

1 **Implementation of risk triaging in primary healthcare facilities in Sub-Saharan Africa: A systematic**
2 **review**

3
4 Short title: Risk triaging in primary healthcare in Africa
5

6
7 Mhairi Maskew^{1*}, Linda Alinafe Sande^{1*}, Mariet Benade², Vinolia Ntjiekelane¹, Nancy Scott², David
8 Flynn³, Sydney Rosen^{2†}
9

10 ¹Health Economics and Epidemiology Research Office, Department of Internal Medicine, School of
11 Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South
12 Africa.

13 ²Department of Global Health, Boston University School of Public Health, Boston, Massachusetts, USA.

14 ³Alumni Medical Library, Boston University, Boston, MA, USA
15

16 *Equally contributing authors
17

18 †Author for correspondence: Sydney Rosen, Boston University School of Public Health, 801
19 Massachusetts Ave, Boston, MA, USA. sbrozen@bu.edu
20
21

22 **ABSTRACT**

23

24 **Background:** One challenge facing treatment programs for HIV and other chronic conditions in sub-
25 Saharan Africa (SSA) is how to target interventions to optimize retention in care and other outcomes.
26 Most efforts to target interventions have identified predictive features among high risk patients after
27 negative outcomes have already been observed. An alternative for identifying patients at high risk of
28 negative outcomes is “risk triaging,” or identifying vulnerable or higher risk patients before they
29 experience an interruption in care or other negative outcome. We conducted a systematic review of the
30 use of risk triaging tools at the primary healthcare (PHC) level in SSA.

31 **Methods:** We searched PubMed and other databases for publications after 1 January 2012 that
32 reported development or implementation of risk triaging tools for PHC use in SSA. We extracted
33 information on three outcomes: 1) characterization of the risk triaging tools; 2) tool performance
34 metrics (sensitivity, specificity, positive and negative predictive value, area under the curve); and 3)
35 health system effects (efficiency, acceptability, resource utilization, cost). We report outcomes for each
36 eligible study and identify lessons for use of risk triaging.

37 **Results:** Of 1,876 articles identified, 28 were eligible for our review. Thirteen addressed HIV, 10 TB, 1
38 TB/HIV, and 4 other conditions. Approximately 60% used existing, retrospective data to identify
39 important risk factors for an outcome and then construct a scoring system, but no implementation of
40 these tools was reported. The remaining 40% designed a tool using existing data or experience and
41 reported implementation results. More than half (16/28, 58%) of the tools achieved sensitivities >80%;
42 specificity was much lower. Only one tool, the World Health Organization’s 4-symptom screen for
43 tuberculosis, had been scaled up widely. While most studies claimed that their tools could increase the
44 efficiency of healthcare delivery, none of the studies provided examples of tangible health system
45 impacts.

46 **Conclusion:** Most of the tools identified were at least somewhat successful in identifying potential risks
47 but uptake by health systems has been minimal. Although well-designed risk triaging tools have the
48 potential to improve health outcomes, implementation will require commitment at the policy,
49 operational, and funding levels.

50

51 **Keywords:** risk triaging, risk scoring, primary healthcare, sub-Saharan Africa

52

53 **INTRODUCTION**

54
55 One of the challenges facing treatment programs for HIV and other chronic conditions in sub-Saharan
56 Africa is how to target interventions to maximize long-term retention in care [1–4]. Ideally, interventions
57 aimed at improving adherence and retention should be offered to patients at higher risk of
58 disengagement from care or poor adherence, while not adding to the burden of care or utilizing
59 resources unnecessarily for lower risk patients who do not require additional attention. Identifying high
60 risk patients before they experience negative outcomes, however, remains a puzzle.

61
62 Traditionally, most efforts to target interventions aimed at improving retention in care or reducing
63 interruptions in treatment have identified high risk patients based on observed behaviour after negative
64 events have occurred. Patients who are observed to miss clinical visits or medication refills, for example,
65 are identified after the event and targeted for tracing, counselling, home-based care, and other services
66 that may help them resume scheduled care [5–10]. An alternative to this post-hoc strategy for
67 identifying high risk patients is “risk triaging,” or the process of identifying vulnerable or higher risk
68 patients before they experience an interruption in care. Predictive models developed into risk scoring or
69 triaging tools have a long history of use in hospitals, in high income countries, and for non-
70 communicable conditions [11–13].

71
72 While risk triaging or risk scoring can be a straightforward approach for evaluating clinical symptoms,
73 gauging a patient’s risk of a behaviour such as disengaging from chronic disease care is much more
74 challenging. Some broad (and largely unmodifiable) patient characteristics, such as sex and age, have
75 consistently been associated with higher loss to follow up rates in the literature[14,15] but these do not
76 identify subgroups that allow for practical or efficient targeting of interventions. Recognizing this,

77 multiple attempts have been made to identify patients who would benefit from early intervention.
78 These include tools developed to identify patients at risk of ART default and poor viral load outcomes
79 [16–19]; patients qualifying for same day ART initiation [20,21]; adults and children in need of HIV
80 testing [22,23]; and patients likely to return to care after disengagement [24]Results of these tools have
81 been mixed, in terms of both accuracy and uptake by healthcare providers. While much risk triaging is
82 conducted in hospitals and other high-care settings, it is at primary healthcare clinics that most patients
83 in SSA initiate and sustain ART. To our knowledge, the only example of HIV-related risk triaging
84 commonly used at the primary healthcare clinic level in SSA is the World Health Organization’s symptom
85 screen for tuberculosis, which triages patients with cough, fever, night sweats, or weight loss for TB
86 diagnostic tests [25–27].

87
88 The absence of risk triaging for patients on ART precludes providers from being able to target specific
89 types of support to specific patients who need that support. Patients who anticipate that transport fares
90 will pose an obstacle to visit attendance, for example, could be offered different interventions from
91 those who fear accidental disclosure of their HIV status, if these obstacles were identified in advance. As
92 part of an effort to design differentiated models of HIV treatment delivery for patients in their first six
93 months after antiretroviral therapy initiation, we are developing a risk triaging tool that could be used
94 by clinical and lay providers to match retention interventions to levels and types of risks[28,29]. To lay
95 the groundwork for this effort, we conducted a systematic review of the use of risk triaging tools at the
96 primary healthcare level in SSA. We asked two main questions: 1) Can risk triaging tools accurately
97 stratify patients into higher and lower risk groups in primary healthcare settings? and 2) How has risk
98 triaging been implemented in primary healthcare facilities in SSA?

99

100 **METHODS**

101

102 *Search strategy and study selection*

103

104 For this review, we defined risk triaging or a risk scoring “tool” as any score, system, test, or algorithm
105 that was designed to be used at point of care to differentiate participants into risk categories based on
106 their individual risk of the pre-defined study outcome. We note that we were seeking tools that predict
107 risk of an existing or future outcome, not diagnostic tools that confirm the presence of an existing
108 condition. (For example, we regarded the World Health Organization symptom screen for tuberculosis
109 to be a risk triaging tool, and thus included it in our review, while TB diagnostic tests were excluded.)

110 Risk may pertain to an existing condition such as TB, in which case risk triaging selects a subset of all
111 potential patients for further diagnostic investigation. This category of risk triaging usually happens
112 when the diagnostic process is too expensive, invasive, or otherwise undesirable to offer to all patients.
113 Alternatively, risk may pertain to a condition that does not exist yet but is more likely in some
114 individuals than in others, such as future disengagement from treatment. Risks in this category may be
115 amenable to intervention before the condition occurs. (We also acknowledge that the use of the term
116 “risk” has been criticized for appearing to place blame on or stigmatize those labeled as high risk. Since
117 “risk” remains a standard term in the literature, however, we have chosen to continue to use it here.)

118

119 We searched peer-reviewed publications and conference abstracts that reported development and/or
120 implementation of risk triaging tools published after 1 January 2012. We limited the review to studies
121 reporting entirely or primarily outcomes from sub-Saharan Africa and restricted our search to tools
122 designed for or implemented in primary health care settings; any tools that required hospital in-patient
123 admissions or required use of specialized laboratory services not typically available in outpatient settings

124 were excluded. We considered community-based and other off-site service delivery models as
125 outpatient primary health services and included these. Mathematical models were included if they
126 served as a data source or as part of the process for developing a risk triaging tool but were not defined
127 as tools themselves, and thus studies reporting only the development of mathematical models were
128 excluded. Studies that reported purely qualitative data or were limited to theoretical concepts were
129 excluded. A full list of inclusion and exclusion criteria and our search strategy are provided in S1 Table.
130 The protocol for this systematic review is included as S1 Text and was registered on PROSPERO
131 (CRD42022328209) and followed PRISMA guidelines.

132
133 The database search was conducted on 24 May 2022 and updated on 25 July 2022. We searched the
134 following databases: NLM's PubMed, Wiley's Cochrane Library, the Cumulative Index of Nursing and
135 Allied Health Literature (CINAHL) via Ebsco; EMBASE via Elsevier; Clarivate's Web of Science (specifically
136 the Web of Science Core Collection); and PsycINFO via Ebsco. Composite search strings were developed
137 using keywords according to the Cochrane Collaboration guidelines' PICO model criteria [30]. We then
138 utilized the PICO components (populations, interventions, comparators, outcomes) to develop the
139 composite search terms. The full list of search terms for each database searched is provided in S2 Table.

140
141 After the search was completed and results deduplicated, two authors (LS and VN) conducted an initial,
142 blinded, independent screening of article titles and abstracts for eligibility. Conflicts were reviewed after
143 unblinding and resolved between the screening authors and a third author (MM). Full texts of
144 potentially eligible articles were retrieved, imported to a reference management program, and
145 evaluated for inclusion in the review. Reasons for exclusion after full-text review were recorded. Using a
146 snowball approach, we also manually searched reference lists of articles considered relevant to identify

147 other sources for potential inclusion. The search results are summarized in accordance with the PRISMA-
148 P reporting protocol (S3 Table).

149

150 *Outcomes*

151

152 For purposes of this review, we were interested in both the tool's statistical performance metrics and
153 the extent to which the tool was taken up for routine practice, including the method of implementation,
154 its usefulness in practical stratification of patient populations, and effects on the healthcare system. The
155 primary outcomes for this review were thus threefold: 1) Characterization of the risk triaging tools; 2)
156 tool performance metrics (sensitivity, specificity, positive and negative predictive value (PPV and NPV),
157 and area under the curve); and 3) health system considerations including efficiency of clinic operations,
158 patient waiting times, provider acceptability, resource utilization, cost, and scalability.

159

160 *Data extraction and analysis*

161

162 We developed a standardized form to extract key information from each eligible article, including (1)
163 article details; (2) patient population; (3) description of the risk triaging tool; (4) risk triaging tool
164 performance metrics and (5) effect of the risk triaging tool implementation on the health system. After
165 reporting this information descriptively, we plotted performance metrics to present point estimates for
166 sensitivity and false positive rates (1-specificity). Where available, negative predictive value (NPV),
167 positive predictive value (PPV), and area under the curve (AUC) were also plotted as individual points.

168

169 Because of the tremendous heterogeneity in the sources included in this review, which varied by
170 condition addressed, risk triaging approach, types of tools, specific outcomes predicted by the tools,

171 population studied, and many other parameters, we did not attempt to conduct any pooled analyses or
172 generate aggregate values to represent the full set of sources, which we believe would present
173 misleading summary values. We instead present the results individually, with sufficient information
174 about each tool to allow results to be compared as appropriate to each reader's needs.

175

176 *Methodological quality assessment*

177

178 Risk of bias for each study was assessed using the Joanna Briggs Institute (JBI) critical appraisal checklist
179 for diagnostic test accuracy studies [31]. The JBI checklist asks a set of 10 closed ended questions for
180 every study included to evaluate risk of bias in the design, conduct and analysis in the studies meeting
181 the inclusion criteria. A "yes" response indicates the criterion for risk of bias is low or not present while
182 "no" responses indicating risk of bias is present. We assigned a score of 1 point to each "No" response
183 and calculated a bias risk percentage score using the total number of questions for each study for which
184 we were able to assess a response as the denominator. We then set the following thresholds for risk
185 assessment: 1) score of 20 percentage points or less indicated low risk of bias; 2) score of 20-50
186 percentage points indicated moderate risk of bias and; 3) score of more than 50 percentage points
187 indicated high risk of bias.

188

189 **RESULTS**

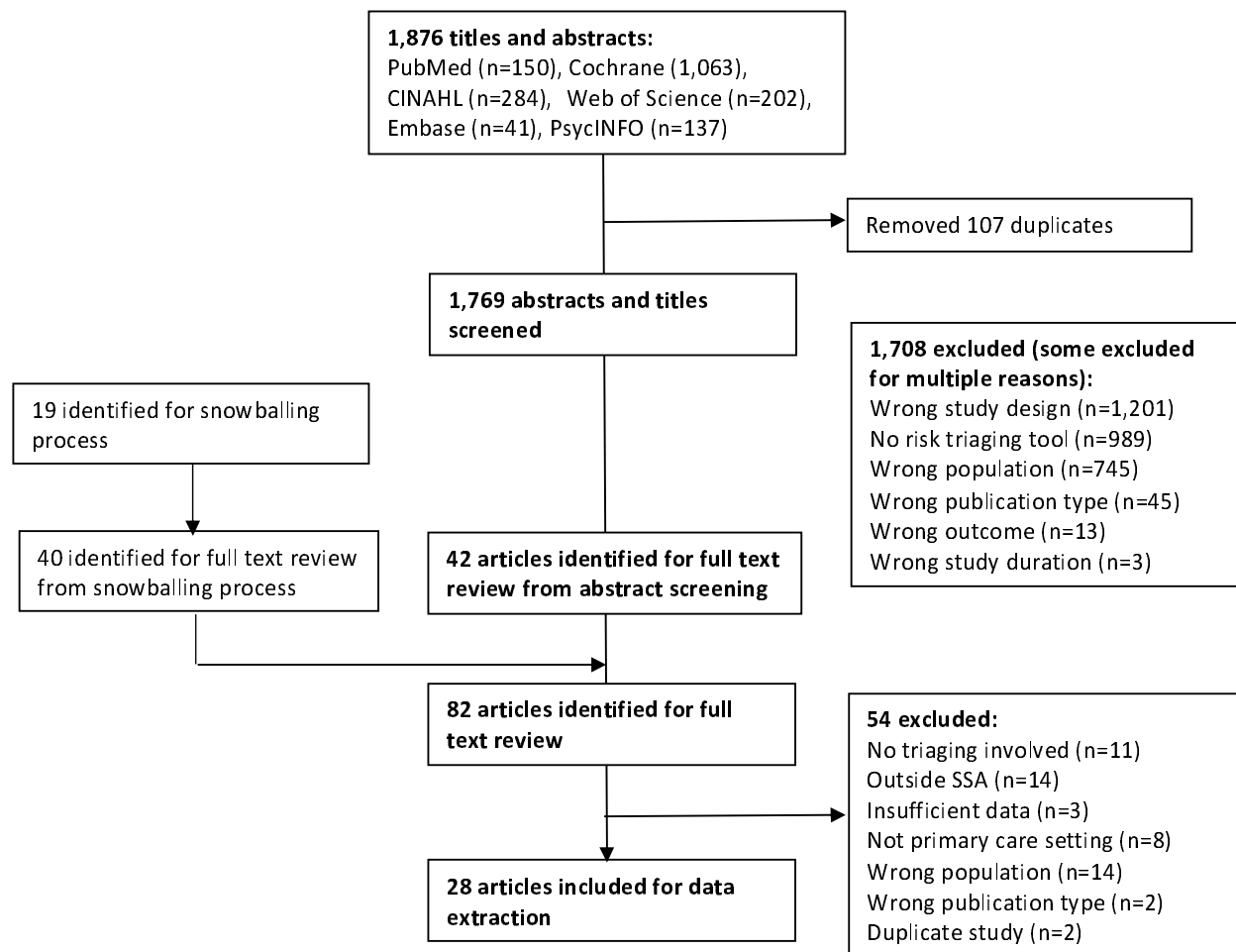
190

191 *Search results*

192

193 Our search strategy yielded a total of 1,876 articles from the six databases searched (Figure 1). We
194 removed 107 articles duplicated across databases and screened abstracts and titles from 1,769 articles.

195 Of these, 42 articles met the inclusion criteria. Of 1,708 articles that did not meet the inclusion criteria,
196 some were excluded for multiple reasons, but having the wrong study design (n=1,201), no evidence of
197 a risk triaging tool being implemented in the study (n=989), or involving the wrong study population (n=
198 745) were the most frequent reasons for article exclusion. From abstracts, we also identified 19
199 published systematic reviews to include in the snowballing process. We reviewed the reference lists of
200 these articles and identified a further 40 articles meeting our inclusion criteria, for a total of 82 articles
201 extracted for full text review. We excluded 54 of these; half of these exclusions were due to either study
202 locations outside sub-Saharan Africa (n=14) or not in a primary health setting (n=12).
203



204 **Figure 1: PRISMA flow chart of article screening process**

205

206 The final data set included 28 articles (Table 1) that reported on risk triaging tools from 14 observational

207 cohorts, 6 clinical trials, and 8 other study designs, with data collected and analysed between 2004 and

208 2020. The most frequent implementation countries were South Africa (n=12), Kenya (n=7), and Uganda

209 (n=6); 8 were multi-country studies. Risk triaging tools were implemented in a diverse range of facility

210 settings within the context of primary health care, including health facility-based outpatient clinics,

211 community-based sites, and dedicated research facilities

212

213 *Characterization of risk triaging tools evaluated*

214

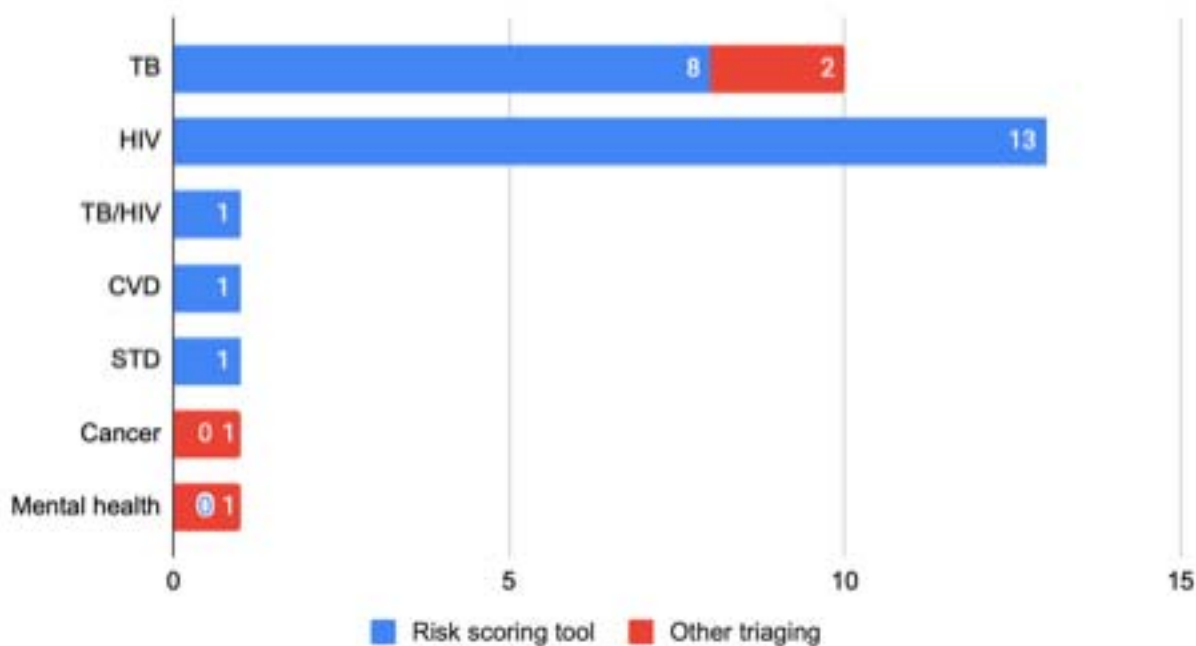
215 Table 1 and Figure 2 describe the risk triaging tools identified in our search. Of the 28 studies, 13
216 addressed HIV, 10 TB, 1 TB/HIV, and 4 other conditions. In total, 24 (86%) of the triaging tools generated
217 risk scores, while the remaining 4 utilized clinical algorithms, symptom screens, or clinical checks and
218 diagnostics. The risk score approach was most frequently applied to HIV (n=13) and TB (n=8), while the
219 other triaging approaches were applied to HIV and TB combined, sexually transmitted diseases, cancer,
220 and mental health conditions.

221

222 **Figure 2. Characterization of triaging tools stratified by condition for which the tool was developed**
223 **(n=28)**

224

225



226

227

Table 1: Description of risk scoring and triaging tools

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|--|-----------|--|---|---|--|---|--|--|--|
| Alamo 2012, Uganda[32] | TB | Community-based site and outpatient HIV clinic, 2007-2010 | HIV-positive adults (≥16) with self-reported two-week history of cough. | 749 | To assist in diagnosis of pulmonary TB in absence of simple diagnostic tools for smear negative PTB among HIV+ people in resource limited settings. | WHO's 2007 algorithm for diagnosing pulmonary TB, which recommended TB diagnosis for HIV-positive individuals with either 1) cough > 2 weeks; or 2) no cough or cough < 2 weeks, weight loss, night sweats, fever, breathlessness, enlarged glands, chronic headache, or abnormal chest x-ray | No, used presence of symptoms but no score | Implementation and evaluation of an existing tool in a study setting. | Validated with sputum culture |
| Auld 2020, Botswana and South Africa[33] | HIV | 22 health facilities (5 hospitals and 17 clinics), 2012-14 | HIV-positive, ART-naïve adolescents and adults (≥12 years). | Derivation and temporal validation datasets: $n = 5,553$ External validation dataset: $n = 1,077$ | To identify patients with the highest risk of early (6-month) ART mortality for care intensification | Risk scoring tool predicting risk of early HIV treatment mortality using sex, pregnancy status, number WHO TB symptoms, WHO stage, fever, CD4 count if available, hemoglobin. | Yes | Tool was developed and validated using existing datasets, not implemented. | Internally validated by splitting the dataset 1:1 into derivation and temporal validation datasets. Externally validated on a different dataset from a different country (South Africa). |
| Auld 2021, Botswana and South Africa[34] | TB | 22 health facilities (5 hospitals and 17 clinics), 2012-14 | HIV-positive adolescents and adults (≥12 years). | Internal derivation and validation datasets: $n = 5,418$ External validation datasets: XPHACTOR (SA:Gauteng) :1807; TB | To better facilitate early TB diagnosis, and to allow for TB risk differentiation into low, moderate, and high-risk patients with the aim of providing | Clinical risk score composed of sex, number of WHO TB symptoms, smoking habit, temperature, body mass index and hemoglobin level. | Yes | Tool was developed and validated using existing datasets, not implemented. | Internally validated by splitting the dataset 1:1 into derivation and validation datasets. Externally validated on 3 separate datasets from South Africa and |

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|--|----------------------|---|--|--|---|--|--|---|--|
| | | | | Fast Track (SA: Gauteng, Limpopo, North West); Gugulethu Cohort (SA: Cape Town): 488 | differentiated models of care. | | | | performance also compared with WHO TB symptom screen. |
| Aunsborg 2020, Guinea-Bissau[35] | TB | HIV clinic of Hospital Nacional Simão Mendes (HNSM) in Bissau, 2014 | HIV-positive, ART-naïve adults (≥15 years), non-pregnant, no TB treatment in last year | 164 | To improve TB case finding among people living with HIV | TBscore, a TB screening tool composed of TB symptoms and signs using checklist and physical examination (self-reported cough, dyspnoea, night sweats, haemoptysis, chest pain, anaemia, tachycardia, positive finding on lung auscultation, fever, BMI<18, low MUAC. | Yes | Implementation and evaluation of an existing tool in a study setting. | Smear microscopy, Xpert MTB/RIF and chest X-ray among patients with scores above a given threshold |
| Awolude 2021, Nigeria[36] | Cervical cancer | Large ART treatment programme in Ibadan, Nigeria, 2020 | HIV-positive women on ART | 98 | To identify patients at high risk of pre-cancerous cervical lesions for same-day treatment and to prevent unnecessary treatment | Screening for HPV using visual inspection to identify individuals for same day treatment. | No, used results of visual inspection but no score | Implementation and evaluation of an existing tool in a study setting. | Histological confirmation of biopsies at a pathology laboratory |
| Baik 2020, South Africa and Uganda[37] | Tuberculosis | South Africa: 56 primary care clinics in Limpopo Province, 2016-2018 Uganda: 4 healthcare facilities in Kampala, 2018-2019 | Adults (>15) presenting with TB symptoms | Derivation cohort: n = 1,387 External validation cohort: n = 387 | To reduce pretreatment loss to follow-up associated with waiting for microbiological tests by predicting active TB among patients | Prediction score requiring only information accessible to clinicians in resource-limited settings (age, sex, HIV status, TB symptoms and duration, comorbidities, education, smoking) | Yes | Tool was developed using an existing dataset, not implemented | 10-fold cross validation and external validation on a different dataset from a different country (Uganda). |
| Balcha 2014, Ethiopia[38] | Tuberculosis/ HIV | 5 health centres in Oromia | HIV-positive, ART-naïve adults (≥18) | 791 | To further categorize presumptive TB | Clinical scoring algorithm added to standard TB symptom screen, including | Yes | Tool was developed using an existing | Smear microscopy, culture, and |

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|--|----------------------|---|--|--|---|--|----------------------------|--|--|
| | | Region, 2011-2013 | | | patients after a positive screen by the WHO TB symptom screening tool | cough, Karnofsky scores ≤ 80 , MUAC < 20 cm, hemoglobin < 10 g/dL, and peripheral lymphadenopathy. Score categorized study subjects into 3 risk groups depending on likelihood of TB. | | dataset, not implemented | Xpert MTB/RIF |
| Balkus 2016, Malawi, South Africa, Kenya, Uganda, Zimbabwe, Zambia, and Tanzania[39] | HIV | Developed, not implemented, 2005-2011 | HIV-negative women not pregnant or breastfeeding; age range varied for 3 cohorts but all ≥ 18 | Derivation and internal validation cohort: $n = 4,834$; External validation cohort 1: $n = 2,848$; External validation cohort 2: $n = 1,804$ | To identify African women at highest risk of HIV acquisition over a period of one year to prioritize prevention interventions | VOICE risk scoring tool predicting risk of HIV acquisition using age, marital status, alcohol use, source of income, partner's sexual risks, STIs, HSV-2 status (modified tool excluded final two indicators, which are laboratory-based). | Yes | Tool was developed using an existing dataset, not implemented | Developed and internally validated using 10-fold cross validation; externally validated with two different datasets. |
| Balkus 2018, Malawi, South Africa, Uganda, and Zimbabwe[40] | HIV | 15 research clinics, 2012-2015 | > 18 | 2,539 | To identify women at high risk of becoming HIV positive with the aim of targeting PrEP | Adaption of an existing risk score predicting HIV acquisition (VOICE risk score). Tool consisted of demographic, behavioural, clinical, and partner characteristics. | Yes | Tool was developed using an existing dataset, not implemented. | Clinical trial design. |
| Brown 2012, Malawi[41] | HIV | Kamuzu Central Hospital and Bwaila Hospital outpatient STI clinics, Lilongwe, dates uncertain | HIV-positive, adult (≥ 18) first-time testers with partners | 329 | To identify partners who are unlikely to report for testing on their own after partner notification and should receive provider-assisted notification | Risk score to identify partners unlikely to report for testing on their own, including index patient characteristics (education, age, STI status) and partner characteristics (sex, main or non-main partner, relationship duration). | Yes | Tool was developed using an existing dataset, not implemented | Trial design and bootstrap sampling. |
| Hanifa 2017, South Africa[42] | Tuberculosis/ HIV | 2 hospital-based and 2 community health | HIV+ adults (≥ 18) with positive WHO symptom | 1,048 | To predict probability of TB among people living | Clinical score composed of ART status, WHO TB symptoms, BMI, and CD4 count. | Yes | Tool was developed using an existing dataset, not | Split data 50:50 into derivation and validation datasets by |

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|---|--------------|--|--|--------------|---|---|--|--|--|
| | | centres in Gauteng Province, 2012-2014 | screen | | with HIV attending routine HIV care after a positive screen from the WHO TB symptom screen. | | | implemented | median date of enrolment. |
| Kahle 2013, Botswana, Kenya, Rwanda, South Africa, Tanzania, Uganda, and Zambia[43] | HIV | Multiple clinics in study countries from 3 clinical trial cohorts, 2004-2010 | Heterosexual, HIV-serodiscordant adult (≥ 18) couples with HIV+ individual not on ART | 8,651 | To identify couples at highest risk for HIV-1 transmission among serodiscordant couples. | Risk scorecard composed of age of HIV-negative partner, marital status, number of children, unprotected sex, circumcision status of male, and viral load. | Yes | Tool was developed using an existing dataset, not implemented | Internal and external validation, 10-fold cross validation on the internal validation dataset. |
| Kerschberger 2021, Eswatini[44] | HIV | Outpatient ward in Nhlngano facility, Shiselweni region, 2019-2020 | Adults (18-49) at risk for acute and early HIV infection based on discordant HIV test results or symptoms | 795 | To predict acute and early HIV infection. | Symptom- and behavior-based predictors risk score (PRS). Indicators are serodiscordant RDT result, female sex, self-reported swollen glands, and fatigue. | Yes | Tool was developed using an existing dataset, not implemented | Compared predictive ability with other predictors risk scores in similar settings. |
| Khan 2014, South Africa[45] | Tuberculosis | 2 HIV clinics in Eastern Cape Province, 2011-2012 | HIV-positive adults (≥ 18) on or off ART | 737 | To detect TB cases earlier and to exclude TB before IPT without laboratory tests. | Standard WHO four-symptom screen questionnaire prior to widespread adoption | Yes | Implementation and evaluation of an existing tool in a study setting | Sputum specimen and chest X-ray. |
| Maskew 2020, South Africa[46] | HIV | 3 public-sector primary clinics in Gauteng Province, 2018 | HIV-positive adults (≥ 18) not yet on ART | 296 | To accelerate same-day ART initiation among people living with HIV who also have mild TB symptoms | SLATE II clinical algorithm. Tool was composed of 4 screens: symptom report, medical history, brief physical examination, and patient readiness assessment. | No, used results of screening algorithm but no score | Adaptation and evaluation of an existing tool in a clinical trial | Nonblinded individually randomized clinical trial design. |
| Mbu 2018, Cameroon[47] | Tuberculosis | Regional hospital and TB reference laboratory in | HIV-positive adults (≥ 18) diagnosed with HIV in | 1,14\9 | To identify TB cases. | Standard WHO four-symptom screen questionnaire. | Yes | Implementation and evaluation of an existing tool in a study setting | Validated with sputum culture. |

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|--------------------------------|---|--|---|--------------|--|--|--|--|--|
| | | Bamenda, 2012-2013 | past month | | | | | | |
| Mellins 2017, South Africa[48] | Psychiatric and substance use disorders | Asenze study clinic in KwaZulu Natal Province, period unclear | Adult caregivers of preschool children | 322 | To identify mental health concerns using lay counselors. | Lay version of the Client Diagnostic Questionnaire (CDQ), brief diagnostic mental health screening tool to identify patients to be referred for clinical diagnosis. | No | Implementation and evaluation of an existing tool in a study setting | Compared results with a clinical CDQ administered by a clinical psychologist. |
| Mlisana 2013, South Africa[49] | Acute HIV infection | Cohort study Durban, 2004-2005 | HIV-negative women at high risk of infection (3 sexual partners in past 3 months or self-identified as sex workers) | 245 | To identify individuals with acute HIV infection | Locally developed Clinical Evaluation Tool to estimate risk score composed of age plus multiple signs and symptoms and physical exam (no laboratory tests required) | Yes | Tool was developed using an existing dataset, not implemented | Compared performance in two groups. |
| Modi 2016, Kenya[50] | Tuberculosis | 15 HIV clinics in Nyanza Province, 2011-2012 | Persons living with HIV aged ≥ 7 years attending HIV care | 738 | To identify TB cases as part of intensified case finding | Paper-based comparison of 3 TB symptom screening algorithms screening for TB as part of intensified case finding: WHO screening algorithm, Kenya MoH clinical screening algorithm and the Improving Diagnosis of TB in HIV-infected persons (ID-TB/HIV) study algorithm. | No, used screening algorithms but no score | Implementation and evaluation of three existing tools | Smear microscopy, mycobacterial culture, and Xpert MTB/RIF |
| Njuguna 2022, Kenya[51] | HIV | Public health facilities providing ART (national), 2015-2017 | HIV-positive adults (≥ 18) with ≥ 1 viral load result after ART initiation | 3,968 | To predict non-suppression of viral load suppression 18 months after ART initiation. | Scorecard to identify virally unsuppressed patients, using age, sex, BMI, education, HIV disclosure status, years between HIV diagnosis and ART initiation. | Yes | Tool was developed using an existing dataset, not implemented | 10-fold cross validation on the derivation dataset and external validation another dataset |
| Peebles 2020, South Africa[52] | HIV | 9 public health facilities providing ART in 5 provinces, 2015-2018 | HIV-negative women 18-35 years | 5,573 | To predict risk of HIV acquisition with the aim of targeting PrEP provision. | ECHO risk score to predict the risk of becoming HIV positive within 1 year, using age, marital status, alcohol intake, HIV prevalence, province, number partners, partner having >1 partners, | Yes | Implementation and evaluation of an existing tool | External validation on a different dataset and compared to VOICE risk score |

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|---|-------------------------|---|---|--------------|--|---|---------------------------------------|---|---|
| Rosen 2019, South Africa, and Kenya[53] | HIV | South Africa: 3 public-sector primary care clinics in Gauteng Province, 2017-2018 Kenya: 3 public-sector HIV outpatient clinics in Kericho, Kapsabet, and Kombewa Districts, 2017-2018 | HIV-positive adults (≥18) not yet on ART. | 1,089 | To determine eligibility for same-day ART initiation in settings without use of point of care (POC) instruments. | condom use, STIs. SLATE clinical algorithm. Tool was composed of 4 screens: symptom report, medical history, brief physical examination, and patient readiness assessment. | No, screening algorithm | Implementation and evaluation of a newly developed tool. | Nonblinded individually randomized clinical trial design. |
| Semitala 2019, Uganda[54] | Tuberculosis | 2 urban clinics in Kampala, 2013-2016 | HIV-positive adults (≥18) initiating ART with a pre-ART CD4 count ≤350 cells/uL | 1,839 | To reduce high rates of unnecessary Xpert testing by adding C-reactive protein (CRP) using a point of care (POC) assay to TB screening algorithm | Point of care C-reactive protein-based test plus standard WHO symptom screen. Used capillary blood using a standard CRP POC assay. | No, POC test plus screening algorithm | Implementation and evaluation of an existing tool | Xpert MTB/RIF. |
| Skogmar 2014, Ethiopia[55] | HIV/ tuberculosis | Outpatient clinics at 6 health centres, 1 zonal hospital, and 1 regional hospital in Oromia Region, 2010-2012 | TB-diagnosed adults (>18) presenting at TB outpatient clinics | 1,116 | To assess immunosuppression among PLHIV patients coinfecting with TB to determine the timing of ART initiation without requiring laboratory facilities | Clinically based scoring system. Contents include mid-upper arm circumference, OHL, gingivitis, shortness of breath, conjunctival pallor, gender, age, and oral candidiasis. | Yes | Tool was developed using an existing dataset, not implemented | Tool developed using regression analysis observing significant association with CD4 cell count. |
| Surka 2014, South Africa[56] | Cardiovascular diseases | Nyanga District, Cape | 35-75 | 477 | To screen for the risk of | Non-laboratory screening tool to generate CVD risk | Yes | Implementation and evaluation of | Compared the performance of |

| Study, country | Condition | Setting and dates of data collection | Population | Sample size* | Identified goal of tool | Tool description | Quantitative scoring tool? | Implementation status | Tool validation process |
|------------------------|---|---|--|--------------|---|---|----------------------------|---|--|
| | | Town, dates uncertain | | | cardiovascular diseases | score using age, gender, diabetes status, systolic blood pressure and BMI. Paper and digital versions tested. | | an existing tool | paper-based and digital versions of the tool |
| Wahome 2013, Kenya[57] | HIV | 2 research clinics in coastal Kenya, 2008-2012 | Men who have sex with men (MSM) with sexually transmitted diseases | 449 | To identify acute/early HIV infection among MSM attending an STD clinic | Cohort-derived risk screening score (CDRSS) based on existing University of North Carolina-Malawi Risk Screening Core† to identify risk of acute/early HIV infection among MSM. Contents include fever, fatigue, symptomatic STI, diarrhea, age<30, discordant HIV tests. | Yes | Implementation and evaluation of an existing tool | Compared performance with another tool. |
| Wahome 2018, Kenya[58] | HIV | 2 research clinics in coastal Kenya, 2005-2016 | Adult men (18-49) who have sex with men (MSM) | 753 | To target the provision of PrEP among MSM | Empiric risk score composed of age, sex of partner in past 3 months, sexual exposure and condom use in the past week, receptive anal intercourse in past 3 months and group sex in the past 3 months. | Yes | Tool was developed using an existing dataset, not implemented | None reported. |
| Wall 2021, Rwanda[59] | Neisseria gonorrhoeae (NG) and Chlamydia trachomatis (CT) | Centre for Family Health Research (CFHR), Kigali, 2016-2020 | Sexually transmitted infection-symptomatic women | 791 | To facilitate diagnosis of NG/CT among symptomatic women | Risk algorithm that includes age, education level, employment, condom use, genital itching, candida, and bacterial vaginosis | No, screening algorithm | Tool was developed using an existing dataset, not implemented | Derivation and validation cohorts, compared with standard of care and applied 10-fold cross validation on the internal validation cohort |

The studies included in the review fell into two methodological categories. First, many studies used existing, retrospective data from cohort studies (n=9) or clinical trials (n=8) to identify important risk factors for an outcome and then used these to construct a scoring system. The published reports for these tools show no evidence of actual administration of the tool to a patient cohort and appear not to have been implemented at all. A second approach was to design a tool in advance—presumably based on existing data and/or clinical experience—and then report a study in which the tool was implemented in a trial or observational cohort setting (n=11). Most of the studies in both categories developed new tools, but others compared an existing or adapted tool to standard of care or to one or more other tools, such as comparing WHO's latest TB symptom screening algorithm to its previous version. The tools that were administered in implementation studies used digital and/or paper formats for data collection; most did not report the format utilized.

Performance metrics

Performance metrics for the risk triaging tools evaluated are detailed in Table 2. Sensitivity of the applied tools varied between studies but also within studies that considered multiple thresholds were considered. More than half (n=16; 57%) achieved sensitivities >80%. As would be expected for risk triaging tools (in contrast to diagnostic tools), results for specificity were generally low. Of the 20 articles that reported specificity, 8 (40%) had a specificity <50%. Not all studies reported measures for negative and positive predictive value; those that did indicated negative predictive values ranging from 4%-99% and positive predictive values ranging from 1%-99%. Area under the curve (AUC) was also not universally reported, but for studies where this metric was available (n=16), most AUC values were >0.7, indicating that the tools correctly identified the outcome of interest in 7 or more of 10 patients.

Table 2: Performance metrics of risk scoring tools

| Study | Sensitivity | Specificity | Negative predictive value (95% CI) | Positive predictive value (95% CI) | Area under the curve | Study's conclusion on performance |
|---------------|--|--|---|--|---|---|
| Alamo, 2012 | Urban ASO: WHO07=97% (95% CI: 80-100%) UgWHO03=94% (95% CI: 77-99%); Rural hospital: WHO07=96% (95% CI: 78-100%) UgWHO03=75% (95% CI: 43-93%) | Urban ASO: WHO07=77% (95% CI: 67-84%) UgWHO03=79% (95% CI: 70-86%); Rural hospital: WHO07=98% (95% CI: 93-100%) UgWHO03=97% (95% CI: 90-100%) | Urban ASO: WHO07=99% (95% CI: 92-100%) UgWHO03=98% (95% CI: 92-100%); Rural hospital: WHO07=99% (95% CI: 94-100%) UgWHO03=96% (95% CI: 89-99%) | Urban ASO: WHO07=54% (95% CI: 40-67%) UgWHO03=54% (95% CI: 40-68%); Rural hospital: WHO07=93% (95% CI: 74-99%) UgWHO03=82% (95% CI: 48-97%) | | The developed risk score successfully identified persons with persistently elevated VL or resistance and who need immediate ART regimen change. |
| Auld 2020 | Model A: 86%; Model B: 92% | Model A: 66%; Model B: 63% | | | Model A - Derivation:0.874; Validation:0.822. Model B - Derivation: 0.887; Validation: 0.836 | Both models had good model fit and excellent discrimination. Model A had adequate prediction performance. Model B overestimated mortality risk in the validation dataset in the highest risk group. |
| Auld 2021 | 88% | 55% | | | 0.800 (95% CI: 0.775 – 0.826) | Excellent discrimination in the development dataset and acceptable discrimination in the internal validation dataset. |
| Aunsborg 2020 | 95.5% | 36.9% | 97.6% (41/42) | 23.1% (22/118) | 0.77 | Tool contributed to increased TB screening and alertness about TB. |
| Awolude, 2021 | VIA: 50.0%; GeneXpert: 95.5% | VIA: 25.0%; GeneXpert: 75.0% | VIA: 55.0%; GeneXpert: 87.5% | VIA: 21.4%; GeneXpert: 90.0% | | Feasible and reduced overtreatment by 47.4%. |
| Baik 2020 | | | | | Derivation population- C-statistic: 0.82 (95% CI: 0.81-0.82); Validation population- C-statistic: 0.75 (95% CI: 0.69-0.80) | Tool had reasonable predictive accuracy and is transportable across SSA primary care settings. Positive net benefit when compared to a treat-all or treat-none strategy. |
| Balcha 2014 | | | For the low-risk group: 92% | | 0.74 | Tool can further classify PLHIV after a positive WHO-TB screen and would reduce the number of patients in need of further TB investigations before ART. |

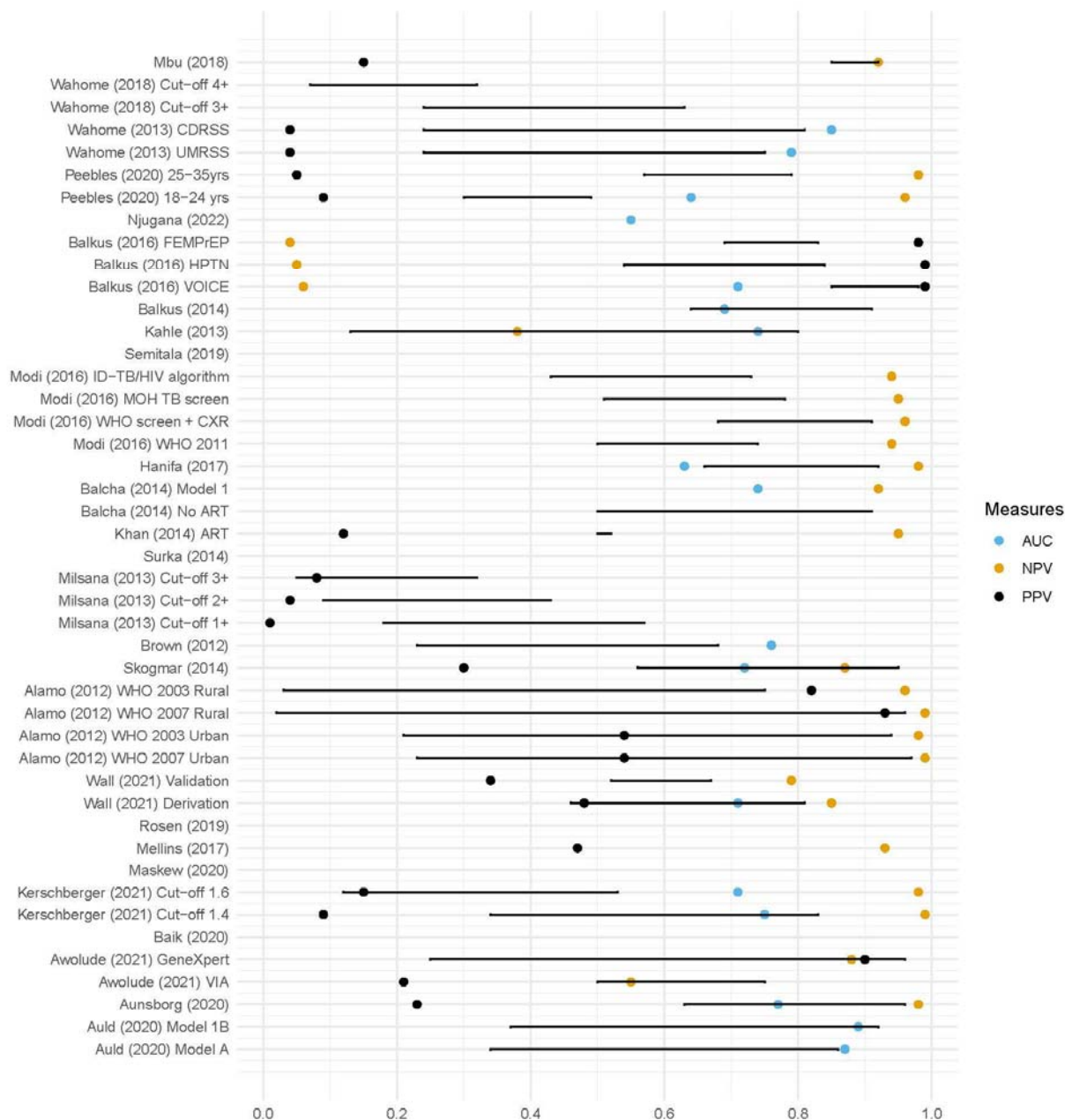
| Study | Sensitivity | Specificity | Negative predictive value (95% CI) | Positive predictive value (95% CI) | Area under the curve | Study's conclusion on performance |
|--------------------|---|--|--|---|--|--|
| Balkus 2016 | VOICE: 98%/91% HPTN: 84%/58% FEMPrEP: 83% | VOICE: 15%/38% HPTN: 46%/58% FEMPrEP: 31% | VOICE: 6% HPTN: 5% FEMPrEP: 4% | VOICE: 99% HPTN: 99% FEMPrEP: 98% | 0.71 (95% CI: 0.68 to 0.74) | Higher risk score associated with higher risk of HIV-1 seroconversion. Authors recommend use to provide targeted PrEP in populations with high burden. |
| Balkus 2018 | 91% | 36% | | | 0.69 (95% CI: 0.64-0.74) | Higher risk score associated with higher risk of HIV-1 seroconversion. Authors recommend use to provide targeted PrEP in populations with high burden. |
| Brown 2012 | 68% | 77% | | | 0.76 (95% CI: 0.67 - 0.84) | HIV partner notification is feasible in the setting. At a score cutoff of ≥ 2 , < 2/3 of resources can be used to yield more than 90% of partners tested under universal provider-assisted referral |
| Hanifa 2017 | Cutoff score ≥ 3 : 91.8% (95% CI: 85-96.2) | Cutoff score ≥ 3 : 34.3% (95% CI: 31.3-37.5) | Cutoff score ≥ 3 : 97.3% (95% CI: 94.9-98.7) | | Cutoff score ≥ 3 : 63.1% (95% CI: 60.1-66.1) | With a cut-off score ≥ 3 , 68% of symptomatic individuals would be tested, avoiding 32% of tests but missing 3% of TB cases. |
| Kahle, 2013 | In the external validation dataset and at cut-off point of ≥ 6 : the score predicted 80% of seroconversion from 37% of the population and 55% of seroconversions in another cohort | Cutoff ≥ 6 : 0.87 | Cutoff ≥ 6 : 0.38 | | Validation cohort1: 0.74 (95% CI: 0.70-0.78); Validation cohort2: 0.70 (95% CI: 0.64-0.76) | Tool was composed of well-established risk factors for HIV-1 and measurable in clinical settings. The tool also had good predictive ability in internal and external validation |
| Kerschberger, 2021 | Cut-off score ≥ 1.4 : 83.3%; Cut-off score ≥ 1.6 : 53.3% | Cut-off score ≥ 1.4 : 65.8%; Cut-off score ≥ 1.6 : 88.1% | Cut-off score ≥ 1.4 : 99.0% (95% CI: 97.7-99.7); Cut-off score ≥ 1.6 : 98.0% (95% CI: 96.6-98.9) | Cut-off score ≥ 1.4 : 8.7% (95% CI: 5.7-12.6); Cut-off score ≥ 1.6 : 15.0% (95% CI: 8.8-23.1) | Cut-off score ≥ 1.4 : 0.75 (95% CI: 0.68-0.82); Cut-off score ≥ 1.6 : 0.71 (95% CI: 0.62-0.80) | The tool showed potential to predict patients at risk of AEHI. Patients achieved favorable outcomes; tool performed better than PRS in other settings. |
| Khan, 2014 | ART: 51.6% (95%CI:33.1-69.9); No ART: 91.2% | 49.6% (95% CI: 45.7-53.4) | 94.9% (92.0-96.9) | 12.2% (95%CI: 9.1-15.6) | | Lower sensitivity and higher specificity among participants on ART compared to those not on ART. |

| Study | Sensitivity | Specificity | Negative predictive value (95% CI) | Positive predictive value (95% CI) | Area under the curve | Study's conclusion on performance |
|---------------|---|--|--|--|--|--|
| Maskew, 2020 | | | | | | The proportion of patients initiating ART was higher in the intervention arm at every time point. |
| Mbu, 2018 | WHO symptom screening: 92% (95% CI: 86–96%) | WHO symptom screening: 15% (95% CI: 12–17%) | WHO symptom screening: 92% (95% CI: 86-96%) | WHO symptom screening: 15% (95% CI: 14-17%) | | Sensitivity was high for people with symptoms. |
| Mellins, 2017 | 73% | 81% | 93% | 47% | | The validation study supports the validity of the isiZulu CDQ as a screening tool. |
| Mlisana, 2013 | Cutoff >=1: 57.1%; Cutoff >=2: 42.9%; Cutoff >=3: 32.1% | Cutoff >=1: 81.5%; Cutoff >=2: 91.4%; Cutoff >=3: 95.3% | | Cutoff >=1: 1.4%; Cutoff >=2: 3.9%; Cutoff >=3: 7.5% | | Tool can be used to enhance detection of acute HIV infection. It did not yield as high a predictive value or sensitivity as another algorithm developed elsewhere. |
| Modi, 2016 | WHO screening: 74.1% (95% CI: 64.1-82.2); WHO screening+chest radiograph: 90.9% (95% CI: 86.4-93.9); MOH TB screening algorithm: 77.5%(95% CI: 68.6-84.5); ID-TB/HIV algorithm: 72.5% (95% CI: 60.9-81.7) | WHO screening: 49.5% (95% CI: 45.1-53.9); WHO screening + chest radiograph: 32.0% (95% CI: 27.5-36.8); MoH TB screening algorithm: 49.4% (95% CI: 43.3-55.5); ID-TB/HIV algorithm: 56.5% (95% CI: 52.5-60.5) | WHO screening: 93.8% (95% CI: 91.4-95.6); WHO screening+chest radiograph: 96.1% (95% CI: 94.4-97.3); MoH TB screening algorithm: 94.7% (95% CI: 92.9-96.0); ID-TB/HIV algorithm: 94.3% (95% CI: 91.6-96.2) | | | Algorithms performed similarly but were variable across sub-sets of PLHIV such as severely immunosuppressed patients and pregnant women. |
| Njuguna, 2022 | 55% | | | | EMR: 0.55 (95% CI: 0.52 to 0.56) Partners: 0.56 (95% CI: 0.48 to 0.57). | Modest accuracy in predicting unsuppressed viral load. |
| Peebles, 2020 | 18-24: 49% 25-35: 79% | 18-24: 70% 25-35: 43% | 18-24: 96% 25-35: 98% | 18-24: 9% 25-35: 5% | 0.64 (95% CI: 0.60 to 0.67) | Authors recommend use to provide targeted PrEP in populations with high burden and stratify risk by age. |
| Rosen, 2019 | | | | | | Half and 55% of the screened patients in the intervention arm were eligible for same-day initiation in South Africa and Kenya, respectively. |

| Study | Sensitivity | Specificity | Negative predictive value (95% CI) | Positive predictive value (95% CI) | Area under the curve | Study's conclusion on performance |
|----------------|--|---|---|---|---|--|
| Semitala, 2019 | Among patients new to care, POC-based ICF detected 93% of all TB cases | | | | | Among patients new to care, POC CRP-based screening can improve ICF efficiency without compromising yield. |
| Skogmar, 2014 | 95% | 44% | 87% | 30% | 0.721 | Scoring system developed based on variables that can be collected by health professionals with limited training. |
| Surka, 2014 | | | | | | Mobile application of a non-blood based CVD used by CHWs was associated with a major reduction in training time, reaching adequate proficiency, screening for CVD risk and eliminating risk score calculation error. |
| Wahome, 2013 | UMRSS: 75.3% CDRSS: 80.8% | UMRSS: 76.4% CDRSS: 76% | | UMRSS: 3.5% CDRSS: 3.7% | UMRSS: 0.79 CDRSS: 0.85/ 0.77 if limited to those <30yrs | Risk score with cutoff point of 2 maximized sensitivity and specificity of predicting AEHI risk. |
| Wahome, 2018 | Cutoff>=3: 62.9% Cutoff>=4: 32% | Cutoff>=3: 76.0% Cutoff>=4: 92.8% | | | | Detected 31 out of 88 person-years for those with score ≥4. |
| Wall, 2021 | Derivation cohort: 81% External validation cohort: 67% | Derivation cohort: 54% External validation cohort: 48% | Derivation cohort: 85% External validation cohort: 79% | Derivation cohort: 48% External validation cohort: 34% | 0.71 (95% CI: 0.55-0.86) | Reasonable discrimination in derivation cohort. The validated risk algorithm outperformed the existing Rwandan National criteria (SoC). |

Figure 3 illustrates the summary performance of tools evaluated in the included studies. For each tool, we plotted a solid line connecting the point estimate for sensitivity with the estimate for false positive rate (1-specificity). Next NPV, PPV, and AUC (where available) were plotted as individual points around the plotted line. In this way, a tool with high performance metrics would be visualised as a wide solid line connecting point estimates for sensitivity and false positive rate with AUC and other metrics clustered near the right side of the axis. Several tools demonstrated these patterns[32,36]. There was a frequently observed trade-off, however, between sensitivity and specificity. Three TB screening algorithms assessed in Kenya, for example, demonstrated sensitivity between 73-91%, while specificity was <50% for all three [50]. The effect of outcome prevalence on PPV and NPV was also observed, such as for the VOICE risk scoring tool predicting risk of HIV acquisition[39].

Figure 3. Summary performance metrics of risk triaging tools*



*For each study, the solid line represents a plotted line connecting the point estimate for sensitivity with the point estimate for false positive rate (1-specificity) for that specific tool. Where available, negative predictive value (NPV), positive predictive value (PPV), and area under the curve (AUC) were plotted as individual points around the plotted line as indicated in the legend.

Health system effects

As noted previously, 14 of the 28 studies included in this review did not implement the risk triaging tools they developed. The remaining 14 implemented the tools in patient cohorts, either in clinical trial (n=4) or observational (n=10) settings. Few of the studies reported effects of tool implementation beyond the accuracy of the risk predictions made. We found no discussion of health system effects as defined in our outcomes in 4 of the implementation studies. Other studies posited (but did not provide evidence of) potential impacts, should their tool be implemented at scale. Comments on implementation fell into 3 categories: 1) the potential for safely shifting screening services to lower cadre staff, even if higher cadre staff are needed to provide treatment; 2) improving efficiency and reducing service delivery costs; and 3) the use of existing or easily collected indicators, without the need for laboratory tests or equipment that is not typically available in primary clinics. None was able to provide tangible evidence of potential impact of the tool in question.

Studies that reported potential improved efficiency and cost savings with risk triaging attributed such gains to fewer patients requiring further investigation after risk screening[33]; task shifting as less expensive cadres including lay health workers conduct the screening[42,48,56]; targeting post-risk screening care to patient groups with the highest benefit[38,54]; and reduced resources required for patient tracing due to improved yield[41]. Studies that both implemented and developed risk scoring tools further reported potential health system benefits through portability, ease of transferability[37], and scalability of risk scoring tools in resource limited settings when compared to advanced microbiological diagnostics[35]. They also noted improved staff alertness of the condition in question when a risk scoring tool is actively integrated in routine clinic processes[35]. In some cases, other

indirect benefits to implementation of triaging tools were observed, such as general improvements in awareness and uptake of TB screening reported after implementation of the TBscore tool[35].

Assessment of risk of bias

We evaluated each included article for risk of bias utilising the JBI checklist for diagnostic test accuracy which considers important potential areas of bias for the measurement and interpretation of test accuracy results [31]. This includes features of study design that impact test metrics such as avoiding a case-control approach (in which the people with the disease (cases) are selected from a different population than the control persons without the disease) in favour of cross-sectional or cohort designs in which all participants undergo the test or triaging process. The latter approach reflects reality better than the case–control design and is more likely to provide valid estimates of diagnostic accuracy [60]. Overall, the median bias score was 12% (IQR 0-25%). Nineteen studies were classified as low risk of bias and 9 with moderate risk of bias. No articles indicated high risk of bias. Details are shown in S4 Table.

DISCUSSION

In settings that face a scarcity of resources—a situation that prevails in many public sector healthcare settings globally—identifying individuals at higher risk of negative outcomes now or in the future is a critical step in ensuring that available resources are allocated where they can do most good. While risk triaging is undoubtedly conducted informally by most if not all healthcare providers, based on their own experience and judgment, widespread use of clinical triaging tools appears to be rare in primary healthcare clinic settings in sub-Saharan Africa. There is even less use of structured instruments to identify future behavioural risks, such as attrition from care. In resource-scarce settings, many facilities

are high-volume, staff-constrained sites, where effective triaging tools could improve the quality of clinical care while also increasing the efficiency of health care resource utilization.

The systematic review reported here is, to our knowledge, the first attempt to synthesize experience in developing, testing, and, in a few cases, implementing risk triaging tools in these settings. Our findings offer a mixed message, in terms of the future role of risk triaging in routine care. On one hand, most of the tools we identified were at least somewhat successful in identifying potential risk, with the majority of tools correctly identifying the outcome of interest in upwards of 70% of screened patients. The relatively low specificity of most of the tools is not a serious concern, since most were developed as screening tools rather than diagnostics, making high sensitivity the priority. For purposes of identifying groups of patients who are likely to benefit from additional testing or monitoring, versus those who clearly have no need for additional services, even a tool with modest sensitivity and low specificity has the potential to improve resource allocation, particularly in settings where prevalence of the outcome or condition screened for is high. In addition, an encouraging proportion of the tools we reviewed were explicitly designed to rely on indicators readily available at the primary healthcare level, rather than laboratory tests or imaging assays, though many also emphasized risk factors that are not easily addressed through practical interventions, such as age and sex.

On the other hand, based on the published reports, very few of the risk triaging tools identified were ever implemented beyond their initial studies or scaled up outside the research setting. Some developed and internally validated their risk scores using retrospective data and some tested risk triaging tools in small study populations, but none reported routine adoption of the prediction tools. There are likely multiple reasons for this, ranging from lack of logistical feasibility (e.g. an instrument required too much data to be feasible to administer during a routine clinical encounter) to lack of dissemination on the part

of the developers. Without a clearly demonstrated need and demand for a triaging approach, as well as motivated and empowered champions willing and able to advocate for scale-up, many potentially well-performing and useful tools will not be utilised beyond their development setting. The lack of published information about how improvements to routine procedures, such as a new risk triaging tool, can be adopted at large scale is a gap in our understanding of how to improve primary healthcare performance.

While few of the tools we reviewed appear to have been adopted for widespread use, there is one (and perhaps only one) that has: the WHO's TB symptom screening tool. The four- or five-item screening questionnaire is in widespread, if not near-universal, use throughout sub-Saharan Africa for people living with HIV. Several of the studies included here evaluated the TB symptom screen, alone or in combination with other indicators, and generated mixed results in terms of sensitivity and specificity for different populations. Exploration of how the TB symptom screen came to be so widely accepted and utilized suggests that a combination of meta-analysis of data from many sources, early and substantial engagement of policy makers, and international agency leadership allowed the TB symptom screen first to be incorporated into WHO guidelines[61] and then integrated into national guidelines and practice throughout the region. Better understanding of this process may be useful in introducing other risk triaging tools into widespread use.

Half of the studies included in our review addressed risks associated with HIV, whether transmission, treatment initiation, viral suppression, or mortality, and another third focused on TB case-finding. This may be explained by the very large numbers of individuals requiring these services in many sub-Saharan African countries. It may also reflect the large amounts of funding that have been available for HIV- and TB-related research and interventions in general, compared to other conditions that could be amenable to risk triaging, such as NCDs. Interestingly, however, none of the HIV-related studies included here

offered a risk triaging tool aimed directly at adherence to HIV treatment, with retention in care as the outcome of interest. Researchers may have had more success in assessing risks of HIV acquisition, readiness for treatment, viral suppression, and mortality than in finding consistent risk predictors of the broader outcome of “retention,” but it is retention that remains the single most difficult challenge for national HIV programs. Continued research in this area may thus be warranted.

The studies reviewed here do provide a number of general and specific lessons about the characteristics of a promising risk triaging tool and how such a tool may be adopted into widespread use. First, tools must be brief and easy to integrate into routine care without disrupting established practices and rely on risk factors that are either already routinely recorded or are easy and quick to collect. Expanding existing data collection requirements is simply not feasible in busy, understaffed clinics where computer access, and even electricity, cannot be taken for granted. In the authors’ experience, moreover, primary healthcare clinic staff may resist innovations that they believe will increase their workload or complicate their tasks, even if the ultimate effect may be the opposite, such as an intervention that increases the number of candidates for early screening but reduces long-term treatment demand. Convincing healthcare staff that a new tool is worth the effort is likely to be much more successful if the tool itself is simple to understand and implement. Tools that can be implemented, wholly or in part, by lay cadres such as counsellors, peer supporters or community healthcare workers, rather than solely by scarcer cadres such as nurses, may also be more promising for widespread adoption, particularly as provision of primary care move towards community-based and other out-of-facility service delivery models.

Second, where risk scores are generated and utilised, interpretation of the output scores should be simple and straightforward. Decision cut-off points (e.g. having a risk score above or below a threshold) must be clear and well justified. We also note that while none of the studies we reviewed discussed

offering healthcare providers discretion in interpreting risk scores, it seems reasonable that these scores should be seen as guidelines rather than absolute rules. If a clinician interacting with a patient believes that the risk score generated does not lead to the correct management pathway, clinician judgment may be preferred, as it is in many clinical situations. Similarly, individual patients may have strong reasons for preferring one pathway to another. Since patient cooperation is essential to achieving good outcomes, such preferences should be taken into account, no matter what a risk triaging tool suggests.

Third, in most cases, assessing risk is not an end in itself. It does little good, for example, to designate an HIV treatment patient as having a low or high risk of treatment interruption if there is not a clear pathway for responding to the conditions that create the assessed level of risk. A client who is determined to be at high risk of disengagement due to fears of disclosure and stigma will require an entirely different intervention strategy from a client at high risk of disengagement due to an employment situation that does not allow clinic visits during working hours. Risk triaging tools should thus come with intervention strategies or programs designed for the specific risk factors that contribute to the score or other outcome.

Fourth, performance metrics should not be the sole decision-making criterion determining implementation of a potential tool. A tool with low specificity may be acceptable if the purpose of risk triaging is to reduce the population of patients who may require additional diagnosis, care, or support, because even a relatively non-specific tool has the potential to identify those patients who do not require additional services. In any group of 100 clients initiating ART, for example, it is likely that at least 60, if not more, are at very low risk of disengagement from care in their first year [14,62]. Standard of care guidelines in many settings would require all one hundred of them to be assigned multiple adherence counselling sessions in the months after initiation. A triaging tool that could identify even half (30) of the true “low risk” patients would allow the facility to focus counselling resources more

efficiently among those who are most likely to derive benefit from such an intervention. While not ideal, this would still represent a potentially valuable improvement in facility performance.

Finally, no risk triaging tool should be expected to perform consistently across all populations and settings. Setting matters when evaluating tool performance, even for a screening tool for TB risk, which relies on presumably objective symptom and laboratory indicators[63] and will be affected by TB prevalence rates in the setting the tool is implemented. For a behavioral risk tool, community and facility factors may be as important in predicting risk as individual characteristics. Travelling long distances to an ART clinic, for example, is frequently identified as a barrier to treatment adherence, and many programs offer community-based service delivery to address this barrier. In communities where stigma around HIV is high, however, individuals requiring HIV care may in fact prefer to travel further to facilities based in communities where they are less likely to be recognized by those to whom they have not disclosed their status. Local refinement of established risk triaging instruments may thus be desirable.

This review had a number of limitations. First, our search terms excluded all inpatient settings. This may have systematically excluded entire medical conditions that frequently use risk triaging approaches, such as disease due to the Ebola virus [64]. Second, while we included terms for performance test metrics such as sensitivity, specificity, and accuracy to increase the yield of studies reporting methodological results, we acknowledge that the terminology around test metrics is not universally standard and may have led to relevant studies being missed, particularly among older studies. Third, though we were unable to address this limitation through search methodology, we acknowledge that a systematic review will not observe tools that have not been described and published in academic literature. There may be some inherent reporting bias if results of tools that did not work were less likely

to be published, but it is also possible that implementing partners and organizations have designed and implemented successful triaging tools and systems in primary health contexts but not evaluated or published these. Fourth, as discussed above, generalizability of the performance of any given risk triaging tool is uncertain. Tools that rely on clinical symptoms or test results may be applicable across multiple settings and populations, but tools that assess behavioural risks seem likely to be highly context-specific. Finally, in determining which papers to include in the analysis, the distinction between risk triaging and diagnosis was sometimes unclear and required author discretion; it is possible that some sources that others may regard as examples of risk triaging were categorized as diagnosis and excluded.

Conclusion

Despite the limitations listed above, our review highlights some important considerations for the implementation of risk triaging tools in the context of HIV retention. Many studies have achieved relatively impressive performance in identifying the risks they target within the data sets or small patient cohorts in which they were designed and validated. The lack of large-scale adoption and implementation of any of these instruments, however, points to serious gaps in previous efforts. Bringing implementers, policy makers, funders, and potentially clients themselves into the process of identifying needs, developing tools and intervention strategies, and assessing feasibility may be an essential component of improving health outcomes through risk stratification.

Ethics approval and consent to participate

This is a systematic review of published information and no direct contact with participants or access to participant data beyond what is reported in the published studies occurred.

Availability of data and materials

All data used in this paper are included in the cited manuscripts and publicly available in the cited sources.

Competing interests

The authors declare that they have no competing interests.

Funding

Funding for the study was provided the Bill & Melinda Gates Foundation through INV-031690 to Boston University. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

MM, SR, and LS conceived of and designed the study. DF defined and ran the search. MM, LS, and MB identified and reviewed sources and extracted data. MM, LS, and MB analyzed the data and drafted the manuscript. All authors reviewed and edited the manuscript.

Acknowledgements

None

REFERENCES

1. Mirzazadeh A, Eshun-Wilson I, Thompson RR, Bonyani A, Kahn JG, Baral SD, et al. Interventions to reengage people living with HIV who are lost to follow-up from HIV treatment programs: A systematic review and meta-analysis. *PLoS Med.* 2022;19: e1003940. doi:10.1371/journal.pmed.1003940
2. Muhula S, Gachohi J, Kombe Y, Karanja S. Interventions to improve early retention of patients in antiretroviral therapy programmes in sub-Saharan Africa: A systematic review. *PLoS One.* 2022;17: e0263663. doi:10.1371/journal.pone.0263663
3. Hariprasad S, et al. Strategies to Improve Antiretroviral Therapy (ART) Initiation and Early Retention Among Men in sub-Saharan Africa: A Systematic Review. *AIDS* 2022. 2022;(submission: 1–2.
4. Damulak PP, Ismail S, Manaf RA, Said SM, Agbaji O. Interventions to Improve Adherence to Antiretroviral Therapy (ART) in Sub-Saharan Africa: An Updated Systematic Review. *Int J Environ Res Public Health.* 2021;18: 1–18. doi:10.3390/IJERPH18052477
5. MacKellar D, Hlophe T, Ujamaa D, Pals S, Dlamini M, Dube L, et al. Antiretroviral therapy initiation and retention among clients who received peer-delivered linkage case management and standard linkage services, Eswatini, 2016–2020: retrospective comparative cohort study. *ARCHIVES OF PUBLIC HEALTH.* 2022;80. doi:10.1186/s13690-022-00810-9 WE - Science Citation Index Expanded (SCI-EXPANDED) WE - Social Science Citation Index (SSCI)

6. Hlongwa M, Cornell M, Malone S, Pitsillides P, Little K, Hasen N. Uptake and Short-Term Retention in HIV Treatment Among Men in South Africa: The Coach Mpilo Pilot Project. *Glob Health Sci Pract*. 2022;10: 1–7. doi:10.9745/GHSP-D-21-00498
7. Lifson AR, Hailemichael A, Workneh S, MacLehose RF, Horvath KJ, Hilk R, et al. A three-year randomized community trial of community support workers in rural Ethiopia to promote retention in HIV care. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*. 2022. doi:10.1080/09540121.2022.2029819
8. Shushtari ZJ, Salimi Y, Sajjadi H, Paykani T. Effect of Social Support Interventions on Adherence to Antiretroviral Therapy Among People Living with HIV: A Systematic Review and Meta-Analysis. *AIDS and Behavior*. Springer; 2022. doi:10.1007/s10461-022-03894-0
9. Amstutz A, Lejone TI, Khesa L, Kopo M, Kao M, Muhairwe J, et al. Offering ART refill through community health workers versus clinic-based follow-up after home-based same-day ART initiation in rural Lesotho: The VIBRA cluster-randomized clinical trial. *PLoS Med*. 2021;18. doi:10.1371/JOURNAL.PMED.1003839
10. Fahey CA, Njau PF, Katabaro E, Mfaume RS, Ulenga N, Mwenda N, et al. Financial incentives to promote retention in care and viral suppression in adults with HIV initiating antiretroviral therapy in Tanzania: a three-arm randomised controlled trial. *Lancet HIV*. 2020;7: e762–e771. doi:10.1016/S2352-3018(20)30230-7
11. Backus BE, Six AJ, Kelder JC, Bosschaert MAR, Mast EG, Mosterd A, et al. A prospective validation of the HEART score for chest pain patients at the emergency department. *Int J Cardiol*. 2013;168: 2153–2158. doi:10.1016/j.ijcard.2013.01.255
12. Kavousi M, Leening MJG, Nanchen D, Greenland P, Graham IM, Steyerberg EW, et al. Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European cohort. *JAMA*. 2014;311: 1416–1423. doi:10.1001/jama.2014.2632
13. Aujesky D, Auble TE, Yealy DM, Stone RA, Obrosky DS, Meehan TP, et al. Prospective comparison of three validated prediction rules for prognosis in community-acquired pneumonia. *American Journal of Medicine*. 2005;118: 384–392. doi:10.1016/j.amjmed.2005.01.006
14. Inzaule SC, Kroeze S, Kityo CM, Siwale M, Akanmu S, Wellington M, et al. Long-term HIV treatment outcomes and associated factors in sub-Saharan Africa: multicountry longitudinal cohort analysis. *AIDS*. 2022;36: 1437–1447. doi:10.1097/QAD.0000000000003270
15. Wekesa P, McLigeyo A, Owuor K, Mwangi J, Ngugi E. Survival probability and factors associated with time to loss to follow-up and mortality among patients on antiretroviral treatment in central Kenya. *BMC Infect Dis*. 2022;22: 1–10. doi:10.1186/s12879-022-07505-0
16. Maskew M, Sharpey-Schafer K, De Voux L, Bor J, Rennick M, Crompton T, et al. Machine learning to predict retention and viral suppression in South African HIV treatment cohorts. *medRxiv*. 2021; 2021.02.03.21251100.

17. Siika A, McCabe L, Bwakura-Dangarembizi M, Kityo C, Mallewa J, Berkley J, et al. Late Presentation with HIV in Africa: Phenotypes, Risk, and Risk Stratification in the REALITY Trial. *Clinical Infectious Diseases*. 2018;66: S140–S146. doi:10.1093/cid/cix1142
18. Mbengue MAS, Chasela C, Onoya D, Mboup S, Fox MP, Evans D. Clinical predictor score to identify patients at risk of poor viral load suppression at six months on antiretroviral therapy: Results from a prospective cohort study in Johannesburg, south Africa. *Clin Epidemiol*. 2019;11: 359–373. doi:10.2147/CLEP.S197741
19. Brathwaite R, Ssewamala FM, Neilands TB, Okumu M, Mutumba M, Damulira C, et al. Predicting the individualized risk of poor adherence to ART medication among adolescents living with HIV in Uganda: the Suubi+Adherence study. *J Int AIDS Soc*. 2021;24. doi:10.1002/jia2.25756
20. Rosen S, Maskew M, Larson BA, Brennan AT, Tsikhutsu I, Fox MP, et al. Simplified clinical algorithm for identifying patients eligible for same-day HIV treatment initiation (SLATE): results from an individually randomized trial in South Africa and Kenya Title. *PLoS Med*. 2019;16: e1002912.
21. Maskew M, Brennan AT, Fox MP, Vezi L, Venter WDF, Ehrenkranz P, et al. A clinical algorithm for same-day HIV treatment initiation in settings with high TB symptom prevalence in South Africa: The SLATE II individually randomized clinical trial. *PLoS Med*. 2020;17. doi:10.1371/JOURNAL.PMED.1003226
22. Moucheraud C, Hoffman RM, Balakasi K, Wong V, Sanena M, Gupta S, et al. Screening Adults for HIV Testing in the Outpatient Department: An Assessment of Tool Performance in Malawi. *AIDS Behav*. 2013;26: 478–486. doi:10.1007/s10461-021-03404-8
23. Moucheraud C, Chasweka D, Nyirenda M, Schooley A, Dovel K, Hoffman RM. Simple screening tool to help identify high-risk children for targeted HIV testing in malawian inpatient wards. *J Acquir Immune Defic Syndr (1988)*. 2018;79: 352–357. doi:10.1097/QAI.0000000000001804
24. Beres LK, Schwartz S, Simbeza S, McGready J, Eshun-Wilson I, Mwamba C, et al. Patterns and Predictors of Incident Return to HIV Care Among Traced, Disengaged Patients in Zambia. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2020;Publish Ah: 313–322. doi:10.1097/qai.0000000000002554
25. Auld AF, Kerkhoff AD, Hanifa Y, Wood R, Charalambous S, Liu Y, et al. Derivation and external validation of a risk score for predicting HIV-associated tuberculosis to support case finding and preventive therapy scale-up: A cohort study. *PLoS Med*. 2021;18. doi:10.1371/JOURNAL.PMED.1003739
26. Hanifa Y, Fielding KL, Chihota VN, Adonis L, Charalambous S, Foster N, et al. A clinical scoring system to prioritise investigation for tuberculosis among adults attending HIV clinics in South Africa. *PLoS One*. 2017;12. doi:10.1371/JOURNAL.PONE.0181519
27. Aunsborg J, Hønge B, Jespersen S, ... FR-IJ of, 2020 undefined. A clinical score has utility in tuberculosis case-finding among patients with HIV: A feasibility study from Bissau. Elsevier.

28. Esra R, Carstens J, Le Roux S, Mabuto T, Eisenstein M, Keiser O, et al. Validation and improvement of a machine learning model to predict interruptions in antiretroviral treatment in South Africa. *J Acquir Immune Defic Syndr*. 2022. doi:10.1097/QAI.0000000000003108
29. Maskew M, Sharpey-Schafer K, De Voux L, Bor J, Rennick M, Crompton T, et al. Machine learning to predict retention and viral suppression in South African HIV treatment cohorts. medRxiv. 2021; 2021.02.03.21251100. Available: <http://medrxiv.org/content/early/2021/02/05/2021.02.03.21251100.abstract>
30. Cochrane Collaboration: Cochrane Screening and Diagnostic Test Methods Group. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 2. 2022.
31. Joanna Briggs Institute. Checklist for Diagnostic Test Accuracy Studies. *JbiGlobal*. 2017; 1–6.
32. Alamo S-T, Kunutsor S, Walley J, Thoulass J, Evans M, Muchuro S, et al. Performance of the new WHO diagnostic algorithm for smear-negative pulmonary tuberculosis in HIV prevalent settings: a multisite study in Uganda. *Wiley Online Library*. 2012;17: 884–895. doi:10.1111/j.1365-3156.2012.03003.x
33. Auld AF, Fielding K, Agizew T, Maida A, Mathoma A, Boyd R, et al. Risk scores for predicting early antiretroviral therapy mortality in sub-Saharan Africa to inform who needs intensification of care: a derivation and external validation cohort study. *BMC Med*. 2020;18: 1–19. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=146892922&site=ehost-live&scope=site>
34. Auld AF, Kerkhoff AD, Hanifa Y, Wood R, Charalambous S, Liu Y, et al. Derivation and external validation of a risk score for predicting HIV-associated tuberculosis to support case finding and preventive therapy scale-up: A cohort study. *PLoS Med*. 2021;18. doi:10.1371/journal.pmed.1003739
35. Aunsborg J, Hønge B, Jespersen S, ... FR-IJ of, 2020 undefined. A clinical score has utility in tuberculosis case-finding among patients with HIV: A feasibility study from Bissau. *International Journal of Infectious Diseases*. 2020.
36. Awolude OA, Oyerinde SO, Ayeni AO, Adewole IF. Human papillomavirus-based cervical precancer screening with visual inspection with acetic acid triage to achieve same-day treatments among women living with human immunodeficiency virus infection: test-of-concept study in Ibadan, Nigeria. *Pan Afr Med J*. 2021;40: 48.
37. Baik Y, Rickman HM, Hanrahan CF, Mmolawa L, Kitonsa PJ, Sewelana T, et al. A clinical score for identifying active tuberculosis while awaiting microbiological results: Development and validation of a multivariable prediction model in sub-Saharan Africa. *PLoS Med*. 2020;17: e1003420. doi:10.1371/journal.pmed.1003420
38. Balcha TT, Skogmar S, Sturegård E, Schön T, Winqvist N, Reepalu A, et al. A Clinical Scoring Algorithm for Determination of the Risk of Tuberculosis in HIV-Infected Adults: A Cohort Study Performed at Ethiopian Health Centers. *Open Forum Infect Dis*. 2014;1. doi:10.1093/OFID/OFU095

39. Balkus JE, Brown E, Palanee T, Nair G, Gafoor Z, Zhang J, et al. An Empiric HIV Risk Scoring Tool to Predict HIV-1 Acquisition in African Women. *J Acquir Immune Defic Syndr*. 2016;72: 333. doi:10.1097/QAI.0000000000000974
40. Balkus JE, Brown ER, Palanee-Phillips T, Matovu Kiweewa F, Mgodhi N, Naidoo L, et al. Performance of a validated risk score to predict HIV-1 acquisition among African women participating in a trial of the dapivirine vaginal ring. *J Acquir Immune Defic Syndr*. 2018;77: e8. doi:10.1097/QAI.0000000000001556
41. Brown LB, Miller WC, Kamanga G, Kaufman JS, Pettifor A, Dominik RC, et al. Predicting partner HIV testing and counseling following a partner notification intervention. *AIDS Behav*. 2012;16: 1148–55. doi:10.1007/s10461-011-0094-9
42. Hanifa Y, Fielding KL, Chihota VN, Adonis L, Charalambous S, Foster N, et al. A clinical scoring system to prioritise investigation for tuberculosis among adults attending HIV clinics in South Africa. *PLoS One*. 2017;12. doi:10.1371/JOURNAL.PONE.0181519
43. Kahle EM, Hughes JP, Lingappa JR, John-Stewart G, Celum C, Nakku-Joloba E, et al. An empiric risk scoring tool for identifying high-risk heterosexual HIV-1 serodiscordant couples for targeted HIV-1 prevention. *J Acquir Immune Defic Syndr*. 2013;62: 339. doi:10.1097/QAI.0B013E31827E622D
44. Kerschberger B, Aung A, Mpala Q, Ntshalintshali N, Mamba C, Schomaker M, et al. Predicting, Diagnosing, and Treating Acute and Early HIV Infection in a Public Sector Facility in Eswatini. *J Acquir Immune Defic Syndr*. 2021;88: 506–517.
45. Khan F, Verkuil S, Parrish A, ... FC-A (London, 2014 undefined. Performance of symptom-based tuberculosis screening among people living with HIV: not as great as hoped. *ncbi.nlm.nih.gov*. 2014 [cited 31 Oct 2022]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5116236/>
46. Maskew M, Brennan AT, Fox MP, Vezi L, Venter WDF, Ehrenkranz P, et al. A clinical algorithm for same-day HIV treatment initiation in settings with high TB symptom prevalence in South Africa: The SLATE II individually randomized clinical trial. *PLoS Med*. 2020;17. doi:10.1371/JOURNAL.PMED.1003226
47. Mbu ET, Sauter F, Zoufaly A, Bronsvort BM de C, Morgan KL, Noeske J, et al. Tuberculosis in people newly diagnosed with HIV at a large HIV care and treatment center in Northwest Cameroon: Burden, comparative screening and diagnostic yields, and patient outcomes. *PLoS One*. 2018;13: e0199634. doi:10.1371/journal.pone.0199634
48. Mellins CA, Kauchali S, Nestadt DF, Bai D, Aidala A, Myeza N, et al. Validation of the Client Diagnostic Questionnaire to Assess Mental Health in South African Caregivers of Children. *Clin Psychol Psychother*. 2017;24: 245–254. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=121062751&site=ehost-live&scope=site>
49. Mlisana K, Sobieszczyk M, Werner L, Feinstein A, van Loggerenberg F, Naicker N, et al. Challenges of Diagnosing Acute HIV-1 Subtype C Infection in African Women: Performance of a Clinical Algorithm and the Need for Point-of-Care Nucleic-Acid Based Testing. *PLoS One*. 2013;8. doi:10.1371/JOURNAL.PONE.0062928

50. Modi S, Cavanaugh JS, Shiraishi RW, Alexander HL, McCarthy KD, Burmen B, et al. Performance of clinical screening algorithms for tuberculosis intensified case finding among people living with HIV in Western Kenya. *PLoS One*. 2016;11. doi:10.1371/JOURNAL.PONE.0167685
51. Njuguna N, Mugo N, Anzala O, Mureithi M, Irungu E, Wamicwe J, et al. An empiric tool to identify Kenyans living with HIV who will have unsuppressed viremia 18 months following treatment initiation to guide differentiated care models. *PLoS One*. 2022;17: e0271520. doi:10.1371/JOURNAL.PONE.0271520
52. Peebles K, Palanee-Phillips T, Balkus JE, Beesham I, Makkan H, Deese J, et al. Age-Specific Risk Scores Do Not Improve HIV-1 Prediction Among Women in South Africa. *J Acquir Immune Defic Syndr*. 2020;85: 156–164. doi:10.1097/QAI.0000000000002436
53. Rosen S, Maskew M, Larson BA, Brennan AT, Tsikhutsu I, Fox MP, et al. Simplified clinical algorithm for identifying patients eligible for same-day HIV treatment initiation (SLATE): Results from an individually randomized trial in South Africa and Kenya. Newell M-L, editor. *PLoS Med*. 2019;16: e1002912. doi:10.1371/journal.pmed.1002912
54. Semitala FC, Cattamanchi A, Andama A, Atuhumuza E, Katende J, Mwebe S, et al. Brief Report: Yield and Efficiency of Intensified Tuberculosis Case-Finding Algorithms in 2 High-Risk HIV Subgroups in Uganda. *J Acquir Immune Defic Syndr*. 2019;82: 416–420. doi:10.1097/QAI.0000000000002162
55. Skogmar S, Balcha TT, Jemal ZH, Björk J, Deressa W, Schön T, et al. Development of a clinical scoring system for assessment of immunosuppression in patients with tuberculosis and HIV infection without access to CD4 cell testing. *Taylor & Francis*. 2014;7: 1. doi:10.3402/gha.v7.23105
56. Surka S, Edirippulige S, Steyn K, Gaziano T, Puoane T, Levitt N. Evaluating the use of mobile phone technology to enhance cardiovascular disease screening by community health workers. *Int J Med Inform*. 2014;83: 648–654. doi:10.1016/J.IJMEDINF.2014.06.008
57. Wahome E, Fegan G, Okuku HS, Mugo P, Price MA, Mwashigadi G, et al. Evaluation of an empiric risk screening score to identify acute and early HIV-1 infection among MSM in Coastal Kenya. *AIDS*. 2013;27: 2163. doi:10.1097/QAD.0B013E3283629095
58. Wahome E, Thiong'o AN, Mwashigadi G, Chirro O, Mohamed K, Gichuru E, et al. An Empiric Risk Score to Guide PrEP Targeting Among MSM in Coastal Kenya. *AIDS Behav*. 2018;22: 35–44. doi:10.1007/S10461-018-2141-2/FIGURES/1
59. Wall KM, Nyombayire J, Parker R, Ingabire R, Bizimana J, Mukamuyango J, et al. Developing and validating a risk algorithm to diagnose *Neisseria gonorrhoeae* and *Chlamydia trachomatis* in symptomatic Rwandan women. *BMC Infect Dis*. 2021;21: 1–11. Available: <https://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=150043388&site=ehost-live&scope=site>
60. Leeflang MMG. Systematic reviews and meta-analyses of diagnostic test accuracy. *Clinical Microbiology and Infection*. 2014;20: 105–113. doi:10.1111/1469-0691.12474

61. Getahun H, Kittikraisak W, Heilig CM, Corbett EL, Ayles H, Cain KP, et al. Development of a standardized screening rule for tuberculosis in people living with HIV in resource-constrained settings: Individual participant data meta-analysis of observational studies. *PLoS Med.* 2011;8. doi:10.1371/JOURNAL.PMED.1000391
62. Tiendrebeogo T, Messou E, Arikawa S, Ekouevi DK, Tanon A, Kwaghe V, et al. Ten-year attrition and antiretroviral therapy response among HIV-positive adults: a sex-based cohort analysis from eight West African countries. *J Int AIDS Soc.* 2021;24: 1–10. doi:10.1002/jia2.25723
63. van't Hoog AH, Onozaki I, Lonroth K. Choosing algorithms for TB screening: A modelling study to compare yield, predictive value and diagnostic burden. *BMC Infect Dis.* 2014;14: 1–12. doi:10.1186/1471-2334-14-532/FIGURES/5
64. Oza S, Sesay AA, Russell NJ, Wing K, Boufkhed S, Vandi L, et al. Symptom- and Laboratory-Based Ebola Risk Scores to Differentiate Likely Ebola Infections. *Emerg Infect Dis.* 2017;23: 1792. doi:10.3201/EID2311.170171

Supporting information

S1 Text. Review protocol

S1 Table. Search strategy

S2 Table. Search terms-PubMed

S3 Table. Prisma checklist.

S4 Table. Summary of bias using the Joanna Briggs Checklist for Diagnostic Test Accuracy