada*
[24]*Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada*
[25]*Department of Computer Science, University of Toronto, Toronto, ON, Canada*
[26]*Vector Institute for Artificial Intelligence, Toronto, ON, Canada*
[27]*Department of Medicine, University of Cape Town, Cape Town, South Africa*
[28]*National Institute for Health and Care Excellence, London, UK*
[29]*The BMJ, London, UK*
[30]*Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
[31]*Department of Intelligent Medical Systems, German Cancer Research Centre, Heidelberg, Germany*
[32]*Wellcome Trust, London, UK*
[33]*Alan Turing Institute, London, UK*
[34]*Department of Bioethics, Hospital for Sick Children Toronto, ON, Canada*
[35]*Genetics and Genome Biology, SickKids Research Institute, Toronto, ON, Canada*
[36]*Australian Institute for Machine Learning, University of Adelaide, Adelaide, SA, Australia*
[37]*Medicines and Healthcare products Regulatory Agency, London, UK*
[38]*Department of Health Policy and Center for Health Policy, Stanford University, Stanford, CA, USA*
[39]*Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA*

[40]*Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands*

ORCID

Gary S. Collins: https://orcid.org/0000-0002-2772-2316
Karel G. M. Moons: https://orcid.org/0000-0003-2118-004X
Paula Dhiman: https://orcid.org/0000-0002-0989-0623
Richard D. Riley: https://orcid.org/0000-0001-8699-0735
Andrew L. Beam: https://orcid.org/0000-0002-6657-2787
Ben Van Calster: https://orcid.org/0000-0003-1613-7450
Marzyeh Ghassemi: https://orcid.org/0000-0001-6349-7251
Xiaoxuan Liu: https://orcid.org/0000-0002-1286-0038
Johannes B. Reitsma: https://orcid.org/0000-0003-4026-4345
Maarten van Smeden: https://orcid.org/0000-0002-5529-1541
Anne-Laure Boulesteix: https://orcid.org/0000-0002-2729-0947
Jennifer Catherine Camaradou: https://orcid.org/0000-0002-5742-2840
Leo Anthony Celi: https://orcid.org/0000-0001-6712-6626
Spiros Denaxas: https://orcid.org/0000-0001-9612-7791
Alastair K. Denniston: https://orcid.org/0000-0001-7849-0087
Ben Glocker: https://orcid.org/0000-0002-4897-9356
Robert M. Golub: https://orcid.org/0009-0000-3270-0632
Hugh Harvey: https://orcid.org/0000-0003-4528-1312
Georg Heinze: https://orcid.org/0000-0003-1147-8491
Michael M. Hoffman: https://orcid.org/0000-0002-4517-1562
André Pascal Kengne: https://orcid.org/0000-0002-5183-131X
Lena Maier-Hein: https://orcid.org/0000-0003-4910-9368
Bilal A. Mateen: https://orcid.org/0000-0003-4423-6472
Melissa D. McCradden: https://orcid.org/0000-0002-6476-2165
Lauren Oakden-Rayner: https://orcid.org/0000-0001-5471-5202
Johan Ordish: https://orcid.org/0000-0001-6911-2367
Richard Parnell: https://orcid.org/0000-0003-0044-3496
Sherri Rose: https://orcid.org/0000-0002-9076-8472
Karandeep Singh: https://orcid.org/0000-0001-8980-2330
Laure Wynants: https://orcid.org/0000-0002-3037-122X
Patricia Logullo: https://orcid.org/0000-0001-8708-7003

TRIPOD+AI working group/consensus meeting participants

Gary Collins (University of Oxford, UK), Karel Moons (UMC Utrecht, Netherlands), Johannes Reitsma (UMC Utrecht, Netherlands), Andrew Beam (Harvard School of Public Health, USA), Ben Van Calster (KU Leuven, Belgium), Paula Dhiman (University of Oxford, UK), Richard Riley (University of Birmingham, UK), Marzyeh Ghassemi (Massachusetts Institute of Technology, USA), Patricia Logullo (University of Oxford, UK), Maarten van Smeden (UMC Utrecht, Netherlands), Jennifer Catherine Camaradou (Health Data Research [HDR] UK public and patient involvement group, NHS England Accelerated Access Collaborative evaluation advisory group member, National Institute for Health and Care Excellence covid-19 expert panel), Richard Parnell (HDR UK public and patient involvement group), Elizabeth Loder (The BMJ), Robert Golub (Northwestern University Feinberg School of Medicine, USA [JAMA, at the time of the consensus meeting]), Naomi Lee (National Institute for Health and Clinical Excellence, UK; The Lancet, at the time of consensus meeting), Johan Ordish (Roche, UK; Medicine and Healthcare products Regulatory Agency, UK at the time of consensus meeting), Laure Wynants (KU Leuven, Belgium), Leo Celi (Massachusetts Institute of Technology, USA), Bilal Mateen (Wellcome Trust, UK), Alastair Denniston (University of Birmingham, UK), Karandeep Singh (University of Michigan, USA), Georg Heinze (Medical University of Vienna, Austria), Lauren Oaken-Rayner (University of Adelaide, Australia), Melissa McCradden (Hospital for Sick Children, Canada), Hugh Harvey (Hardian Health, UK), Andre Pascal Kengne (University of Cape Town, South Africa), Viknesh Sounderajah (Imperial College London, UK), Lena Maier-Hein (German Cancer Research Centre, Germany), Anne-Laure Boulesteix (University of Munich, Germany), Xiaoxuan Liu (University of Birmingham, UK), Emily Lam (HDR UK public and patient involvement group), Ben Glocker (Imperial College London, UK), Sherri Rose (Stanford University, US), Michael Hoffman (University of Toronto, Canada), and Spiros Denaxas (University College London, UK). The last seven participants in this list did not attend the virtual consensus meeting.

Authors' contributions

GSC and KGMM conceived the study and this paper and are joint first authors. GSC, PL, PD, RDR, ALB, BVC, XL, JBR, MvS, and KGMM designed the surveys carried out to inform the guideline content. PL analysed the survey results and free text comments from the surveys. GSC designed the materials for the consensus meeting with input from KGMM. All authors except SR, MMH, XL, SD, BG, and ALB attended the consensus meeting. PL took consolidated notes from the consensus meeting. GSC

drafted the manuscript with input and edits from KGMM. All authors were involved in revising the article critically for important intellectual content and approved the final version of the article. GSC is the guarantor of this work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflict of interest

All authors have completed the ICMJE uniform disclosure form at https://www.icmje.org/disclosure-of-interest/ and declare: support from the funding bodies listed above for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. GSC is a National Institute for Health and Care Research (NIHR) senior investigator, the director of the UK EQUATOR Centre, editor-in-chief of *BMC Diagnostic and Prognostic Research*, and a statistics editor for *The BMJ*. KGMM is director of Health Innovation Netherlands and editor-in-chief of *BMC Diagnostic and Prognostic Research*. RDR is an NIHR senior investigator, a statistics editor for *The BMJ*, and receives royalties from textbooks *Prognosis Research in Healthcare* and *Individual Participant Data Meta-Analysis*. AKD is an NIHR senior investigator. EWL is the head of research at *The BMJ*. BG is a part time employee of HeartFlow and Kheiron Medical Technologies and holds stock options with both as part of the standard compensation package. SR receives royalties from Springer for the textbooks *Targeted Learning: Causal Inference for Observational and Experimental Data and Targeted Learning: Causal Inference for Complex Longitudinal Studies*. JCC receives honorariums as a current lay member on the UK NICE COVID-19 expert panel and a citizen partner on the COVID-END COVID-19 Evidence Network to support decision making; was a lay member on the UK NIHR AI AWARD panel in 2020-22 and is a current lay member on the UK NHS England AAC Accelerated Access Collaborative NHS AI Laboratory Evaluation Advisory Group; is a patient fellow of the European Patients' Academy on Therapeutic Innovation and a EURORDIS rare disease alumni; reports grants from the UK National Institute for Health and Care Research, European Commission, UK Cell Gene Catapult, University College London, and University of East Anglia; reports patient speaker fees from MEDABLE, Reuters Pharma events, Patients as Partners Europe, and EIT Health Scandinavia; reports consultancy fees from Roche Global, Smith, the Future Science Group and Springer Healthcare (scientific publishing), outside of the scope of the present work; and is a strategic board member of the UK Medical Research Council IASB Advanced Pain Discovery Platform initiative, Plymouth Institute of Health, and EU project Digipredict Edge AI-deployed Digital Twins for COVID-19 Cardiovascular Disease. ALB is a paid consultant for Generate Biomedicines, Flagship Pioneering, Porter Health, FL97, Tessera, FL85; has an equity stake in Generate Biomedicines; and receives research funding support from Smith, National Heart, Lung, and Blood Institute, and National Institute of Diabetes and Digestive and Kidney Diseases. No other conflicts of interests with this specific work are declared.

Funding

This research was supported by Cancer Research UK programme grant (C49297/A27294), which supports GSC and PL; Health Data Research UK, an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities, which supports GSC; an Engineering and Physical Sciences Research Council grant for "Artificial intelligence innovation to accelerate health research" (EP/Y018516/1), which supports GSC, PD, and RDR; Netherlands Organisation for Scientific Research (which supports KGMM); and University Hospitals Leuven (COPREDICT grant), Internal Funds KU Leuven (grant C24M/20/064), and Research Foundation– Flanders (grant G097322N), which supports BVC and LW. The funders 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.

Data availability

Aggregated Delphi survey responses are available on the Open Science Framework TRIPOD+AI repository https://osf.io/zyacb/.

Acknowledgments

We thank the TRIPOD+AI Delphi panel members for their time and valuable contribution in helping to develop TRIPOD+AI statement. Full list of Delphi participants are as follows (in alphabetical order of first name): Abhishek Gupta, Adrian Barnett, Adrian Jonas, Agathe Truchot, Aiden Doherty, Alan Fraser, Alex Fowler, Alex Garaiman, Alistair Denniston, Amin Adibi, André Carrington, Andre Esteva, Andrew Althouse, Andrew

Beam, Andrew Soltan, Ane Appelt, Anne-Laure Boulesteix, Ari Ercole, Armando Bedoya, Baptiste Vasey, Bapu Desiraju, Barbara Seeliger, Bart Geerts, Beatrice Panico, Ben Glocker, Ben Van Calster, Benjamin Fine, Benjamin Goldstein, Benjamin Gravesteijn, Benjamin Wissel, Bilal Mateen, Bjoern Holzhauer, Boris Janssen, Boyi Guo, Brooke Levis, Catey Bunce, Charles Kahn, Chris Tomlinson, Christopher Kelly, Christopher Lovejoy, Clare McGenity, Conrad Harrison, Constanza Andaur Navarro, Daan Nieboer, Dan Adler, Danial Bahudin, Daniel Stahl, Daniel Yoo, Danilo Bzdok, Darren Dahly, Darren Treanor, David Higgins, David McClernon, David Pasquier, David Taylor, Declan O'Regan, Emily Bebbington, Erik Ranschaert, Evangelos Kanoulas, Facundo Diaz, Felipe Kitamura, Flavio Clesio, Floor van Leeuwen, Frank Harrell, Frank Rademakers, Gael Varoquaux, Garrett Bullock, Gary Collins, Gary Weissman, Georg Heinze, George Fowler, George Kostopoulos, Georgios Lyratzaopoulos, Gianluca Di Tanna, Gianluca Pellino, Girish Kulkarni, Giuseppe Biondi Zoccai, Glen Martin, Gregg Gascon, Harlan Krumholz, Herdiantri Sufriyana, Hongqiu Gu, Hrvoje Bogunovic, Hui Jin, Ian Scott, Ijeoma Uchegbu, Indra Joshi, Irene Stratton, James Glasbey, Jamie Miles, Jamie Sergeant, Jan Roth, Jared Wohlgemut, Javier Carmona Sanz, Jean-Emmanuel Bibault, Jeremy Cohen, Ji Eun Park, Jie Ma, Joel Amoussou, Johan Ordish, Johannes Reitsma, John Pickering, Joie Ensor, Jose L Flores-Guerrero, Joseph LeMoine, Joshua Bridge, Josip Car, Junfeng Wang, Karel Moons, Keegan Korthauer, Kelly Reeve, Laura Ación, Laura Bonnett, Laure Wynants, Lena Maier-Hein, Leo Anthony Celi, Lief Pagalan, Ljubomir Buturovic, Lotty Hook, Luke Farrow, Maarten Van Smeden, Marianne Aznar, Mario Doria, Mark Gilthorpe, Mark Sendak, Martin Fabregate, Marzyeh, Ghassemi, Matthew Sperrin, Matthew Strother, Mattia Prosperi, Melissa McCradden, Menelaos Konstantinidis, Merel Huisman, Michael Harhay, Michael Hoffman, Miguel Angel Luque, Mohammad Mansournia, Munya Dimairo, Musa Abdulkareem, Myura Nagendran, Niels Peek, Nigam Shah, Nikolas Pontikos, Nurulamin Noor, Oilivier Groot, Pall Jonsson, Patricia Logullo, Patrick Bossuyt, Patrick Lyons, Patrick Omoumi, Paul Tiffin, Paula Dhiman, Peter Austin, Quentin Noirhomme, Rachel Kuo, Ram Bajpal, Ravi Aggarwal, Richard Riley, Richiardi Jonas, Robert Golub, Robert Platt, Rohit Singla, Roi Anteby, Rupa Sakar, Safoora Masoumi, Sara Khalid, Saskia Haitjema, Seong Park, Shravya Shetty, Spiros Denaxas, Stacey Fisher, Stephanie Hicks, Susan Shelmerdine, Tammy Clifford, Tatyana Shamliyan, Teus Kappen, Tim Leiner, Tim Liu, Tim Ramsay, Toni Martinez, Uri Shalit, Valentijn de Jong, Valentyn, Bezshapkin, Veronika Cheplygina, Victor Castro, Viknesh Sounderajah, Vineet Kamal, Vinyas Harish, Wim Weber, Wouter Amsterdam, Xioaxuan Liu, Zachary Cohen, Zakia Salod, and Zane Perkins.

We thank Sophie Staniszewska (University of Warwick, UK) for chairing the HDR UK patient and public involvement and engagement meeting, where the TRIPOD+AI study and drak (pre-consensus meeting) checklist was presented and discussed; and Jennifer de Beyer for proofreading the manuscript (University of Oxford, UK).

Supplementary materials

The online version contains supplementary material available at https://doi.org/10.12771/emj.2025.00668

Supplement 1. TRIPOD+AI Expanded Checklist (Explanation & Elaboration Light).

Supplement 2. Fillable TRIPOD+AI checklist.

References

1. van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. J Clin Epidemiol 2021;132:142-145. https://doi.org/10.1016/j.jclinepi.2021.01.009
2. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, Lockowandt U. EuroSCORE II. Eur J Cardiothorac Surg 2012;41:734-745. https://doi.org/10.1093/ejcts/ezs043
3. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879-1886. https://doi.org/10.1093/jnci/81.24.1879
4. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117:743-753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579
5. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, Murray GD, Marmarou A, Roberts I, Habbema JD, Maas AI. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 2008;5:e165. https://doi.org/10.1371/journal.pmed.0050165
6. Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, Burckhardt P, Cooper C, Christiansen C, Cummings S, Eisman

JA, Fujiwara S, Glüer C, Goltzman D, Hans D, Krieg MA, La Croix A, McCloskey E, Mellstrom D, Melton LJ, Pols H, Reeve J, Sanders K, Schott AM, Silman A, Torgerson D, van Staa T, Watts NB, Yoshimura N. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int 2007;18:1033-1046. https://doi.org/10.1007/s00198-007-0343-y

7. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C, Schlussel MM, Gerry S, Black JA, Heus P, van der Schouw YT, Peelen LM, Moons KG. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416. https://doi.org/10.1136/bmj.i2416
8. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. BMJ 2019;367:l5358. https://doi.org/10.1136/bmj.l5358
9. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Bonten MM, Dahly DL, Damen JA, Debray TP, de Jong VM, De Vos M, Dhiman P, Haller MC, Harhay MO, Henckaerts L, Heus P, Kammer M, Kreuzberger N, Lohmann A, Luijken K, Ma J, Martin GP, McLernon DJ, Andaur Navarro CL, Reitsma JB, Sergeant JC, Shi C, Skoetz N, Smits LJ, Snell KI, Sperrin M, Spijker R, Steyerberg EW, Takada T, Tzoulaki I, van Kuijk SM, van Bussel B, van der Horst IC, van Royen FS, Verbakel JY, Wallisch C, Wilkinson J, Wolff R, Hooft L, Moons KG, van Smeden M. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. BMJ 2020;369:m1328. https://doi.org/10.1136/bmj.m1328
10. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med 2010;8:20. https://doi.org/10.1186/1741-7015-8-20
11. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med 2011;9:103. https://doi.org/10.1186/1741-7015-9-103
12. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. Open Med 2008;2:e49-e50.
13. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. Lancet 2014;383:267-276. https://doi.org/10.1016/S0140-6736(13)62228-X
14. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, Voysey M, Wharton R, Yu LM, Moons KG, Altman DG. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol 2014;14:40. https://doi.org/10.1186/1471-2288-14-40
15. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:1-12. https://doi.org/10.1371/journal.pmed.1001221
16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63. https://doi.org/10.7326/M14-0697
17. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-W73. https://doi.org/10.7326/M14-0698
18. Heus P, Reitsma JB, Collins GS, Damen JA, Scholten RJ, Altman DG, Moons KG, Hooft L. Transparent reporting of multivariable prediction models in journal and conference abstracts: TRIPOD for abstracts. Ann Intern Med 2020 Jun 2 [Epub]. https://doi.org/10.7326/M20-0193
19. Debray TP, Collins GS, Riley RD, Snell KI, Van Calster B, Reitsma JB, Moons KG. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. BMJ 2023;380:e071018. https://doi.org/10.1136/bmj-2022-071018
20. Debray TP, Collins GS, Riley RD, Snell KI, Van Calster B, Reitsma JB, Moons KG. Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster): explanation and elaboration. BMJ 2023;380:e071058. https://doi.org/10.1136/bmj-2022-071058
21. Snell KI, Levis B, Damen JA, Dhiman P, Debray TP, Hooft L, Reitsma JB, Moons KG, Collins GS, Riley RD. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ 2023;381:e073538.

https://doi.org/10.1136/bmj-2022-073538
22. Dhiman P, Whittle R, Van Calster B, Ghassemi M, Liu X, McCradden MD, Moons KG, Riley RD, Collins GS. The TRIPOD-P reporting guideline for improving the integrity and transparency of predictive analytics in healthcare through study protocols. Nat Mach Intell 2023;5:816-817. https://doi.org/10.1038/s42256-023-00705-6
23. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38:1276-1296. https://doi.org/10.1002/sim.7992
24. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, Collins GS. Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. Stat Med 2019;38:1262-1275. https://doi.org/10.1002/sim.7993
25. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, Moons KG, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441. https://doi.org/10.1136/bmj.m441
26. van Smeden M, de Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, Reitsma JB. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol 2016;16:163. https://doi.org/10.1186/s12874-016-0267-3
27. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, Reitsma JB. Sample size for binary logistic prediction models: beyond events per variable criteria. Stat Methods Med Res 2019;28:2455-2474. https://doi.org/10.1177/0962280218784726
28. Snell KI, Archer L, Ensor J, Bonnett LJ, Debray TP, Phillips B, Collins GS, Riley RD. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol 2021;135:79-89. https://doi.org/10.1016/j.jclinepi.2021.02.011
29. Archer L, Snell KI, Ensor J, Hudda MT, Collins GS, Riley RD. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. Stat Med 2021;40:133-146. https://doi.org/10.1002/sim.8766
30. Riley RD, Debray TP, Collins GS, Archer L, Ensor J, van Smeden M, Snell KI. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021;40:4230-4251. https://doi.org/10.1002/sim.9025
31. Riley RD, Collins GS, Ensor J, Archer L, Booth S, Mozumder SI, Rutherford MJ, van Smeden M, Lambert PC, Snell KI. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. Stat Med 2022;41:1280-1295. https://doi.org/10.1002/sim.9275
32. Riley RD, Snell KI, Archer L, Ensor J, Debray TP, van Calster B, van Smeden M, Collins GS. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. BMJ 2024;384:e074821. https://doi.org/10.1136/bmj-2023-074821
33. Wawira Gichoya J, McCoy LG, Celi LA, Ghassemi M. Equity in essence: a call for operationalising fairness in machine learning for healthcare. BMJ Health Care Inform 2021;28:e100289. https://doi.org/10.1136/bmjhci-2020-100289
34. McDermott MB, Wang S, Marinsek N, Ranganath R, Foschini L, Ghassemi M. Reproducibility in machine learning for health research: still a ways to go. Sci Transl Med 2021;13:eabb1655. https://doi.org/10.1126/scitranslmed.abb1655
35. UNESCO. UNESCO recommendation on Open Science [Internet]. UNESCO; 2023 [cited 2025 Jul 10]. Available from: https://www.unesco.org/en/open-science/about?hub=686
36. Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, Van Calster B, van Klaveren D, Venema E, Steyerberg E, Paulus JK, Kent DM. External validations of cardiovascular clinical prediction models: a large-scale review of the literature. Circ Cardiovasc Qual Outcomes 2021;14:e007858. https://doi.org/10.1161/CIRCOUTCOMES.121.007858
37. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JA, Hooft L, Kirtley S, Riley RD, Van Calster B, Moons KG, Collins GS. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. BMC Med Res Methodol 2022;22:101. https://doi.org/10.1186/s12874-022-01577-x
38. Andaur Navarro CL, Damen JA, van Smeden M, Takada T, Nijman SW, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KG, Hooft L. Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models. J Clin Epidemiol 2023;154:8-22. https://doi.org/10.1016/j.jclinepi.2022.11.015
39. Andaur Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KG, Hooft L. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. BMC Med Res Methodol 2022;22:12. https://doi.org/10.1186/s12874-021-01469-6

40. Rech MM, de Macedo Filho L, White AJ, Perez-Vega C, Samson SL, Chaichana KL, Olomu OU, Quinones-Hinojosa A, Almeida JP. Machine learning models to forecast outcomes of pituitary surgery: a systematic review in quality of reporting and current evidence. Brain Sci 2023;13:495. https://doi.org/10.3390/brainsci13030495
41. Munguia-Realpozo P, Etchegaray-Morales I, Mendoza-Pinto C, Mendez-Martinez S, Osorio-Pena AD, Ayon-Aguilar J, Garcia-Carrasco M. Current state and completeness of reporting clinical prediction models using machine learning in systemic lupus erythematosus: a systematic review. Autoimmun Rev 2023;22:103294. https://doi.org/10.1016/j.autrev.2023.103294
42. Kee OT, Harun H, Mustafa N, Abdul Murad NA, Chin SF, Jaafar R, Abdullah N. Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review. Cardiovasc Diabetol 2023;22:13. https://doi.org/10.1186/s12933-023-01741-7
43. Song Z, Yang Z, Hou M, Shi X. Machine learning in predicting cardiac surgery-associated acute kidney injury: a systemic review and meta-analysis. Front Cardiovasc Med 2022;9:951881. https://doi.org/10.3389/fcvm.2022.951881
44. Yang Q, Fan X, Cao X, Hao W, Lu J, Wei J, Tian J, Yin M, Ge L. Reporting and risk of bias of prediction models based on machine learning methods in preterm birth: a systematic review. Acta Obstet Gynecol Scand 2023;102:7-14. https://doi.org/10.1111/aogs.14475
45. Groot OQ, Ogink PT, Lans A, Twining PK, Kapoor ND, DiGiovanni W, Bindels BJ, Bongers ME, Oosterhoff JH, Karhade AV, Oner FC, Verlaan JJ, Schwab JH. Machine learning prediction models in orthopedic surgery: a systematic review in transparent reporting. J Orthop Res 2022;40:475-483. https://doi.org/10.1002/jor.25036
46. Lans A, Kanbier LN, Bernstein DN, Groot OQ, Ogink PT, Tobert DG, Verlaan JJ, Schwab JH. Social determinants of health in prognostic machine learning models for orthopaedic outcomes: a systematic review. J Eval Clin Pract 2023;29:292-299. https://doi.org/10.1111/jep.13765
47. Li B, Feridooni T, Cuen-Ojeda C, Kishibe T, de Mestral C, Mamdani M, Al-Omran M. Machine learning in vascular surgery: a systematic review and critical appraisal. NPJ Digit Med 2022;5:7. https://doi.org/10.1038/s41746-021-00552-y
48. Groot OQ, Bindels BJ, Ogink PT, Kapoor ND, Twining PK, Collins AK, Bongers ME, Lans A, Oosterhoff JH, Karhade AV, Verlaan JJ, Schwab JH. Availability and reporting quality of external validations of machine-learning prediction models with orthopedic surgical outcomes: a systematic review. Acta Orthop 2021;92:385-393. https://doi.org/10.1080/17453674.2021.1910448
49. Andaur Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KG, Hooft L. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ 2021;375:n2281. https://doi.org/10.1136/bmj.n2281
50. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22. https://doi.org/10.1016/j.jclinepi.2019.02.004
51. Yusuf M, Atal I, Li J, Smith P, Ravaud P, Fergie M, Callaghan M, Selfe J. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. BMJ Open 2020;10:e034568. https://doi.org/10.1136/bmjopen-2019-034568
52. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, Wang Y, Douiri A, Wolfe CD, Bray B. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One 2020;15:e0234722. https://doi.org/10.1371/journal.pone.0234722
53. Miles J, Turner J, Jacques R, Williams J, Mason S. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. Diagn Progn Res 2020;4:16. https://doi.org/10.1186/s41512-020-00084-1
54. Dhiman P, Ma J, Navarro CA, Speich B, Bullock G, Damen JA, Kirtley S, Hooft L, Riley RD, Van Calster B, Moons KG, Collins GS. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. J Clin Epidemiol 2021;138:60-72. https://doi.org/10.1016/j.jclinepi.2021.06.024
55. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JA, Hooft L, Kirtley S, Riley RD, Van Calster B, Moons KG, Collins GS. Risk of bias of prognostic models developed using machine learning: a systematic review in oncology. Diagn Progn Res 2022;6:13. https://doi.org/10.1186/s41512-022-00126-w
56. Araujo AL, Moraes MC, Perez-de-Oliveira ME, Silva VM, Saldivia-Siracusa C, Pedroso CM, Lopes MA, Vargas PA, Ko-

channy S, Pearson A, Khurram SA, Kowalski LP, Migliorati CA, Santos-Silva AR. Machine learning for the prediction of toxicities from head and neck cancer treatment: a systematic review with meta-analysis. Oral Oncol 2023;140:106386. https://doi.org/10.1016/j.oraloncology.2023.106386

57. Sheehy J, Rutledge H, Acharya UR, Loh HW, Gururajan R, Tao X, Zhou X, Li Y, Gurney T, Kondalsamy-Chennakesavan S. Gynecological cancer prognosis using machine learning techniques: a systematic review of the last three decades (1990-2022). Artif Intell Med 2023;139:102536. https://doi.org/10.1016/j.artmed.2023.102536
58. Collins GS, Whittle R, Bullock GS, Logullo P, Dhiman P, de Beyer JA, Riley RD, Schlussel MM. Open science practices need substantial improvement in prognostic model studies in oncology using machine learning. J Clin Epidemiol 2024;165:111199. https://doi.org/10.1016/j.jclinepi.2023.10.015
59. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JA, Hooft L, Kirtley S, Riley RD, Van Calster B, Moons KG, Collins GS. Overinterpretation of findings in machine learning prediction model studies in oncology: a systematic review. J Clin Epidemiol 2023;157:120-133. https://doi.org/10.1016/j.jclinepi.2023.03.012
60. Andaur Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KG, Hooft L. Systematic review finds "spin" practices and poor reporting standards in studies on machine learning-based prediction models. J Clin Epidemiol 2023;158:99-110. https://doi.org/10.1016/j.jclinepi.2023.03.024
61. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in healthcare. Annu Rev Biomed Data Sci 2021;4:123-144. https://doi.org/10.1146/annurev-biodatasci-092820-114757
62. Ganapathi S, Palmer J, Alderman JE, Calvert M, Espinoza C, Gath J, Ghassemi M, Heller K, Mckay F, Karthikesalingam A, Kuku S, Mackintosh M, Manohar S, Mateen BA, Matin R, McCradden M, Oakden-Rayner L, Ordish J, Pearson R, Pfohl SR, Rostamzadeh N, Sapey E, Sebire N, Sounderajah V, Summers C, Treanor D, Denniston AK, Liu X. Tackling bias in AI health datasets through the STANDING Together initiative. Nat Med 2022;28:2232-2233. https://doi.org/10.1038/s41591-022-01987-w
63. Kadakia KT, Beckman AL, Ross JS, Krumholz HM. Leveraging Open Science to accelerate research. N Engl J Med 2021;384:e61. https://doi.org/10.1056/NEJMp2034518
64. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, Altman DG, Moher D, Barber R, Denegri S, Entwistle A, Littlejohns P, Morris C, Suleman R, Thomas V, Tysall C. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. BMJ 2017;358:j3453. https://doi.org/10.1136/bmj.j3453
65. Camaradou JC, Hogg HD. Commentary: Patient perspectives on artificial intelligence; what have we learned and how should we move forward? Adv Ther 2023;40:2563-2572. https://doi.org/10.1007/s12325-023-02511-3
66. Finlayson SG, Beam AL, van Smeden M. Machine learning and statistics in clinical research articles-moving past the false dichotomy. JAMA Pediatr 2023;177:448-450. https://doi.org/10.1001/jamapediatrics.2023.0034
67. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, Moons K, Collins G, Moher D, Bossuyt PM, Darzi A, Karthikesalingam A, Denniston AK, Mateen BA, Ting D, Treanor D, King D, Greaves F, Godwin J, Pearson-Stuttard J, Harling L, McInnes M, Rifai N, Tomasev N, Normahani P, Whiting P, Aggarwal R, Vollmer S, Markar SR, Panch T, Liu X. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021;11:e047709. https://doi.org/10.1136/bmjopen-2020-047709
68. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020;2:e200029. https://doi.org/10.1148/ryai.2020200029
69. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, Denniston AK, Faes L, Geerts B, Ibrahim M, Liu X, Mateen BA, Mathur P, McCradden MD, Morgan L, Ordish J, Rogers C, Saria S, Ting DS, Watkinson P, Weber W, Wheatstone P, McCulloch P. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med 2022;28:924-933. https://doi.org/10.1038/s41591-022-01772-9
70. Hawksworth C, Elvidge J, Knies S, Zemplenyi A, Petyko Z, Siirtola P, Chandra G, Srivastava D, Denniston A, Chalkidou A, Delaye J. Protocol for the development of an artificial intelligence extension to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022. medRxiv [Preprint] 2023 Jun 1. https://doi.org/10.1101/2023.05.31.23290788
71. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guide-

lines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ 2020;370:m3210. https://doi.org/10.1136/bmj.m3210
72. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020;370:m3164. https://doi.org/10.1136/bmj.m3164
73. Cacciamani GE, Chu TN, Sanford DI, Abreu A, Duddalwar V, Oberai A, Kuo CJ, Liu X, Denniston AK, Vasey B, McCulloch P, Wolff RF, Mallett S, Mongan J, Kahn CE, Sounderajah V, Darzi A, Dahm P, Moons KG, Topol E, Collins GS, Moher D, Gill IS, Hung AJ. PRISMA AI reporting guidelines for systematic reviews and meta-analyses on AI in healthcare. Nat Med 2023;29:14-15. https://doi.org/10.1038/s41591-022-02139-w
74. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, Harrell FE, Martin GP, Moons KG, van Smeden M, Sperrin M, Bullock GS, Riley RD. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ 2024;384:e074819. https://doi.org/10.1136/bmj-2023-074819
75. Riley RD, Archer L, Snell KI, Ensor J, Dhiman P, Martin GP, Bonnett LJ, Collins GS. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ 2024;384:e074820. https://doi.org/10.1136/bmj-2023-074820
76. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. BMC Med 2023;21:70. https://doi.org/10.1186/s12916-023-02779-w
77. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010;7:e1000217. https://doi.org/10.1371/journal.pmed.1000217
78. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. Lancet 2019;393:1577-1579. https://doi.org/10.1016/S0140-6736(19)30037-6
79. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, van Smeden M, Riley RD, Moons KG. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008. https://doi.org/10.1136/bmjopen-2020-048008
80. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, Winchester CC, Tovey D, Goldman K, Hungin AP, Harrison N. ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLoS Med 2024;21:e1004326. https://doi.org/10.1371/journal.pmed.1004326
81. Olczak J, Pavlopoulos J, Prijs J, Ijpma FF, Doornberg JN, Lundstrom C, Hedlund J, Gordon M. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop 2021;92:513-525. https://doi.org/10.1080/17453674.2021.1918389
82. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, Arnaout R, Kohane IS, Saria S, Topol E, Obermeyer Z, Yu B, Butte AJ. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020;26:1320-1324. https://doi.org/10.1038/s41591-020-1041-y
83. Hernandez-Boussard T, Bozkurt S, Ioannidis JP, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020;27:2011-2015. https://doi.org/10.1093/jamia/ocaa088
84. Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health Care Inform 2021;28:e100251. https://doi.org/10.1136/bmjhci-2020-100251
85. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J. Artificial intelligence in dental research: checklist for authors, reviewers, readers. J Dent 2021;107:103610. https://doi.org/10.1016/j.jdent.2021.103610
86. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. NPJ Digit Med 2020;3:41. https://doi.org/10.1038/s41746-020-0253-3
87. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. Circ Cardiovasc Qual Outcomes 2020;13:e006556. https://doi.org/10.1161/CIRCOUTCOMES.120.006556
88. Kwong JC, McLoughlin LC, Haider M, Goldenberg MG, Erdman L, Rickard M, Lorenzo AJ, Hung AJ, Farcas M, Goldenberg L, Nguan C, Braga LH, Mamdani M, Goldenberg A, Kulkarni GS. Standardized Reporting of Machine Learning Applications in Urology: The STREAM-URO Framework. Eur

Urol Focus 2021;7:672-682. https://doi.org/10.1016/j.euf.2021.07.004
89. de Hond AA, Leeuwenberg AM, Hooft L, Kant IM, Nijman SW, van Os HJ, Aardoom JJ, Debray TP, Schuit E, van Smeden M, Reitsma JB, Steyerberg EW, Chavannes NH, Moons KG. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022;5:2. https://doi.org/10.1038/s41746-021-00549-7
90. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51-58. https://doi.org/10.7326/M18-1376
91. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1-W33. https://doi.org/10.7326/M18-1377
92. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021;3:e260-e265. https://doi.org/10.1016/S2589-7500(20)30317-4
93. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. Lancet Digit Health 2020;2:e221-e223. https://doi.org/10.1016/S2589-7500(20)30065-0
94. Mccradden M, Odusi O, Joshi S, Akrout I, Ndlovu K, Glocker B, Maicas G, Liu X, Mazwi M, Garnett T, Oakden-Rayner L, Alfred M, Sihlahla I, Shafei O, Goldenberg A. What's fair is… fair?: presenting JustEFAB, an ethical framework for operationalizing medical ethics and social justice in the integration of clinical machine learning: JustEFAB. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency; 2023 Jun 12-15; Chicago, USA. Association for Computing Machinery; 2023. p. 1505-1519. https://doi.org/10.1145/3593013.3594096
95. Thibault RT, Amaral OB, Argolo F, Bandrowski AE, Davidson AR, Drude NI. Open Science 2.0: towards a truly collaborative research ecosystem. PLoS Biol 2023;21:e3002362. https://doi.org/10.1371/journal.pbio.3002362
96. Riley RD, Tierney JF Stewart LA. Individual participant data meta-analysis: a handbook for healthcare research. Wiley; 2021.