The Clean Paper

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

Clinical AI da ya yi kama da safe kuma ya inganta paperwork — amma bai inganta patient outcomes ba

Pragmatic cluster-randomized trial a Kenyan primary care facilities 16 ya gwada generative-AI decision-support tool da aka kara ga record da clinical officers suke amfani da shi. A patients 9,691, strict 14-day expert-adjudicated composite na treatment failure ya kasance 2.2% da tool vs 2.0% ba tare da shi ba (adjusted odds ratio 0.77, 95% CI 0.55–1.08, P = 0.13) — babu significant difference. Tool bai nuna safety signal ba, ya improve documentation a duk rated domains, prescribing da satisfaction ba su canza ba, kuma antibiotic costs sun trend down. Wannan ba failed study ba ne; rare, honest one ne — measured on patients, ba benchmarks ba — kuma yana nuna unglamorous truth cewa tool da ya look safe kuma ya help process bai demonstrably help patients cikin weeks biyu ba, kuma detect any such benefit zai bukaci far larger trial.

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

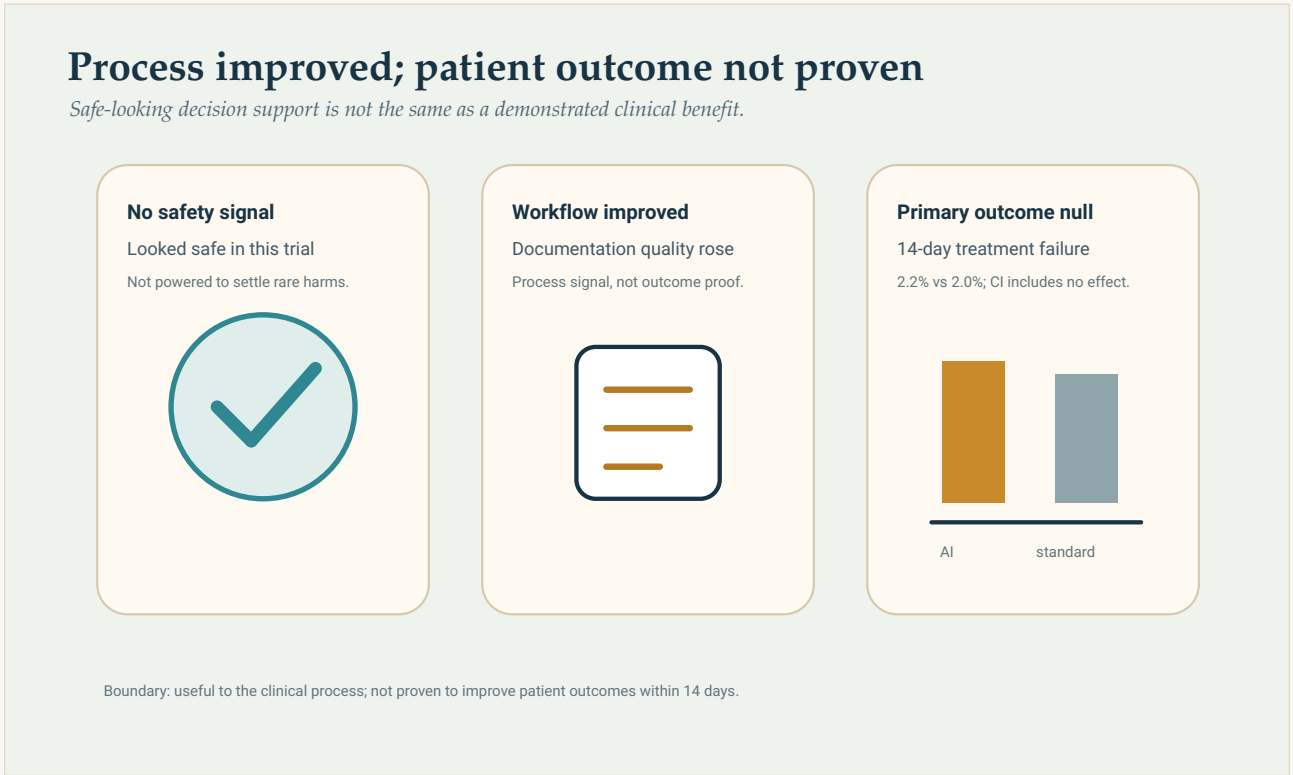
Kasancewa lafiya kuma mai amfani ba daidai yake da tab-batar da inganci ba

Yawancin kanun labarai game da AI a likitanci suna fitowa ne daga irin binciken da bai amsa tambayar da ta fi muhimmanci ba. Wani samfuri ya yi kyau a tambayoyin jarrabawa, ko ya fi likitoci a wasu misalan marasa lafiya da aka shirya da kyau, sai a ce injin ya shirya shiga asibiti.

Wannan gwajin ya yi abu mafi wahala kuma da ba kasafai ake yi ba. Ya saka wani kayan tallafa wa yanke shawara na generative AI cikin kulawar farko ta gaske, a cibiyoyi na gaske, tare da marasa lafiya na gaske, sannan ya tambayi abin da ya fi muhimmanci: **shin marasa lafiya sun samu sakamako mafi kyau?**

Amsar da ta fi gaskiya ita ce a'a — ba a ga wani bambanci da za a iya aunawa cikin kwanaki 14 ba. Kayan bai nuna wata alamar matsalar aminci ba. Ya inganta yadda aka rubuta bayanan asibiti. Wataƙila ma ya rage wasu kuɗin magunguna. Amma bai rage gazawar magani da muhimmancin kididdiga ba, kuma marubutan sun yi taka-tsantsan wajen cewa duk wata fa'ida, idan tana akwai, mai yiwuwa ƙarama ce.

Wannan ba gazawar binciken ba ce. Wannan ne binciken yake yi idan an tsara shi da kyau. Haka hujja mai alhaki game da AI a likitanci take idan ana auna abin da ya faru ga **marasa lafiya**, ba maki a benchmark ba.



Kayan bai nuna wata alamar matsalar aminci ba kuma ya inganta rubutun bayanan asibiti, amma babban sakamakon da aka kayyade tun farko na kwanaki 14 bai inganta da muhimmancin kididdiga ba. Taimakon tsarin aiki ba daidai yake da tabbatacciyar fa'ida ga mara lafiya ba.

Abin da marubutan suka yi

Tawagar ta gudanar da gwaji na aikace-aikacen yau da kullum, tare da bazuwar rarrabawa bisa rukunin ma'aikata, a **cibiyoyin kulawar farko 16** na cibiyar lafiya mai zaman kanta Penda Health a Nairobi da Kiambu, Kenya. A waɗannan cibiyoyi, **clinical officers** — ma'aikatan lafiya na matsakaicin mataki da ke da difloma ta shekaru uku — ne ke ba da babban ɓangaren kulawa, sau da yawa ba tare da sauƙin samun shawarar manyan likitoci ba.

An yi bazuwar rarrabawar ne bisa **clinician**, ba bisa mara lafiya ba. An rarraba **clinical officers 103**: 52 zuwa rukunin da zai yi amfani da kayan, 51 kuma zuwa rukunin kwatantawa. Duk rukunan biyu sun yi amfani da kundin lafiya na lantarki iri ɗaya. Rukunin farko ya ƙara samun **AI Consult** (version 2.0), kayan tallafa wa yanke shawara da aka gina a kan babban samfurin harshe **GPT-4o na OpenAI** kuma aka haɗa shi cikin kundin. Yana karanta abin da ma'aikacin lafiya ya rubuta, sannan yana iya nuna yiwuwar matsala a ganewar cuta ko shirin magani. Ma'aikatan lafiya sun ci gaba da riƙe cikakken ikon yanke shawara: za su iya karɓa, gyara ko yin watsi da shawarwarinsa.

> **Wane samfuri aka yi amfani da shi, kuma me ya sa cikakkun bayanan suke da muhimmanci**
>
> Kayan shi ne **AI Consult 2.0**, yana amfani da **GPT-4o na OpenAI** (sakin Mayu 2025), ta API na kasuwanci na OpenAI ƙarƙashin lasisin enterprise, tare da ƙaramin randomness (temperature 0.1). An saka shi cikin kundin lafiya na musamman, EasyClinic's EMR, kuma system prompts ɗinsa an rubuta su ne domin su dace da jagororin magani na ƙasar Kenya. Marubutan sun wallafa cikakken prompt ɗin umarni.
>
> Me ya sa ake bayyana wannan dalla-dalla? Saboda sakamakon ya shafi **wani tsari takamaimai** — sigar samfuri ɗaya, prompt ɗaya, kundin lafiya ɗaya da yanayin saiti ɗaya — ba "LLMs a likitanci" gaba ɗaya ba. Marubutan ma sun jaddada haka: sun kira sakamakonsu *ma'aunin wani lokaci, ba dindindindin ma'aunin ƙarfin fasahar ba*. Sabon samfuri, wani prompt ko cibiyar lafiya da ba ta da cikakken tsarin dijital na iya ba da wani sakamako dabam.
>
> Game da 'yancin binciken: daga baya OpenAI ta bayar da tallafi a matsayin cloud-compute credits da shawarar fasaha kan amfani da API. Amma marubutan sun ce an yanke shawarar amfani da OpenAI **kafin** wannan tayin, kuma OpenAI ba ta shiga tsara gwajin, tattara bayanai, nazari ko yanke shawarar wallafawa ba.

Tsakanin 22 Afrilu da 16 Yuli 2025, an shigar da **marasa lafiya 9,691**. **Babban sakamakon gwajin an tsara shi ne ya shafi mara lafiya kai tsaye kuma ya kasance mai tsauri:** wani kwamitin ƙwararru ya tantance haɗaɗɗen ma'aunin *gazawar magani* cikin kwanaki 14 bayan ziyarar. Masu tantancewar ba su san wane rukuni mara lafiyar yake ba, kuma suna duba ko ya sami mummunan sakamako kamar rashin warkewa ko ƙara tsanantar rashin lafiya. An yi rajistar gwajin tun kafin farawa (Pan-African Clinical Trials Registry 202502499779176).

Wannan zaɓin ma'auni shi ne muhimmin abu. Abu ne mai sauƙi a nuna cewa kayan AI ya canza abin da ma'aikacin lafiya yake rubutawa. Abu ne mafi wahala — kuma mafi muhimmanci — a nuna cewa ya canza abin da ya faru ga mara lafiya.

Abin da suka gano

Babban sakamakon bai inganta ba. Gazawar magani ta faru a marasa lafiya 102 cikin 4,693 (2.2%) a rukunin AI da 94 cikin 4,654 (2.0%) a rukunin kwatantawa. Idan aka kalli kaso kawai, rukunin AI ya ɗan fi girma; amma bayan gyara bisa bambance-bambancen rukunan clinicians, kiyasin ya karkata kaɗan zuwa fa'ida: **adjusted odds ratio 0.77 (95% CI 0.55 zuwa 1.08, P = 0.13)**. Wannan **ba sakamako mai muhimmancin kididdiga ba ne**. Bambancin da ke tsakanin lambobin ɗanye da na bayan gyara ba kuskuren lissafi ba ne; gyaran yana la'akari da bambance-bambance tsakanin rukunin clinicians. A kowane hali, confidence interval ya ƙunshi "babu tasiri" a fili, don haka ba za a iya cewa an tabbatar da fa'ida ba. Kuma idan aka kalli adadi kai tsaye, bambancin ƙarami ne sosai.

Don bayani mai sauƙi game da yadda ake karanta odds ratio, confidence interval da P value tare, duba jagorar karanta sakamakon asibiti.

Ba a ga alamar matsalar aminci ba — amma wannan tabbacin yana da iyaka. Ba a samu serious adverse event da aka danganta da kayan ba, kuma wani bita mai zaman kansa bai ga alamar matsalar aminci ba. Marubutan sun bayyana iyakar wannan kwanciyar hankali: gwajin bai isa ya gano matsaloli masu tsanani da ba kasafai suke faruwa ba, kuma ba a tsara masa tun farko gwajin noninferiority ko cikakken tsarin aminci ba. Saboda haka ba zai iya **tabbatar da aminci** ga matsalolin da ba sa yawan faruwa ba.

Rubutun bayanan asibiti ya inganta. Daga ziyarori 2,000 da ƙwararru masu makanta ga rukunin gwajin suka duba, clinicians masu amfani da AI Consult sun rubuta bayanai mafi kyau a duk bangarorin da aka tantance — ganewar cuta, shirin magani da cikar bayanan gaba ɗaya.

Rubuta magunguna bai kusan canzawa ba. Ba a ga wani bambanci mai muhimmanci a rubuta magunguna ba, ciki har da amfani da antibiotics daidai (adjusted odds ratio 0.86, 95% CI 0.48 zuwa 1.55). Kayan bai canza yawan rubuta antibiotics ba.

Marasa lafiya ba su lura da wani bambanci ba. Daga marasa lafiya 826 da suka cika tambayoyin gamsuwa, matakin gamsuwa ya kasance kusan iri ɗaya a rukunan biyu, kuma lokacin konsultation ma ya yi kama.

Kuɗin magunguna ya nuna ɗan karkata zuwa ƙasa. Bayan gyara, kuɗin da ke da alaƙa da antibiotics ya yi ƙasa a rukunin AI — mai yiwuwa saboda zaɓin magunguna masu rahusa, ba saboda an rubuta antibiotics kaɗan ba. A kiyasin marubutan, ajiyar kuɗin antibiotic ga kowane mara lafiya ta yi kama da ta fi kuɗin gudanar da kayan ga kowane mara lafiya. Amma sun ɗauki wannan a matsayin alama kawai, ba hukunci ba: cikakken lissafin dukan kuɗin mallaka da gudanarwa bai kasance cikin gwajin ba.

Takaitaccen bayanin marubutan shi ne mafi tsafta: a cikin iyakokin wannan gwajin, taimakon LLM bai nuna matsalar aminci ba amma bai rage gazawar magani cikin kwanaki 14 ba; duk wata fa'ida mai yiwuwa tana da alamar kasancewa ƙarama.

Me ya sa “babu bambanci mai muhimmancin kididdiga” ba ya nufin “ba ya aiki”

Yana da sauƙi a wuce gona da iri wajen fassara babban sakamako da bai kai muhimmancin kididdiga ba, ko ta hanyar cewa kayan ba ya da amfani, ko kuma ta hanyar neman hujjar da ba ta nan. Abubuwa biyu suna hana wannan sauƙaƙen labari.

Na farko, **an tsara gwajin ne don gano tasiri mafi girma fiye da wanda aka gani.** Mummunan sakamako a kulawar farko ba ya yawan faruwa — kusan 2% a nan — don haka gano ƙaramin fa’ida ta gaske a tsakanin hayaniyar kididdiga yana buƙatar mutane masu yawa sosai. Lissafin power da marubutan suka yi bayan gwajin ya nuna cewa gano tasiri da girman wanda aka gani zai buƙaci **gwaji mafi girma sosai, kimanin sama da marasa lafiya 100,000**. Sakamako marar muhimmancin kididdiga a gwaji irin wannan ba ya kore yiwuwar ƙaramin amfani na gaske; yana nufin gwajin bai iya bambance shi da hayaniya ba.

Na biyu, **kwatantawar rukunan ta ɗan gauraye.** Wani kuskuren saiti na ɗan lokaci ya ba wasu clinicians na rukunin kwatantawa damar samun AI Consult. Haka kuma clinicians a cibiyar aiki ɗaya suna magana da juna kuma suna iya ɗaukar sabbin halaye daga wani rukuni zuwa wani. Duk waɗannan abubuwa suna sa rukunan biyu su yi kama, kuma idan akwai bambanci na gaske suna tura kiyasin zuwa sifili. Bugu da ƙari, cibiyar da aka yi gwajin tana aiki da ƙa’idoji masu kyau tun farko, don haka akwai ƙaramin sarari da kayan zai nuna babban ci gaba.

Babu ɗayan waɗannan bayanan da zai maido da kanun labarin “breakthrough.” Amma suna nufin karatun da ya dace ya zama mai ma’auni: a kan ma’aunin da ya fi wahala kuma ya fi dacewa da mara lafiya, kayan bai nuna tabbatacciyar fa’ida cikin makonni biyu ba — yayin da ya inganta rubutun bayanai kuma bai nuna matsalar aminci ba.

Abin da wannan binciken bai tabbatar ba

- Bai nuna cewa AI ya inganta sakamakon marasa lafiya ba. A babban ma’aunin kwanaki 14, babu tabbatacciyar fa’ida.
- Bai nuna cewa AI ba shi da amfani ba. Kiyasin bayan gyara ya karkata zuwa fa’ida, rubutun bayanai ya inganta, kuma kuɗin magunguna ya nuna karkata zuwa ƙasa; sakamakon da bai kai muhimmancin kididdiga ba har yanzu ya dace da yiwuwar ƙaramin amfani da gwajin bai isa ya tabbatar ba.
- Bai tabbatar da cewa kayan yana da aminci ga matsaloli masu wuya da ba kasafai suke faruwa ba. Babu alamar matsalar aminci da aka gani, amma gwajin bai isa ko aka tsara shi don tabbatar da irin wannan aminci ba.
- Bai nuna cewa “AI ya fi likitoci” ko zai maye gurbin clinicians ba. Wannan kayan **tallafa wa yanke shawara** ne; clinician ya riƙe cikakken ikon karɓa ko ƙin shawararsa.
- Ba za a ɗauki sakamakon kai tsaye zuwa kowane wuri ba. Gwajin ya gudana a cibiyar lafiya mai zaman kanta guda ɗaya a biranen Kenya; karkara, gefen birane da ƙasashe masu kuɗi da yawa na iya ba da wani sakamako dabam.
- Bai tabbatar da ajiyar kuɗi ba. Alamar kuɗin tana da ban sha’awa, amma ba cikakken binciken tattalin arziki ba ne.

Yaya ƙarfin shaidar yake?

Ga babban ikirari — cewa ba a tabbatar da raguwar cutar mara lafiya cikin gajeren lokaci ba — tsarin shaidar yana da ƙarfi, yayin da hukuncin ya kasance mai taka-tsantsan yadda ya kamata. Gwaji na gaba-gaba, da aka yi rajista tun farko, tare da bazuwar rarrabawa bisa rukunin clinicians da kuma ma'aunin mara lafiya da ƙwararru masu makanta suka tantance, yana daga cikin mafi kyawun shaidar rayuwar yau da kullum da za a iya tarawa ga irin wannan kayan. Ya fi maki a benchmark ko binciken vignettes bayani sosai.

Ga sakamako na biyu — ingantaccen rubutun bayanai, rashin canjin rubuta magunguna, gamsuwar marasa lafiya iri ɗaya da ƙananan kuɗin antibiotics — shaidar tana da kyau, amma ya kamata a karanta ta a matsayin **sakandare**, ba babban sakon binciken ba. Haka kuma tana ƙarƙashin matsalolin gaurayewar rukuni da iyakar cibiyar guda ɗaya.

Ga aminci, shaidar tana kwantar da hankali amma tana da iyaka: ba a ga wata alama ba, amma ba wannan ne gwajin da aka tsara don gano matsalolin da ba kasafai suke faruwa ba.

Matsayin da ya fi amfani ba "yana aiki" ko "ya gaza" ba ne. Shi ne: **gwaji na gaske da aka tsara da kyau ya nuna cewa wannan kayan AI bai nuna matsalar aminci ba kuma ya inganta tsarin rubutun kulawa, amma bai tabbatar da fa'ida ga marasa lafiya cikin makonni biyu ba; gano ƙaramin amfani zai buƙaci gwaji mafi girma sosai.**

Me ya sa wannan yake da muhimmanci

Muhawarar AI a likitanci tana fama da ƙarancin irin wannan shaida. Akwai dubban takardu da ke nuna samfura suna cin jarrabawa ko suna daidaita da clinicians a misalan da aka shirya. Amma manyan gwaje-gwaje na aikace-aikacen yau da kullum da aka rarraba bazuwar lokaci, waɗanda suke auna ko marasa lafiya na gaske sun fi samun sauƙi, kaɗan ne sosai. Wannan ɗaya ne daga cikinsu, kuma sakamakonsa ya kawo gaskiyar da ba ta da armashi amma tana da muhimmanci: **cin jarrabawa ba daidai yake da taimaka wa mara lafiya ba.**

Wannan tazara ita ce labarin. Kayan zai iya zama mai amfani ga clinicians — rubutu mafi kyau, ƙananan kuɗin wasu magunguna, wani ido na biyu kan shirin kulawa — amma duk da haka bai motsa wani babban sakamako ga mara lafiya cikin makonni biyu ba. Gaskiyoyin biyu na iya kasancewa tare, kuma tsarin lafiya mai balaga ya kamata ya iya riƙe su tare maimakon ya zaɓi wanda ya fi dacewa da labarin da yake so.

Haka kuma binciken yana mayar da nauyin hujja zuwa wurin da ya dace. Idan kamfani yana son cewa AI ɗinsa na asibiti yana inganta kulawa, hujjar da ta dace ba leaderboard ba ce. Gwaji irin wannan ne, da sakamako da suke da ma'ana ga marasa lafiya — kuma, a mafi kyau, gwaji mafi girma, domin darasin da aka samu a nan shi ne ƙaramin amfani yana buƙatar manyan adadin mutane kafin a iya ganinsa da kyau.

Takaitaccen bayani

Wani gwaji na aikace-aikacen yau da kullum da aka rarraba bazuwar lokaci bisa rukunin clinicians a cibiyoyin kulawar farko **16** a Kenya ya gwada kayan tallafa wa yanke shawara na generative AI, **AI Consult**, wanda aka ƙara cikin kundin lafiya na lantarki da clinical officers suke amfani da shi. Daga **marasa lafiya 9,691**, haɗaɗɗen ma'aunin gazawar magani cikin kwanaki 14 da ƙwararru suka tantance ya kasance 2.2% tare da kayan da 2.0% ba tare da shi ba (adjusted odds ratio 0.77, 95% CI 0.55–1.08, P = 0.13) — **babu bambanci mai muhimmancin kididdiga**. Kayan bai nuna matsalar aminci ba, ya inganta rubutun bayanan asibiti a duk bangarorin da aka tantance, bai canza rubuta magunguna ba, bai canza gamsuwar marasa lafiya ba, kuma ya kasance tare da ɗan raguwar kuɗin antibiotics. Marubutan sun kammala cewa, cikin iyakokin gwajin, kayan bai nuna matsalar aminci ba amma bai rage gazawar magani ba; duk wata fa'ida mai yiwuwa tana da alamar kasancewa ƙarama. Gano tasiri da girman wanda aka gani zai buƙaci gwaji mai kusan sama da marasa lafiya 100,000. Sakamakon bai nuna cewa clinical AI yana inganta sakamakon marasa lafiya ba, kuma bai nuna cewa ba shi da amfani ba; yana nuna cewa kayan da ya taimaka wajen tsarin kulawa bai nuna tabbatacciyar fa'ida ga mara lafiya cikin makonni biyu ba a wannan cibiyar birni guda ɗaya.

Bincike ba tare da ƙarin gishiri ba

Abin da takardar ta nuna: A gwaji na gaske da aka rarraba bazuwar lokaci, ƙara kayan tallafa wa yanke shawara da ke amfani da LLM cikin kundin kulawar farko bai nuna matsalar aminci ba, ya inganta ingancin rubutun bayanan asibiti, bai canza rubuta magunguna ba, kuma bai rage babban haɗaɗɗen ma'aunin gazawar magani na kwanaki 14 da muhimmancin kididdiga ba.

Abin da yake yiwuwa amma ba a tabbatar ba: Kayan yana rage gazawar magani da ɗan ƙaramin kaso da wannan gwajin bai iya ganowa ba; yana iya adana kuɗi idan aka ƙirga duk kuɗin gudanarwa; ingantaccen rubutun bayanai zai iya haifar da ingantacciyar kulawa a dogon lokaci.

Abin da bai nuna ba: Cewa clinical AI ya inganta sakamakon marasa lafiya; cewa an tabbatar da aminci ga matsaloli masu wuya; cewa zai maye gurbin ko ya fi clinicians; cewa sakamakon zai kasance iri ɗaya a karkara ko ƙasashe masu kuɗi da yawa; ko cewa an tabbatar da ajiyar kuɗi.

Manyan iyakoki: An tsara gwajin don tasiri mafi girma fiye da wanda aka gani, yayin da mummunan sakamako ba ya yawan faruwa kuma ƙaramin tasiri yana buƙatar samfurin mutane masu yawa; kuskuren saiti ya sa wasu clinicians na rukunin kwatantawa suka sami kayan, kuma ma'aikata a cibiyar guda suna musayar halaye — duka suna sa bambancin rukuni ya ragu; cibiyar birni mai zaman kanta guda ɗaya ce kuma tana da ƙa'idoji masu kyau tun farko; babu noninferiority ko cikakken tsarin aminci da aka kayyade tun farko; lokacin bibiyar babban sakamako kwanaki 14 ne kawai.

Yawan amincewar da ya dace ga mai karatu na gama gari: Babba cewa kayan bai nuna matsalar aminci a wannan gwajin ba kuma ya inganta rubutun bayanai. Babba cewa bai nuna tabbatacciyar inganta sakamakon mara lafiya cikin kwanaki 14 ba a nan. Ƙasa ga duk wani hukunci cewa "yana aiki" ko "ya gaza" wajen amfanin mara lafiya — wannan tambayar har yanzu ba a warware ta ba kuma tana buƙatar gwaji mafi girma. Matsayin da ya dace shi ne: wannan sakamakon gaske ne kuma mai taka-tsantsan game da kayan da ya taimaka wa tsarin kulawa, ba hukuncin cewa AI zai sauya ko lalata kulawar farko ba.

Majiyoyi

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