

Supplementary Table 1. TRIPOD+AI 확장 체크리스트(간략 설명 및 해설 포함)

섹션/주제	항목 순서	주제 구분	체크리스트항목
<b>제목</b>			
제목	1	D;E	<p>연구가 다변량 예측 모델의 개발 또는 성능 평가임을, 대상 집단 및 예측할 결과와 함께 명시한다.</p> <ul style="list-style-type: none"> <li>정보를 담은 제목은 잠재적 독자와 체계적 문헌 고찰자들이 예측 모델 연구임을 쉽게 식별할 수 있도록 돕는다.</li> <li>모델이 예측하고자 하는 대상 집단과 예측 결과를 명확히 제공하는 정보성 제목을 사용한다.</li> </ul>
<b>초록</b>			
초록	2	D;E	<p>TRIPOD+AI 초록 체크리스트를 참조한다.</p> <ul style="list-style-type: none"> <li>TRIPOD+AI 초록 체크리스트의 각 항목을 포함하여 초록을 작성한다.</li> </ul>
<b>서론</b>			
배경	3a	D;E	<p>예측 모델의 개발 또는 평가가 이루어진 보건의료 맥락(진단 또는 예후 포함)과 그 근거를 설명하고, 기존 모델에 대한 인용을 포함한다.</p> <ul style="list-style-type: none"> <li>모델이 실제로 사용될 보건의료 환경을 설명한다.</li> <li>이미 예측 모델이 존재하는 경우, 새 모델 개발의 명확한 필요성과 근거를 제시한다.</li> <li>기존 모델을 평가하는 연구라면, 평가의 근거와 평가 대상 모델에 대한 인용을 명확히 제시한다.</li> </ul>
	3b	D;E	<p>예측 모델의 대상 집단과, 진료 경로(context of the care pathway) 내에서 모델의 의도된 목적 및 주 사용자를 설명한다(예: 의료인, 환자, 일반인 등).</p> <ul style="list-style-type: none"> <li>개발 또는 평가하는 모델의 목표 집단이 누구인지 설명한다(예: 특정 연령, 특정 국가, 특정 질병을 가진 사람들 등).</li> <li>모델의 사용 목적, 즉 임상적 의사결정이나 지침(예: 추가 검사를 위한 진료의뢰, 입원, 분류, 치료 시작, 생활습관 개선 등)에 사용될 의도와 진료 흐름 중 어느 시점에 사용되는지 구체적으로 설명한다.</li> <li>모델의 주된 사용자가 누구인지(의료진, 환자, 일반인 등)도 명시한다.</li> </ul>
	3c	D;E	<p>사회인구학적 집단 간에 이미 알려진 건강 불평등이 있다면 이를 기술한다.</p> <ul style="list-style-type: none"> <li>모델이 실제로 사용될 보건의료 환경의 맥락에서, 대상 집단 내 사회인구학적 그룹 간에 알려진 건강 불평등이 있다면, 이를 근거 문헌과 함께 기술한다.</li> </ul>
목표	4	D;E	<p>연구 목적을 명확히 기술하며, 본 연구가 예측 모델의 개발, 검증(성능 평가), 또는 두 가지 모두에 해당하는지 포함한다.</p> <ul style="list-style-type: none"> <li>연구의 모든 목적을 명확하게 진술하고, 예측 모델을 개발하는지, 성능을 평가하는지, 또는 둘 다를 수행하는지 명확히 설명한다.</li> </ul>
<b>방법</b>			
데이터	5a	D;E	<p>개발 및 평가 데이터 세트의 출처(예: 무작위 대조시험, 코호트, 일상진료, 레지스트리 등), 해당 데이터 사용 근거, 대표성을 각각 명확히 기술한다.</p>

			<ul style="list-style-type: none"> <li>• 모델 개발 및 성능 평가에 사용한 데이터가 무작위 대조시험, 코호트, 레지스트리, 전자의무기록 등에서 수집된 것인지 명확히 설명한다.</li> <li>• 연구가 기존 데이터를 활용하는지, 아니면 예측 모델 연구 목적을 위해 신규로 데이터를 수집하는지를 명확히 한다.</li> <li>• 기존 데이터(즉, 원래는 다른 목적으로 수집된 데이터)를 사용하는 경우, 해당 데이터 사용의 이유와(특히 환경·국가가 다른 경우) 연구의 의도된 대상 집단 및 맥락에 대한 적합성, 대표성에 대해 논의한다.</li> <li>• 모든 데이터 세트(개발 및 평가)에 대해 각각 데이터 출처를 기술한다.</li> <li>• 합성 데이터를 사용한 경우, 그 이유와 생성 방법(코드 포함, 18f 참고), 연구 내 활용 방식 등 모든 세부 정보를 제공한다.</li> </ul>
	5b	D;E	<ul style="list-style-type: none"> <li>• 연구 대상자 데이터의 수집 기간(등록 시작/종료일)을 명시하고, 필요 시 추적관찰 종료일을 명확히 보고한다.</li> <li>• 연구에 포함된 대상자 또는 데이터가 선정된 기간의 시작과 종료 날짜를 명확히 제시한다.</li> <li>• 예후 예측 모델의 경우 추적관찰 종료일이 중요하므로 반드시 보고한다.</li> </ul>
참가자	6a	D;E	<ul style="list-style-type: none"> <li>• 연구 환경(예: 1차 진료, 2차 진료, 일반 인구 등)의 주요 요소와 기관(센터) 수 및 위치를 명시한다.</li> <li>• 연구 대상자가 모집된 보건의료 환경을 설명한다.</li> <li>• 연구가 수행된 지리적 위치(최소 국가)와 센터(센터 수 포함)를 보고한다.</li> </ul>
	6b	D;E	<ul style="list-style-type: none"> <li>• 연구 대상자 선정(포함/제외) 기준을 명확히 기술한다.</li> <li>• 참가자 선정 기준을 보고해야 예측 모델의 적용 가능성과 일반화 가능성을 이해할 수 있다.</li> <li>• 연속형 변수(예: 연령 등)에 제한이 있다면, 포함된 대상자의 선정 기준에 사용한 범위 등 모든 제한 조건을 명확히 보고한다.</li> </ul>
	6c	D;E	<ul style="list-style-type: none"> <li>• 연구 대상자가 받은 치료 내역과, 모델 개발 또는 평가 과정에서 이러한 치료가 어떻게 다루어졌는지(해당 시)를 구체적으로 기술한다.</li> <li>• 추적관찰 시작 전이나 시작 시점에 받은 모든 치료를 보고하며, 예측 모델 개발 또는 평가 과정에서 이를 어떻게 처리했는지 기술한다(해당되는 경우).</li> <li>• 예측 모델 사용 시점부터 예측 결과(결과 측정)까지의 기간에 받은 치료가 결과에 영향을 줄 수 있다면 해당 치료 내역도 함께 보고한다(해당되는 경우).</li> </ul>
데이터 준비	7	D;E	<ul style="list-style-type: none"> <li>• 데이터 전처리 및 품질 점검 과정 전체를 설명하며, 이러한 과정이 주요 사회인구학적 집단 간에도 동일하게 적용되었는지 명시한다.</li> <li>• 데이터 정제 과정(예: 피쳐 엔지니어링, 원자료 변환, 피쳐 축소, 품질 점검 등)을 구체적으로 설명한다. 데이터 정제에 사용한 모든 코드는 공유한다(18f 참고).</li> <li>• 여러 출처(다른 연구, 코호트, 레지스트리 등)의 데이터를 분석에 사용한 경우, 결과 변수와 예측 변수의 조화(harmonisation) 과정도 설명한다.</li> <li>• 모든 데이터 전처리/정제 과정이 주요 사회인구학적 집단에서 동일하게 수행되었는지 여부를 반드시 명확히 한다.</li> <li>• 데이터 전처리/정제 과정이 매우 방대하다면 부록이나 별도 자료로 제공하는 것을 고려한다.</li> </ul>

결과(outcome)	8a	D;E	<p>예측하고자 하는 결과(outcome)와 예측 시간 범위(time horizon)를 명확히 정의하고, 결과의 평가 방법과 시점, 해당 결과를 선택한 근거, 그리고 평가 방법이 사회인구학적 집단 간에 일관되게 적용되었는지를 명시한다.</p> <ul style="list-style-type: none"> <li>진단 예측 모델의 경우, 결과(질병, 상태 등)를 명확히 정의하고 (널리 받아들여진) 표준 진단 기준(ground truth)이 사용되었는지 설명한다.</li> <li>예후 예측 모델(미래의 결과 예측)의 경우, 예측 시간 범위(예: 수술 후 28일 내 사망 위험, 골다공증 환자의 10년 내 골절 위험 등)와 추적관찰 중 결과 평가 빈도를 반드시 보고한다.</li> <li>표준 정의(예: ICD [1] 코드 등)를 사용했다면 해당 내용을 명확히 언급하고 인용한다.</li> <li>사회인구학적 집단 간에 결과 평가 기준이 다르면 이를 보고한다.</li> <li>특정 예측 변수가 결과 정의에 사용되거나 결과의 대리변수로 사용되지 않았는지 확인이 필요한 경우가 있으면 그 여부를 명확히 밝힌다.</li> </ul>
	8b	D;E	<p>결과 평가에 주관적 해석이 필요한 경우(예: 영상 판독 결과 등), 평가자의 인원 수, 자격, 인구통계적 특성을 모두 기술한다.</p> <ul style="list-style-type: none"> <li>만약 결과 측정 및 해석에 추가 교육이나 별도의 지침이 필요하다면, 해당 내용을 보고한다.</li> <li>교육이나 지침이 매우 방대한 경우, 부록(또는 별도 자료)으로 기술하는 것을 고려한다.</li> </ul>
	8c	D;E	<p>예측 결과 평가 시 눈가림 조치를 시행했다면 해당 내용을 보고한다.</p> <ul style="list-style-type: none"> <li>예측 대상 결과는 예측 변수 정보에 대해 눈가림으로 평가되어야 하며, 이는 주관적 해석이 필요한 결과에서 데이터(레이블) 누출을 방지하는 데 특히 중요하다.</li> <li>적절한 경우 결과 평가자가 어떤 정보에 접근할 수 있었는지와 결과 평가를 눈가림하기 위해 취한 구체적 조치가 무엇인지 명확히 기술한다.</li> </ul>
예측 변수	9a	D	<p>초기 예측 변수 선정 과정(예: 문헌 고찰, 기존 모델 참고, 전체 변수 사용 등)과 모델 구축 전 예측 변수 사전 선별 여부를 구체적으로 기술한다.</p> <ul style="list-style-type: none"> <li>모델 구축에 포함할 예측 변수의 초기 목록이 어떻게 선정되었는지, 예를 들어 문헌의 (체계적) 검토, 임상(도메인 전문가) 의견, 혹은 이용 가능한 데이터 내 모든 변수 사용 등 어떤 기준이 적용되었는지를 상세히 보고한다.</li> <li>모델 구축 전 결측값이 많은 변수나 예측 결과와 임상적으로 관련성이 낮은 변수 등을 제외하는 등 사전 선별을 실시한 경우, 그 방법을 구체적으로 설명한다.</li> <li>초기 예측 변수 목록이 매우 많다면 이를 부록이나 별도 자료로 제공하는 것이 바람직하다.</li> </ul>
	9b	D;E	<p>모델에 포함된 모든 예측 변수를 명확히 정의하고, 측정 방법과 시점(측정 단위, 범주형 변수의 모든 카테고리 포함), 예측 변수 평가 시 결과 또는 다른 예측 변수에 대한 눈가림 적용 여부를 구체적으로 기술한다.</p> <ul style="list-style-type: none"> <li>모든 예측 변수는 측정 단위, 범주형 변수의 경우 모든 카테고리를 명확히 제시하여, 독자가 모델을 재현·구현·성능 평가할 수 있도록 한다.</li> <li>예측 변수 값이 언제, 어떻게 측정되었는지 구체적으로 기술한다. 예측 변수는 모델 사용 시점 이전 또는 해당 시점에 측정되어야 한다.</li> </ul>

			<ul style="list-style-type: none"> <li>• 주관적 해석이 필요한 예측 변수는 다른 예측 변수와 독립적으로 해석하는 것이 바람직하므로(데이터 누출 방지), 이러한 눈가림 조치가 있었다면 반드시 보고한다.</li> <li>• 특히 진단 모델의 경우 예측 변수 측정은 결과(outcome) 값을 모르는 상태에서 수행해야 하며, 그렇지 않으면 예측 변수와 결과 간 연관성이 과대평가될 수 있다. 예측 변수 평가 시 결과에 대한 눈가림 조치가 있었다면 이를 구체적으로 명시한다.</li> <li>• 예측 변수 수가 매우 많은 경우, 본문에 모두 기술하는 것이 적절하지 않다면 모든 예측 변수를 명확히 정의하여 부록 또는 별도 자료로 제공하는 것이 중요하다.</li> </ul>
	9c	D;E	<p>예측 변수 평가에 주관적 해석이 필요한 경우, 평가자의 자격 및 인구통계적 특성을 보고한다.</p> <ul style="list-style-type: none"> <li>• 예를 들어, 영상 판독과 같이 주관적 해석이 필요한 예측 변수라면 평가자의 자격, 인구통계적 특성을 반드시 보고한다.</li> <li>• 측정 및 해석에 추가 교육이나 특정 지침이 필요한 경우 그 내용을 포함하며, 해당 정보가 방대하다면 부록(또는 별도 자료)에서 보고하는 것을 고려한다.</li> </ul>
표본 크기	10	D;E	<p>연구의 표본 크기 산정 방법을(모델 개발과 평가를 각각) 설명하고, 표본 크기가 연구 질문에 충분했는지 근거를 제시한다.</p> <ul style="list-style-type: none"> <li>• 표본 크기 산정은 모델 개발에 필요한 표본과, 모델 성능 평가에 필요한 표본을 구분하여 각각 제시한다(데이터가 신규 수집이든, 기존 데이터 활용이든 무관하다).</li> <li>• 표본 크기 계산에 사용한 모든 근거와 추정값을 명확히 보고한다.</li> <li>• 공식적인 표본 크기 산정을 하지 않았고 사용 가능한 모든 데이터를 활용한 경우라도, 데이터의 크기가 연구 질문에 충분한 근거가 있었는지 설명한다.</li> </ul>
결측값	11	D;E	<p>결측값 처리 방법을 구체적으로 기술하고, 데이터 생략 이유를 명확히 제시한다.</p> <ul style="list-style-type: none"> <li>• 결측값은 흔히 발생하는 문제이므로 모델에 포함된 각 예측 변수별 결측값 개수를 보고한다.</li> <li>• 결측값 처리 방법과 결측 발생 원인에 대한 가정이 있다면 이를 함께 명시한다.</li> <li>• 결측값으로 인해 연구 대상자(혹은 예측 변수)가 제외된 경우, 그 사실과 이유를 설명한다.</li> <li>• 결측값 대체(임putation)를 실시한 경우, imputation 방법 전체를 명확히 보고한다.</li> <li>• 결측값 대체가 훈련 데이터와 평가 데이터(테스트 데이터)에서 각각 독립적으로 실시되었는지(데이터 누출 방지) 여부도 확인하여 명시한다.</li> </ul>
분석 방법	12a	D	<p>데이터가 분석에 어떻게 사용되었는지(예: 모델 개발, 성능 평가)와, 데이터 분할 여부 및 분할 사유, 표본 크기 요건 준수 여부를 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>• 사용 가능한 데이터가 모델 개발과 성능 평가에 각각 어떻게 활용되었는지와 데이터 분할(예: 모델 개발, 하이퍼파라미터 튜닝, 성능 평가, 내부-외부 교차검증 등)의 실시 여부 및 그 사유를 구체적으로 설명한다.</li> <li>• 데이터 분할이 이루어진 경우, 분할 시 표본 크기 요건(10항목 참고)을 고려했는지, 분할된 데이터 크기가 연구 분석과 질문에 충분한지 함께 보고한다.</li> <li>• 데이터를 훈련(및 하이퍼파라미터 튜닝) 데이터와 평가(테스트) 데이터로 분할한 경우, 데이터 누출이 없었는지 반드시 확인하여 명시한다.</li> </ul>

		<ul style="list-style-type: none"> <li>• 데이터에 한 개인의 여러 기록이나 샘플이 포함되어 있고, 데이터가 훈련(및 하이퍼파라미터 튜닝 포함) 및 평가(테스트) 데이터로 분할된 경우, 동일 개인의 데이터가 분할된 데이터 세트에 중복 포함되어 있지 않은지(데이터 누출) 반드시 확인하고, 만약 포함된 경우 분석에서 이를 어떻게 처리했는지 구체적으로 기술한다(12c 항목 참고).</li> </ul>
12b	D	<p>모델 유형에 따라 예측 변수가 분석에서 어떻게 처리되었는지(함수 형태, 재표준화, 변환, 표준화 등) 구체적으로 기술한다.</p> <ul style="list-style-type: none"> <li>• 예측 변수가 분석 과정에서 변환, 재표준화, 또는 표준화된 경우, 그 방법을 상세히 설명한다.</li> <li>• 범주형 예측 변수의 경우, 샘플 크기 부족 또는 결과 사건 수 부족 등으로 범주를 통합(병합)했다면 그 이유와 방법을 구체적으로 기술한다.</li> </ul>
12c	D	<p>모델의 유형, 선정 근거, 모든 모델 구축 단계(하이퍼파라미터 튜닝 포함), 내부 검증 방법을 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>• 개발된 모델(예: 로지스틱 회귀, 콕스 회귀, 랜덤 포레스트, 신경망 등)의 유형과 각 모델 구축 방법의 사용 근거를[2] 명확히 설명한다. 이때 예측하고자 하는 결과의 유형과 모델이 실제로 어떻게 구현될 것인지도 함께 고려한다.</li> <li>• 각 모델별로 모델 구축의 모든 단계를 상세히 보고하며, 하이퍼파라미터 튜닝이 포함된 경우 튜닝한 하이퍼파라미터와 튜닝 방법을 구체적으로 기술한다. 모델 구축 방법이 많고 논문 분량 제한으로 본문에서 모두 서술하지 못하는 경우, 부록(또는 별도 자료)에서 전체 정보를 제공한다.</li> <li>• 둘 이상의 모델을 개발(예: 서로 다른 모델링 방법 적용)한 경우, 최종적으로 제시하는 모델 선정 기준을 명확히 설명한다(모델 성능 평가는 12e, 23항목 참고).</li> <li>• 모델 개발 단계에서 내부 검증(모델 성능 평가)에 어떤 방법(예: k-겹 교차검증, 부트스트래핑 등)을 사용했는지, 그리고 모든 모델 구축 단계(하이퍼파라미터 튜닝 포함)가 내부 검증 과정에서도 반복 적용되었는지 명확히 밝힌다.</li> <li>• 모델의 안정성(예: 예측 변수 선택, 예측 성능, 개별 예측값의 변화 등)을 검토하기 위해 사용한 방법(예: 부트스트래핑 등)도 구체적으로 기술한다(Riley &amp; Collins, Biom J 2023; 65: 2200302 [DOI: 10.1002/bimj.202200302]).</li> <li>• 데이터에 동일 개인의 여러 기록 또는 샘플이 포함된 경우, 모델 구축과 내부 검증 시 이를 어떻게 처리했는지 상세히 보고한다(예: k-겹 교차검증에서 한 개인의 모든 데이터가 동일 폴드에 포함되었는지 등, 데이터 누출 방지 확인 포함).</li> </ul>
12d	D;E	<p>클러스터(예: 병원, 국가 등) 간 모델 파라미터 값 또는 모델 성능 추정치의 이질성을 어떻게 처리하고 정량화했는지 명확히 기술한다(TRIPOD-Cluster [3] 체크리스트 추가 참고).</p> <ul style="list-style-type: none"> <li>• 모델 개발 또는 모델 성능 평가 시 데이터에 군집(예: 여러 연구의 개별 참가자 데이터 결합, 의료기관/병원, 국가 등)이 존재한다면, 군집을 고려한 근거와 방법을 명확히 기술한다.</li> <li>• 모델 파라미터 값 및 성능 추정치의 군집 및 이질성 처리를 위한 구체적 보고 권고는 TRIPOD-Cluster 체크리스트(Debray 등, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018])를 참고한다.</li> </ul>
12e	D;E	<p>모델 성능 평가에 사용한 모든 지표와 그래프(근거 포함)를 구체적으로 명시하며, 여러 모델을 비교하는 경우 그 방법도 함께 기술한다.</p> <ul style="list-style-type: none"> <li>• 모델 성능 평가는 최소한 분별력(discrimination)과 적합도(calibration, 적합도 곡선 포함) 지표를 보고하는 것이 일반적이다.</li> <li>• 예측 모델이 사건 발생 시점(time-to-event)을 예측한다면, 검열(censoring)과 관련하여 어떤 지표와 방법을 사용했는지 명확히 설명한다. 또한 경쟁 위험(competing risks) 처리 방법도 해당 시 보고한다.</li> </ul>

			<ul style="list-style-type: none"> <li>• 예후 예측 모델의 경우, 평가에 사용된 모든 시간점을 명확히 보고한다.</li> <li>• 모델 성능 평가에 사용한 그래프(예: 보정 곡선[평활화된 보정 곡선 포함], 의사결정 곡선 등)의 제작 방법을 상세히 보고한다.</li> <li>• 여러 모델을 비교(기존 모델 또는 여러 모델링 방법 등)하는 경우, 모델 비교에 사용된 방법과 우수성 판정 기준을 명확히 설명한다.</li> </ul>
	12f	E	모델 평가 후 모델 업데이트(예: 재보정 등)를 실시한 경우, 전체 집단 또는 특정 사회인구학적 집단/환경에서 업데이트된 모델의 방법과 세부 내용을 명확히 기술한다.
	12g	E	<p>모델 평가 시, 모델 예측값 산출 방법(수식, 코드, 객체, API 등)을 구체적으로 기술한다.</p> <ul style="list-style-type: none"> <li>• 별도의 데이터 세트에서 기존 모델을 평가(외부 검증)하는 경우, 각 대상자별 모델 예측값 산출 방법을 명확히 제시한다. 만약 모델을 자유롭게(공개적으로) 이용할 수 없는 경우, 예측값 산출 과정을 구체적으로 설명한다.</li> <li>• 회귀 모델 방정식을 평가하는 경우, 해당 방정식의 상세(방정식 자체, 원 개발 논문 인용 등)를 함께 제공한다.</li> <li>• 예측 모델 평가에서 수식이 없는 경우(예: 신경망, 랜덤 포레스트 등), 예측값 산출 방식(코드, 소프트웨어 객체, API 등)과 해당 자료의 위치(URL [4], DOI [5] 등)를 구체적으로 명시한다.</li> <li>• 모델 예측값을 이용해 개발 시 미리 정의되지 않은 위험군이나 분류 그룹을 새로 만든 경우, 그 방법과 이유도 함께 기술한다(15항목 참고).</li> </ul>
클래스 불균형	13	D;E	<p>클래스 불균형 처리를 실시했다면 그 이유와 구체적 방법, 그리고 이후 모델 또는 예측값 재보정 방법도 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>• 언더샘플링, 오버샘플링, SMOTE [6] 등 클래스 불균형 처리 방법을 사용한 경우, 왜 이러한 방법을 선택했는지와 어떤 방식으로 적용했는지(샘플 크기 변화 등)를 상세히 설명한다.</li> <li>• 클래스 불균형 보정은 모델 적합도(캘리브레이션)에 영향을 줄 수 있으므로, 보정 이후 확률 예측치가 과대평가되는 경향과(위험군 분류에 미치는 영향 포함) 모델 또는 예측값의 재보정 방법을 구체적으로 기술한다. (참고: van den Goorbergh 등, J Am Med Inform Assoc 2022; 29: 1525-1534 [DOI: 10.1093/jamia/ocac093])</li> </ul>
공정성	14	D;E	<p>모델의 공정성을 확보하기 위해 사용한 모든 접근법과 그 근거를 명확히 설명한다.</p> <ul style="list-style-type: none"> <li>• 공정성이란 예측 모델이 인종, 성별, 연령 등 개인 속성을 기반으로 특정 개인이나 집단에 불이익을 주지 않도록 하는 것으로, 모든 접근법과 사용 근거를 구체적으로 기술한다.</li> <li>• 모델 개발 및 성능 평가 시 데이터가 대상 집단을 대표할 수 있도록 하는 것이 중요하며, 연구자는 이를 입증할 수 있도록 시도해야 한다.</li> <li>• 만약 개발 데이터에 소수 집단 또는 특정 집단이 포함되지 않았다면, 이러한 집단에 대해 대표성 있는 데이터에서 모델을 평가하여 더 넓은 집단에 일반화 가능성을 높여야 한다.</li> </ul>
모델 출력	15	D	예측 모델의 출력값(예: 확률, 분류 등)을 명확히 제시하고, 분류가 이루어진 경우 그 이유와 임계값 선정 근거를 함께 기술한다.

			<ul style="list-style-type: none"> <li>대부분의 모델은 개별 대상자에 대한 확률 추정치를 출력하지만, 일부 모델은 이를 분류(예: 저위험군, 고위험군 등)로 변환하기도 하므로 이를 반드시 명확히 기술해야 한다. 위험군 등 분류나 그룹이 생성된 경우, 진료 경로(context of care pathway)에서 그 이유와 해당 위험군 분류가 임상적 의사결정에 어떻게 활용되는지 상세히 설명한다.</li> <li>분류 또는 위험군을 도출한 경우, 해당 임계값(예: 추정 확률 범위 등)이 문헌, 임상 지침, 통계적 고려, 임의 설정 등 어떤 근거로 결정되었는지 명확히 명시한다.</li> <li>개별 예측값에 대한 불확실성 구간(uncertainty interval)을 제시했다면, 그 산출 방법(예: 파라미터 추정치의 분산-공분산 행렬 또는 conformal prediction 사용 등)을 구체적으로 설명한다.</li> </ul>
개발-평가 차이	16	D;E	<p>개발 데이터와 평가 데이터 간의 의료 환경, 선정 기준, 예측 결과, 예측 변수 등에서 차이가 있다면 이를 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>예측 모델이 하나의 환경, 기관, 국가에서 개발되었다고 해서, 다른 환경·기관·국가에서도 유용하다고 볼 수 없다. 데이터 출처별로 선정 기준, 예측 결과, 예측 변수 정의가 (의도적으로) 다를 수 있으므로, 개발 데이터와 성능 평가에 사용된 데이터 간 차이를 상세히 보고하는 것이, 모델의 성능 해석 및 일반화 가능성 평가에 유익하다.</li> </ul>
윤리 승인	17	D;E	<p>연구를 승인한 기관연구윤리위원회(IRB) 또는 윤리위원회 명칭과, 참여자 동의 획득 또는 동의 면제 여부를 명확히 명시한다.</p> <ul style="list-style-type: none"> <li>연구에 IRB 혹은 윤리적 승인 없이 진행되었다면, 그 사실과 사유도 함께 명확히 기술한다.</li> </ul>
<b>오픈 사이언스</b>			
연구비	18a	D;E	<p>연구의 자금 출처와 연구비 제공자가 본 연구에서 수행한 역할을 명확히 밝힌다.</p> <ul style="list-style-type: none"> <li>연구비 지원 여부와 연구비 제공자가 연구에 미친 영향(기획, 설계, 해석, 보고 등)을 구체적으로 설명한다.</li> <li>모든 저자의 추가 연구비 수혜 내역도 보고한다.</li> </ul>
이해관계	18b	D;E	<p>모든 저자의 이해상충 및 재정 공개를 명확히 선언한다.</p> <ul style="list-style-type: none"> <li>연구 설계, 진행, 해석, 보고 등에 영향을 줄 수 있다고 독자가 판단할 만한 저자의 관계나 활동(재정적, 개인적, 전문적 등)은 모두 공개해야 한다.</li> </ul>
프로토콜	18c	D;E	<p>연구 프로토콜 접근 경로를 구체적으로 명시하거나, 프로토콜이 준비되지 않았음을 밝힌다.</p> <ul style="list-style-type: none"> <li>프로토콜의 이용 가능성과 접근 방법(예: 별도 자료, 논문 내 부록, Open Science Framework 등 저장소 포함)과 관련된 모든 정보를 URL, DOI 등과 함께 명확히 보고한다.</li> <li>프로토콜이 작성·공개되지 않았다면 그 사실과 사유를 명확히 밝힌다.</li> <li>프로토콜 명시 사항과 실제 연구 수행 내용에 중요한 차이가 있었다면 그 내용과 사유도 요약해 제시한다.</li> </ul>
등록	18d	D;E	<p>연구 등록 정보(등록 기관명, 등록번호 등)를 제공하거나, 연구가 등록되지 않았음을 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>연구가 clinicaltrials.gov, Open Science Framework 등에 등록되었다면 등록번호와 기관명, 등록 링크(또는 DOI 등)까지 상세히 보고한다.</li> <li>연구 등록이 이루어지지 않은 경우 그 사실을 명확히 명시한다.</li> </ul>
데이터 공유	18e	D;E	<p>연구 데이터의 이용 가능성과 접근 방법을 구체적으로 명시한다.</p> <ul style="list-style-type: none"> <li>데이터를 공개하는 경우, 데이터가 위치한 저장소(예: 공개 저장소, URL, DOI 등)와 데이터 획득·이용 방법, 이용 조건 및 제한 사항을 모두 구체적으로 제시한다. 데이터와 함께 데이터 사전(data dictionary)도 반드시 제공한다.</li> </ul>

코드 공유	18f	D;E	<ul style="list-style-type: none"> <li>• 데이터를 공유할 수 없는 경우 그 사유를 명확히 기술한다.</li> <li>• '합리적인 요청 시 데이터 제공'과 같은 막연한 표현은 피하고, 합리적인 요청의 기준과 조건을 구체적으로 명시한다.</li> </ul> <p>분석 코드의 공개 여부와 접근 방법을 구체적으로 명시한다[7].</p> <ul style="list-style-type: none"> <li>• 분석 코드(실행 방법 문서 포함)를 저장한 위치(예: 코드 저장소, DOI, 링크 등), 코드 획득 및 이용 조건·라이선스, 코드 및 관련 소프트웨어·패키지 버전(필요 시 컴퓨팅 환경 정보) 등 모든 세부 정보를 제공한다.</li> <li>• 분석 코드는 연구 결과 재현에 필요한 모든 코드(데이터 정제 코드 포함)이며, 해당 소프트웨어와 패키지, 버전까지 명확히 기술한다. 경우에 따라 하드웨어, 운영체제, CPU, 메모리 등 추가 환경 정보도 보고할 수 있다.</li> </ul>
<b>환자 및 공중 참여</b>			
환자 및 공중 참여	19	D;E	<p>연구의 설계, 진행, 보고, 해석 또는 결과 확산 등 전 과정에서 환자 및 공중이 참여한 내역을 구체적으로 보고하거나, 참여가 없었다면 그 사실을 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>• 환자 또는 공중이 연구 기획, 설계, 수행, 결과 보고 및 확산 과정에 어떻게 참여하였는지 구체적으로 설명한다.</li> <li>• 연구 결과가 환자나 공중에게 제공되었는지도 명시한다.</li> <li>• 환자 및 공중 참여 보고 시 GRIPP2 지침(Staniszewska 등, BMJ 2017; j3453 [DOI: 10.1136/bmj.j3453]) 사용을 고려한다.</li> <li>• 환자 및 공중이 연구 전 과정에 전혀 참여하지 않았다면 그 사실을 분명히 밝힌다.</li> </ul>
<b>결과</b>			
참가자	20a	D;E	<p>연구 대상자의 연구 내 흐름(flow), 결과 보유 여부별 대상자 수, 그리고(해당 시) 추적관찰 기간 요약을 구체적으로 설명한다. 필요하다면 도식을 활용할 수 있다.</p> <ul style="list-style-type: none"> <li>• 참가자 흐름을 설명할 때, 흐름도(flow diagram)를 사용하면 도움이 될 수 있다. 흐름도의 시작점은 참가자 모집 출처이며, 이후 선정 기준, 추적관찰(해당 시), 데이터 가용성 등 주요 단계를 시각적으로 표현할 수 있다.</li> <li>• 흐름도는 결측값이 있는 참가자 수, 결과 발생 수 등도 포함하면 유익하다.</li> <li>• 예후 또는 진단 연구에서 지연된 기준 검사(reference testing)가 있을 경우, 추적관찰 기간(예: 중앙값, 범위 등)을 반드시 요약하여 보고한다.</li> </ul>
	20b	D;E	<p>전체 데이터와(해당되는 경우) 각 데이터 출처 또는 환경별로, 주요 날짜, 주요 예측 변수(인구통계 포함), 받은 치료, 표본 크기, 결과 사건 수, 추적관찰 기간, 결측값 현황 등을 표 등으로 요약해 제시한다.</p> <ul style="list-style-type: none"> <li>• 주요 인구통계적 집단 간 차이도 보고한다. 모든 데이터 세트의 요약(결과 분포, 예측 변수[평균/중앙값, 표준편차/사분위수, 빈도], 치료 내역, 표본 크기, 결과 사건 수, 추적관찰 기간, 각 예측 변수별 결측값 개수/비율 등)은 표로 제시하는 것이 바람직하다.</li> <li>• 관심 있는 주요 인구통계적 집단 간 차이가 있다면 함께 보고하는 것이 유익하다.</li> </ul>
	20c	E	<p>모델 평가에서는, 개발 데이터와 비교하여 주요 예측 변수(인구통계, 예측 변수, 결과 등) 분포를 제시한다.</p> <ul style="list-style-type: none"> <li>• 기존 모델의 성능을 평가하는 연구(또는 개발 연구 내 평가 포함)에서는, 인구통계, 모델 내 예측 변수, 결과 등 주요 변수(평균/중앙값, 표준편차/사분위수, 빈도, 결측값 비율 등)의 분포를 개발 데이터와 평가 데이터 간 비교해 제시한다. 이는 표로 제시하고, 필요 시 결과 상태별로도 보고하는 것이 바람직하다.</li> </ul>

모델 개발	21	D;E	<p>각 분석(예: 모델 개발, 하이퍼파라미터 튜닝, 모델 평가)별 참가자 수와 결과 사건 수를 명확히 기술한다.</p> <ul style="list-style-type: none"> <li>예측 모델 연구에서, 데이터 분할, 하이퍼파라미터 튜닝 등 다양한 분석 단계마다 표본 크기와 결과 사건 수가 달라질 수 있으므로, 각 단계별(모델 개발, 평가 등)로 구체적으로 보고한다.</li> <li>데이터에 한 개인에 대한 여러 샘플이나 기록이 포함된 경우, 개별 대상자 수도 함께 보고한다.</li> </ul>
모델 명세	22	D	<p>예측 모델의 전체 정보를(수식, 코드, 객체, API 등) 제시하여, 새로운 개별 대상에 대한 예측과 제3자의 평가·구현이 가능하도록 한다. 재사용·이용 제한(무료, 독점 등)이 있다면 명시한다[8].</p> <ul style="list-style-type: none"> <li>예측 모델 개발 연구의 '산출물'은 바로 예측 모델이므로, 새로운 개별 대상자에 대해 예측이 가능하도록 해당 모델의 정보를(예: 수식, 코드 등) 구체적으로 제공하는 것이 중요하다(예: 자유롭게 이용 가능한 경우와 독점적으로 제한된 경우). 회귀 모델 등 수식으로 명확히 표현 가능한 경우 수식을 제공하고, 수식 형태로 표현할 수 없는 방법(예: 신경망, 앙상블 등)을 사용했다면 코드, 소프트웨어 객체, API 등의 이용 가능성 및 접근 경로를 명확히 제시하여 타인이 자신의 데이터로 모델을 평가하거나 실제 임상에 구현할 수 있도록 한다.</li> <li>여러 모델을 개발했다면 모든 모델의 이용 가능 정보(수식, 코드 등)를 함께 제공한다.</li> <li>타인이 새로운 개별 대상자에 대해 예측을 수행할 수 있도록 모델 활용 방법을 구체적으로 설명한다.</li> <li>제3자가 모델을 평가·구현·모니터링할 수 있도록 하드웨어 요구사항, 소프트웨어 및 패키지 정보 등도 명확히 보고한다.</li> <li>만약 상업적 이유 등으로 모델을 공개할 수 없다면, 그 사실을 분명히 명시하고 새로운 대상자 예측이나 제3자 평가가 가능한 접근 조건을 설명한다.</li> </ul>
모델 성능	23a	D;E	<p>모델 성능 추정치와 신뢰구간을 주요 하위집단(예: 사회인구학적 집단 등)에 대해서도 보고하며, 시각자료(그래프 등) 활용을 권고한다.</p> <ul style="list-style-type: none"> <li>12e 항목에서 기술한 모든 성능 평가 지표를 신뢰구간과 함께 제시한다.</li> <li>전체 대상뿐 아니라 성별, 인종 등 주요 관심 집단별로도 성능 추정치와 신뢰구간을 보고한다(공정성 평가 측면에서도 중요).</li> <li>보정 곡선(평활화된 보정 곡선 및 예측값 분포 등)과 의사결정 곡선 등 다양한 그래프를 활용해 성능 평가 결과를 시각적으로 제시한다.</li> <li>모든 평가(개발 데이터, 평가 데이터, 내부 검증 등)와, 예후 모델의 경우 평가 시점별로도 성능 추정치를 보고한다.</li> <li>모델의 안정성(성능 추정치, 부트스트랩 샘플 간 개별 예측값 변동성 등) 평가 결과도 함께 제시한다(Riley &amp; Collins, Biom J 2023; 65: 2200302 [DOI: 10.1002/bimj.202200302]).</li> <li>각 성능 추정치가 어떤 데이터에서 산출된 것인지 명확히 구분하여 보고한다.</li> </ul>
	23b	D;E	<p>모델 성능의 군집(클러스터) 간 이질성 평가가 이루어진 경우, 그 결과(신뢰구간 포함)를 명확히 보고한다(자세한 내용은 TRIPOD Cluster를 참고).</p> <ul style="list-style-type: none"> <li>여러 연구 데이터 통합, 의료기관/국가별 데이터 등에서 군집 또는 이질성을 고려해 평가했다면, 그 결과를 신뢰구간과 함께 보고한다(23a 항목 참고).</li> </ul>

모델 업데이트	24	E	<ul style="list-style-type: none"> <li>• 군집 및 이질성 처리를 위한 구체적 보고 권고는 TRIPOD-Cluster 체크리스트(Debray 등, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018])를 참고한다.</li> </ul> <p>업데이트된 모델과 그에 따른 성능을 포함한 모든 결과를 보고한다.</p> <ul style="list-style-type: none"> <li>• 모델 검증 이후 모델을 업데이트(예: 재보정, 재적합)한 경우, 업데이트된 예측 모델 및 이후 성능 평가 결과, 제3자 접근이나 재사용 제한까지 모두 명확히 보고한다(22항목 참고).</li> <li>• 업데이트된 모델의 성능 평가 결과도 반드시 보고한다(23a, 23b 항목 참고).</li> </ul>
<b>고찰</b>			
해석	25	D;E	<p>주요 결과에 대한 전반적인 해석을 제시하며, 연구 목적 및 기존 연구와의 관련성, 공정성 이슈를 함께 논의한다.</p> <ul style="list-style-type: none"> <li>• 연구 결과 해석 시, 본 연구의 결과가 기존 증거들과 어떻게 연결되는지 맥락을 명확히 설명한다. 이미 존재하는 예측 모델이 있다면, 기존 연구들과 비교하여 본 연구 결과를 논의한다.</li> <li>• 기존 예측 모델 성능을 평가한 다른 연구가 있는 경우, 그 결과도 함께 요약·논의하여 맥락을 제시하는 것이 중요하다.</li> <li>• 연구 결과 해석 시, 모델 개발·평가 과정에서 산출된 결과를 벗어나 과도한 해석이나 'spin'을 피해야 한다.</li> <li>• 모델 평가 데이터에서의 성능이 기존 연구 결과와 다를 경우, 그 차이에 대한 가능한 원인을 논의하는 것도 독자의 이해에 도움이 된다.</li> </ul>
한계	26	D;E	<p>연구의 한계(예: 대표성 부족, 표본 크기, 오버피팅, 결측값 등)와 그로 인한 편향, 통계적 불확실성, 일반화 가능성의 영향을 구체적으로 논의한다.</p> <ul style="list-style-type: none"> <li>• 한계 인식은 모든 과학 논문의 핵심 요소로, 연구 설계·수행·분석의 모든 측면에서 논의할 수 있다. 사용한 데이터의 대표성, 표본 크기, 오버피팅, 결측값/데이터 품질 등과 관련된 우려를 충분히 논의한다.</li> </ul>
활용성	27a	D	<p>예측 모델을 실제로 구현할 때 입력 데이터(예측 변수)가 부실하거나 존재하지 않는 경우, 이를 어떻게 평가·처리해야 하는지 구체적으로 설명한다.</p> <ul style="list-style-type: none"> <li>• 실제 진료 과정에서 예측 모델 사용 시 입력 변수가 존재하지 않을 때 어떻게 처리할지(즉시 대체 전략 등) 언급한다. 또한 결측값 대체 전략이 있다면 방법·결과에서 모두 이에 대해 평가·보고한다.</li> <li>• 마찬가지로, 구현 시 입력 데이터 품질이 낮은 경우(예: 영상 해상도, 데이터 형식 등) 어떻게 처리할지도 논의한다.</li> </ul>
	27b	D	<p>사용자가 입력 데이터 처리 또는 모델 사용에 직접 개입해야 하는지, 필요한 전문성 수준은 무엇인지 구체적으로 논의한다.</p> <ul style="list-style-type: none"> <li>• 예측 모델을 의도대로 사용하기 위해 사용자가 데이터 처리에 어떻게 개입해야 하는지 등의 세부 정보를 제공한다.</li> <li>• 모델 사용과 데이터 수집·입력·처리에 전문성 교육이나 훈련이 필요한 경우 그 내용을 상세히 기술한다.</li> </ul>
	27c	D;E	<p>모델의 적용 가능성 및 일반화 가능성 측면에서, 향후 추가 연구 필요성(다른 인구집단 또는 하위집단에서의 평가, 임상시험, 실제 진료과정에서의 적용 등)을 논의한다.</p> <ul style="list-style-type: none"> <li>• 모델의 다음 단계 평가가 필요한지, 예를 들어 다양한 인구집단·하위집단에서의 추가 평가, 임상시험에서의 평가, 진료과정 내 실제 적용 준비 여부 등에 대해 구체적으로 논의한다.</li> </ul>

1. ICD는 International Classification of Diseases (국제질병분류)를 의미한다.

2. 모든 모델 구축 접근법에 대해 별도로 적용한다.

3. TRIPOD-Cluster는 데이터의 군집 구조(예: 서로 다른 병원/센터)나 성능 이질성을 명시적으로 고려한 예측 모델 연구의 보고를 위한 체크리스트이다.
4. URL은 uniform resource locator(통합 자원 위치 지정자)를 의미한다.
5. DOI는 digital object identifier(디지털 객체 식별자)를 의미한다.
6. SMOTE는 synthetic minority oversampling technique(합성 소수집단 과샘플링 기법)를 의미한다.
7. 이는 분석 코드(예: 데이터 정제, 피처 엔지니어링, 모델 구축, 평가 등)에 해당한다.
8. 이는 새로운 개별 대상의 위험도를 산출하기 위한 모델 구현 코드에 해당한다.

원전: Collins GS, Moons KGM, Dhiman P, et al. BMJ 2024;385:e078378. doi:10.1136/bmj-2023-078378