

The Clean Paper

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

Clinical AI kan tó dà bí safe, tó sì improve paperwork — ṣùgbọ́n kò improve patient outcomes

Pragmatic cluster-randomized trial nínú 16 Kenyan primary care facilities test generative-AI decision-support tool tí a fi kún record clinical officers. Nínú 9,691 patients, strict 14-day expert-adjudicated composite of treatment failure jẹ́ 2.2% pẹ̀lú tool versus 2.0% láìsí rẹ̀ (adjusted odds ratio 0.77, 95% CI 0.55-1.08, P = 0.13) — no significant difference. Tool kò fi safety signal hàn, ó improve documentation kọjá all rated domains, prescribing àti satisfaction unchanged, antibiotic costs sì trend lower. Èyí kì í ṣe failed study; ó jẹ́ rare, honest one — measured on patients, not benchmarks — tó fi hàn pé tool tí looked safe, tó help process, kò demonstrably help patients nínú two weeks, benefit kankan sì yóò nilo much larger trial láti detect.

Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

Paper: *Generative AI-enabled clinical decision support system in primary care: a pragmatic, cluster-randomized trial*, Ambrose Agweyu, Paul Mwaniki, Vaishnavi Menon, Robert Korom, Lynda Isaaka, Conrad Wanyama, Xiaoxuan Liu & Bilal A. Mateen (and colleagues), Nature Medicine (2026) · DOI: 10.1038/s41591-026-04503-6

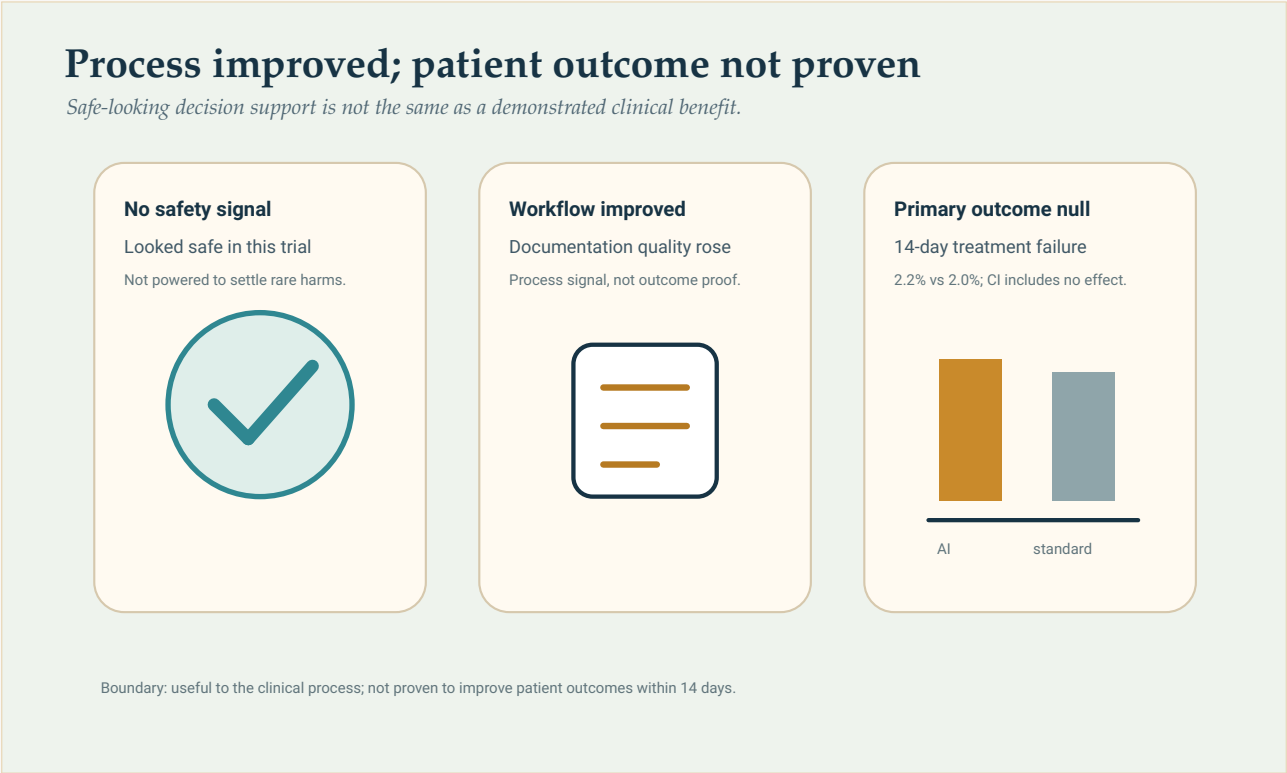
Safe àti helpful kì í dọ́gba pẹ̀lú effective

Ọ̀pọ̀ àkọlé nípa ìṣègùn AI ni a kọ láti wrong kind of ìwádìí. àwòṣe kan score dáadáa lórí exam ìbéèrè, tàbí ṣégun doctors lórí curated vignettes, ìtàn sì kọ ara rẹ̀: machine ti ready fún clinic.

ìdánwò yìí ṣe ohun tó nira, tó sì ṣọ̀wọ́n. Ó fi generative-AI decision-support irinṣẹ́ sínú gidi primary care, nínú gidi facilities, pẹ̀lú gidi aláìsàn, ó sì béèrè ìbéèrè tó matter gidi: ṣé aláìsàn ṣe dáadáa sí i?

Honest answer ni no — kò measurable, kì í ṣe nínú 14 days. irinṣẹ́ náà kò fi àábò àmì hàn. Ó mú sunwọ̀n sí i quality ìṣègùn documentation. Ó lè ti lower díẹ̀ drug costs. Ṣùgbọ́n kò significantly dín kù ìtọju ìkùnà, àwọn olùkọ̀wé sì ṣọ́ra láti sọ pé àǹfàání kankan, tí ó bá wà, probably modest.

Èyí kì í ṣe ìkùnà of ìwádìí. Èyí ni ìwádìí tí ń ṣiṣẹ́. Irú ẹ̀rí responsible nípa AI nínú medicine ni èyí, nígbà tí a bá measure lórí aláìsàn dípò benchmarks.



irinṣẹ́ náà kò fi àábò àmì hàn, ó sì mú sunwọ̀n sí i ìṣègùn documentation, ṣùgbọ́n prespecified 14-day aláìsàn èsì kò significantly mú sunwọ̀n sí i. Process help kì í dọ́gba pẹ̀lú proven aláìsàn àǹfàání.

The Clean Paper CC BY 4.0

Ohun tí àwọn olùkọ̀wé ṣe

ẹgbẹ́ náà run pragmatic, cluster-randomized ìdánwò kọjá **16 primary care facilities** tí private health network (Penda Health) ń operate ní Nairobi àti Kiambu counties, Kenya. Care nínú facilities wọ̀nyí jẹ́ mostly láti ọwọ́

ìṣègùn officers — mid-level practitioners pẹ̀lú three-year diploma — tí wọn kò sábà ní easy access sí senior consultation.

Unit of randomization ni clinician, kì í ṣe aláìsàn. **103 ìṣègùn officers** ni wọ́n randomize: 52 sí intervention arm àti 51 sí control arm. Arms méjèèjì lo kan náà cloud-based electronic ìṣègùn àkọsílẹ̀. Intervention arm ní afikun **"AI Consult"** (ẹ̀dà 2.0), decision-support irinṣẹ́ tí a kọ lórí **OpenAI's GPT-4o** large èdè àwòṣe, tí a embed sínú àkọsílẹ̀ náà. Ó ka information tí clinician document, ó sì lè flag ṣeé ṣe issues pẹ̀lú ìdánimọ̀ àrùn tàbí ìtọju plan. Clinicians pa full autonomy mọ́: wọ́n lè accept, modify tàbí ignore suggestions rẹ̀.

> **Model wo ni, kí ló sì dé tí specifics matter?**
>
> irinṣẹ́ náà jẹ́ **AI Consult 2.0**, tó run **OpenAI's GPT-4o** (May 2025 release), accessed nipasẹ̀ OpenAI commercial API labẹ enterprise licence, pẹ̀lú low-randomness settings (temperature 0.1). Ó wà nínú bespoke electronic àkọsílẹ̀ (EasyClinic's EMR), ètò prompts sì steer rẹ̀ láti align pẹ̀lú Kenyan national ìtọju guidelines; àwọn olùkọ̀wé publish full instruction prompt.
>
> Kí ló dé tá a fi sọ gbogbo èyí? Nítorí èsì jẹ́ nípa *specific ètò kan* — àwòṣe ẹ̀dà kan, prompt kan, àkọsílẹ̀ kan, setting kan — kì í ṣe "LLMs in medicine" ní gbogbogbo. Àwọn olùkọ̀wé ṣe kan náà kókó: wọ́n pe àwárí wọn ní *temporal benchmark rather than fixed ìṣírò àfojúsùn of capability*. Newer àwòṣe, yàtọ̀ prompt, tàbí clinic tí digitization rẹ̀ kéré lè gbogbo wọn yí èsì padà.
>
> Nípa independence: OpenAI nigbamii fún in-kind ṣe àtìlẹ́yìn fún (cloud-compute credits àti technical guidance lórí API), ṣùgbọ́n àwọn olùkọ̀wé sọ pé decision láti lo OpenAI ṣẹlẹ̀ *ṣáájú* offer yẹn, àti pé OpenAI kò ní role nínú ìdánwò àpẹẹrẹ, dátà collection, ìtúpalẹ̀ tàbí decision to publish.

Láàárín 22 April àti 16 July 2025, **9,691 aláìsàn** ni wọ́n enroll. **Primary èsì jẹ́ deliberately patient-centered àti strict**: expert-adjudicated *composite of ìtọju ìkùnà* nínú 14 days lẹ́yìn visit — panel clinicians, blind sí ìwádìí arm, judge bóyá aláìsàn kọọkan ní bad èsì bí unresolved tàbí worsening illness. ìdánwò náà ni a register ṣáájú (Pan-African Clinical Trials Registry 202502499779176).

Design choice yìí ni kókó. Ó rọrùn láti fi hàn pé AI irinṣẹ́ yí ohun tí clinician kọ sílẹ̀ padà. Ó nira púpọ̀, ó sì meaningful sí i, láti fi hàn pé ó yí ohun tí ó ṣẹlẹ̀ sí aláìsàn padà.

Ohun tí wọ́n rí

Primary èsì kò mú sunwọn sí i. ìtọju ìkùnà ṣẹlẹ̀ nínú 102 of 4,693 aláìsàn (2.2%) nínú AI arm àti 94 of 4,654 (2.0%) nínú control arm. Crude percentages díẹ̀ ga pẹ̀lú AI, ṣùgbọ́n lẹ́yìn adjustment kókó ìṣírò àfojúsùn lean toward àǹfààní: **adjusted odds ratio 0.77 (95% ìgbẹ́kẹ̀lé interval 0.55 to 1.08, P = 0.13)** — not statistically significant. Flip láàárín raw àti adjusted numbers kì í ṣe arithmetic slip; adjustment accounts fún differences láàárín clinician clusters. Ní ọ̀nà méjèèjì, ìgbẹ́kẹ̀lé interval comfortably include "no ipa," nítorí náà a kò lè claim àǹfààní, absolute ipa sì tiny.

Fún plain-English way láti ka odds ratios, ìgbẹ́kẹ̀lé intervals àti P values papọ̀, wo guide to ìṣègùn èsì.

Kò sí ààbò àmì — lábẹ́ ààlà. Kò sí serious adverse ìṣẹ̀lẹ̀ tí a judge gẹ́gẹ́ bí related sí irinṣẹ́, independent review sì kò rí ààbò àmì. Àwọn olùkọ̀wé honest nípa ceiling reassurance yìí: ìdánwò kò powered láti detect rare severe harms,

kò sì ní prespecified noninferiority tàbí formal ààbò àgbékalẹ, nítorí náà kò lè *fi ẹ̀rí múlẹ̀* ààbò fún uncommon ìṣẹ̀lẹ̀.

Documentation dára sí i. Nínú encounters 2,000 tí blinded experts review, clinicians tí wọ́n lo AI Consult produce better ìṣègùn documentation kọjá gbogbo rated domains — recorded ìdánimọ̀ àrùn, ìtọju plan àti overall completeness.

Prescribing fẹ̀rẹ̀ má yí. Kò sí significant difference nínú prescribing, pẹlu correct antibiotic lò (adjusted odds ratio 0.86, 95% CI 0.48 to 1.55). irinṣẹ́ kò yí antibiotic prescribing rates padà.

aláìsàn kò notice difference. Nínú aláìsàn 826 tí wọn complete satisfaction survey, satisfaction fẹ̀rẹ̀ identical láàárín arms, consultation times sì similar.

Costs tọ́ka díẹ̀ sí isalẹ. Nínú adjusted ìtúpalẹ̀, antibiotic-related costs kéré nínú AI arm — plausibly nítorí cheaper choices dípò fewer prescriptions — per-patient antibiotic saving sì dà bí ẹni pé ó ju per-patient cost ti running irinṣẹ́. Àwọn olùkọ̀wé flag èyí gẹ́gẹ́ bí suggestive, kì í ṣe settled: full total-cost-of-ownership accounting wà níta ìdánwò.

One-line summary àwọn olùkọ̀wé fúnra wọn ni cleanest ẹ̀dà: LLM assistance láìléwu lábẹ́ ààlà wọ̀nyí, ṣùgbọ́n kò dín kù ìtọju ìkùnà nínú 14 days, àǹfààní kankan probably modest.

Kí ló dé tí “no significant difference” kì í ṣe “it doesn’t iṣẹ́”

Null primary èsì rọrùn láti over-read ní ìtọ́ṣọ́nà méjì. Nǹkan méjì dá simple ìtàn dúró.

Àkọ́kọ́, **ìdánwò náà built láti catch ipa tó tóbi ju ohun tí ó rí lọ.** Serious bad èsì nínú primary care rare — tó 2% níbí — nítorí náà distinguishing small gidi àǹfààní láti ariwo nílò enormous numbers. Post-hoc power calculations àwọn olùkọ̀wé daba pé láti detect ipa ìwọ̀n tí wọ́n observe yóò nílò **ìdánwò tó tóbi gan-an, order of sí i than 100,000 aláìsàn.** Non-significant èsì nínú ìwádìí ìwọ̀n yìí kò òfin out small gidi àǹfààní; ó túmọ̀ sí pé ìwádìí yìí kò lè resolve rẹ̀.

Ẹlẹ́ẹ̀kejì, **ìfiwéra náà blurred díẹ̀.** Configuration àṣìṣe kan fún brief access sí AI Consult fún díẹ̀ control-arm clinicians, clinicians nínú shared network sì ń bá ara wọn sọ̀rọ̀, wọ́n sì carry habits kọjá ààlà. Effects méjèèjì ń mú arms méjèèjì jọra sí i, pushing any gidi difference toward zero. Lórí rẹ̀, host network ti already run sí relatively high standards, tó fi room kéré sílẹ̀ fún irinṣẹ́ láti fi hàn improvement.

Kò sí èyí tó rescue “breakthrough” àkọlé. Ṣùgbọ́n ó túmọ̀ sí pé correct reading jẹ́ calibrated, kì í ṣe deflationary: lórí hardest, most honest endpoint, irinṣẹ́ yìí kò demonstrably help aláìsàn nínú méjì weeks — nígbà tí ó measurably help record-keeping, ó sì looked láìléwu.

Ohun tí èyí kò fi ẹ̀rí múlẹ̀

- Kò fi hàn pé AI mú sunwọ̀n sí i aláìsàn èsì. Lórí primary 14-day endpoint, kò sí significant àǹfààní.
- Kò fi hàn pé AI useless. Point ìṣírò àfojúsùn favored rẹ̀, documentation mú sunwọ̀n sí i jálẹ̀, drug costs trend down; null èsì consistent pẹ̀lú small gidi àǹfààní tí ìdánwò kere ju láti confirm.

- Kò fi ẹ̀rí múlẹ̀ irinṣẹ́ láìléwu fún rare harms. Kò fi ààbò àmì hàn, ṣùgbọ́n ìdánwò kò powered tàbí designed láti certify ààbò fún uncommon severe ìṣẹ̀lẹ̀.
- Kò fi hàn pé “AI beats doctors” tàbí replaces clinicians. Èyí jẹ́ decision *ṣe àtìlẹ́yìn fún*; clinician pa full authority mọ́ láti accept tàbí reject.
- Kò generalize automatically. ìdánwò run nínú single private urban network ní Kenya; rural, periurban àti higher-income settings lè differ ní ìtọ́ṣọ́nà kankan.
- Kò establish cost savings. Cost àmì suggestive, kì í ṣe completed economic evaluation.

Báwo ni ẹ̀rí ṣe lágbára tó?

Fún central claim — no proven reduction nínú short-term aláìsàn harm — ẹ̀rí lágbára *gégé bí àpẹrẹ*, appropriately humble *gégé bí conclusion*. Prospective, pre-registered, cluster-randomized ìdánwò pẹ̀lú blinded patient-level composite èsì sún mọ́ best ayé gidi ẹ̀rí tí a lè gather fún irinṣẹ́ báyìí. Ó informative púpọ̀ ju benchmark score tàbí vignette ìwádìí lọ.

Fún secondary àwárí — better documentation, unchanged prescribing, similar satisfaction, lower antibiotic costs — ẹ̀rí good ṣùgbọ́n yẹ kí a ka wọn gégé bí secondary: supportive àmì, kì í ṣe àkọlé, wọ́n sì vulnerable sí contamination àti single-network ààlà.

Fún ààbò, ẹ̀rí reassuring ṣùgbọ́n bounded: no àmì rí, ṣùgbọ́n kì í ṣe ìwádìí built láti rí rare harms.

Most useful stance kì í ṣe “it works” tàbí “it failed.” Ó jẹ́: careful ayé gidi ìdánwò rí pé AI irinṣẹ́ yìí kò raise ààbò àmì, ó mú sunwọ̀n sí i ìlànà of care, ṣùgbọ́n kò demonstrate patient-outcome ànfààní nínú méjì weeks — àti detecting such ànfààní yóò nilo ìwádìí tó tóbi jù lọ.

Kí nìdí tí ó fi ṣe pàtàkì?

Debate nípa ìṣègùn AI ebi ń pa fún irú ẹ̀rí yìí. Thousands of ìwé ìwádìí wà tí àwòṣe ń ace exams, match clinicians lórí tidy ọ̀ràn. Large pragmatic randomized ìdánwò díẹ̀ ni wà tí ń measure bóyá gidi aláìsàn better off. Èyí jẹ́ ọ̀kan wọn, ó sì land lórí unglamorous truth: passing ìdánwò kì í dọ́gba pẹ̀lú helping aláìsàn.

Gap yẹn ni gbogbo ìtàn. irinṣẹ́ kan lè genuinely useful sí clinicians — clearer notes, lower drug bills, second pair of eyes — ṣùgbọ́n ṣì má move hard aláìsàn èsì nínú fortnight. Facts méjèèjì lè true ní akoko kan náà, mature health ètò gbọ́dọ̀ hold wọn papọ̀ dípò pick convenient ọ̀kan.

Ó tún reset burden of proof lórí useful. Tí company bá fẹ́ claim pé ìṣègùn AI rẹ̀ mú sunwọ̀n sí i care, relevant ẹ̀rí kì í ṣe leaderboard. Ó jẹ́ ìdánwò bí yìí, lórí èsì tí matter sí aláìsàn — ideally bigger ọ̀kan, nítorí honest lesson níbí ni pé modest ànfààní nílò large ìwádìí láti rí.

Àkótán kedere

Pragmatic cluster-randomized ìdánwò nínú 16 Kenyan primary care facilities ìdánwò generative-AI decision-support irinṣẹ́ (“AI Consult”) tí a fi kún electronic àkọsílẹ̀ tí ìṣègùn officers ń lo. Nínú 9,691 aláìsàn, expert-adjudicated composite ìtọju ìkùnà nínú 14 days jẹ́ 2.2% pẹ̀lú irinṣẹ́ versus 2.0% láìsí rẹ̀ (adjusted odds ratio 0.77, 95% CI 0.55–1.08, P = 0.13) — no significant difference. irinṣẹ́ kò fi ààbò àmì hàn, ó mú sunwọ̀n sí i ìṣègùn documentation kọjá gbogbo rated domains, kò yí prescribing padà, aláìsàn satisfaction unchanged, ó sì associated pẹ̀lú somewhat lower antibiotic costs. Àwọn olùkọ̀wé conclude pé ó láìléwu lábẹ́ ààlà wọ̀nyí, ṣùgbọ́n kò dín kù ìtọju ìkùnà, ànfààní kankan probably modest; detecting ipa ìwọ̀n tí wọ́n observe yóò nilo ìdánwò tó tóbi gan-an (order of 100,000 aláìsàn). èsì kò fi hàn pé ìṣègùn AI mú sunwọ̀n sí i aláìsàn èsì, kò sì fi hàn pé useless — ó fi hàn pé irinṣẹ́ tí ó looked láìléwu, tí ó help ìlànà, kò demonstrably help aláìsàn nínú méjì weeks nínú urban private network kan.

Àyẹ̀wò láìsí àṣejù

Ohun tí ìwé ìwádìí fi hàn: Nínú ayé gidi randomized ìdánwò, adding LLM-based decision-support irinṣẹ́ sí primary care àkọsílẹ̀ kò raise ààbò àmì, mú sunwọ̀n sí i ìṣègùn documentation quality, kò yí prescribing padà, kò sì significantly dín kù strict 14-day aláìsàn composite of ìtọju ìkùnà.

Ohun tó ṣeé gbà ṣùgbọ́n tí a kò fi ẹ̀rí múlẹ̀: Pé irinṣẹ́ produce small gidi reduction nínú ìtọju ìkùnà tó kéré ju ìdánwò yìí láti detect; pé ó save money lẹ́yìn full costs; pé better documentation eventually translate sí better care.

Ohun tí kò fi hàn: Pé ìṣègùn AI mú sunwọ̀n sí i aláìsàn èsì; pé ó unsafe fún rare ìṣẹ̀lẹ̀; pé ó replace tàbí outperform clinicians; pé èsì transfer sí rural tàbí higher-income settings; pé cost saving established.

Main limitations: Powered fún larger ipa ju observed lọ (rare èsì nílò huge àpẹẹrẹ); configuration àṣìṣe contaminate control arm, shared-network habits sì blur ìfiwéra, méjèèjì bias toward no difference; single private urban network pẹ̀lú already-high standards; no prespecified noninferiority tàbí ààbò àgbékalẹ̀; short 14-day horizon.

Confidence wo ni gbogbogbò reader yẹ kí ó ní? Gíga pé irinṣẹ́ kò fi ààbò àmì hàn nínú ìdánwò yìí, ó sì mú sunwọ̀n sí i documentation. Gíga pé kò demonstrably mú sunwọ̀n sí i 14-day aláìsàn èsì níbí. Kéré fún claim kankan pé ó “works” tàbí “fails” gẹ́gẹ́ bí patient-benefit intervention — ìbéèrè náà genuinely unresolved, ó sì nílò much larger ìwádìí. Ìpò tó yẹ: careful ayé gidi èsì nípa irinṣẹ́ tí kò raise ààbò àmì, tó mú sunwọ̀n sí i ìlànà of care, kì í ṣe verdict pé AI transforms tàbí wrecks primary care.

Àwọn orísun

Agweyu et al., Nature Medicine (2026)
https://doi.org/10.1038/s41591-026-04503-6

Pan-African Clinical Trials Registry, registration 202502499779176
https://pactr.samrc.ac.za/

EurekAlert / PATH press release on the trial (2026)
https://www.eurekalert.org/news-releases/1133583