



The CLEAN PAPER

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

“محفوظ اور ”helpful“ ہونا ”effective“ ہونے کے برابر نہیں

Kenyan 16 بنیادی-نگہداشت facilities کے cluster-RCT میں “اے آئی ”GPT-4o Consult“ (based) کو کلینیکل officers کے EMR میں add کیا گیا۔ 9,691** مریض 14-day ماہر-adjudicated علاج ناکامی 2% vs 0%۔2 کنٹرول؛ OR adjusted 77.0**، CI 95% 08.1-55.0، 13**۔P=0۔ اہم فرق۔ حفاظت اشارہ نہیں، دستاویزات تمام domains بہتر، prescribing/satisfaction، antibiotic unchanged، لاگت somewhat کم اشارہ۔ مصنفین: limits کے اندر محفوظ، علاج ناکامی reduce نہیں، ممکن فائدہ modest، مشاہدہ شدہ اثر شناخت کرنا کرنے کو 100k+ مریض پیمانہ چاہیے۔ یہ مریض فائدہ ثابت کرنا نہیں کرتا، uselessness بھی نہیں۔ عمل help ہوئی، مریض نتیجہ demonstrably نہیں۔

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: 6-04503-026-s41591/1038.10

“محفوظ اور helpful ہونا” مؤثر ہونے کے برابر نہیں

طبی اے آئی کی زیادہ headlines غلط of kind مطالعہ سے بنتی ہیں۔ ماڈل exam سوالات پر اچھا اسکور کرے یا curated vignettes میں ڈاکٹرز کو سے بہتر رہا کرے، کہانی تیار: مشین کلینک کے لیے ready ہے۔

اس ٹرائل نے زیادہ مشکل اور نایاب کام کیا۔ اے-generative آئی فیصلہ-تائید کو حقیقی بنیادی نگہداشت، حقیقی facilities، حقیقی مریض میں رکھا اور اصل سوال پوچھا: کیا مریض بہتر ہوئے؟

دیانت دار جواب: نہیں—14 دن میں قابل پیمائش طور پر نہیں۔ ٹول نے کوئی حفاظت اشارہ نہیں دکھایا۔ کلینیکل دستاویزات بہتر ہوئی۔ کچھ drug لاگت شاید کم ہوئیں۔ مگر علاج significantly failures کم نہیں ہوئے، اور مصنفین صاف کہتے ہیں اگر فائدہ ہے بھی تو غالباً محدود ہے۔

یہ مطالعے کی ناکامی نہیں؛ مطالعے کا صحیح کام کرنا ہے۔ طب میں اے آئی شواہد مریض نتائج پر پیمائش ہو تو کچھ ایسا ہی دکھتا ہے، بینچ مارک پر نہیں۔

Process improved; patient outcome not proven

Safe-looking decision support is not the same as a demonstrated clinical benefit.

No safety signal

Looked safe in this trial

Not powered to settle rare harms.



Workflow improved

Documentation quality rose

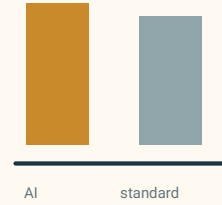
Process signal, not outcome proof.



Primary outcome null

14-day treatment failure

2.2% vs 2.0%; CI includes no effect.



Boundary: useful to the clinical process; not proven to improve patient outcomes within 14 days.

ٹول نے حفاظت اشارہ نہیں دکھایا اور دستاویزات بہتر کی، مگر prespecified 14-دن مریض نتیجہ significantly بہتر کرنا نہیں ہوا۔ عمل مدد ثابت شدہ مریض فائدہ نہیں۔

مصنفین نے کیا کیا

ٹیم نے Kenya Nairobi/Kiambu میں Penda صحت کی 16 بنیادی-نگہداشت facilities پر pragmatic تصادفی-cluster ٹرائل کیا۔ نگہداشت largely کلینیکل officers دیتے ہیں—3 سال diploma والے سطح mid—جنہیں—consultation senior practitioners—آسانی سے دستیاب نہیں ہوتی۔

بے ترتیب تقسیم معالج سطح پر، مریض نہیں۔ 103 کلینیکل officers: 52 مداخلت، 51 کنٹرول۔ دونوں ایک ہی بادل EMR۔ مداخلت کو اضافی طور پر اے آئی Consult 0.2، OpenAI، GPT-4o پر مبنی فیصلہ-تائید، ریکارڈ میں ضم۔ معالج کی documented info پڑھ کر تشخیص/علاج plan میں ممکن issues نشان۔ معالج مکمل خودمختاری: قبول/modify/ignore۔

کون سا ماڈل تھا، اور specifics کیوں ماڈل کرتے ہیں

درجہ کم-تصادف، API enterprise licence تجارتی اجراء، 2025 مئی OpenAI GPT-4o تھا، Consult 0.2 آئی اے ٹول مکمل لیے؛ کے کرنے align اصول رہنما علاج قومی Kenyan ہدایات نظام ضم، میں EasyClinic EMR 1.0۔ حرارت شدہ۔ شائع ہدایت ہدایت

یہ اس لیے اہم ہے کہ نتیجہ ایک مخصوص نظام کے بارے میں ہے—ایک ماڈل ورژن، ایک ہدایت، ایک ریکارڈ، ایک ترتیب—“in LLMs” طب “عمومی نہیں۔ مصنفین اسے temporal بینچ مارک کے بجائے مقرر تخمینہ of صلاحیت کہتے ہیں۔ نئے ماڈل/ہدایت/کم-ڈیجیٹل ریکارڈ والے کلینک نتیجہ بدل سکتے ہیں۔

آزادی: OpenAI نے بعد میں in-kind تائید (بادل کریڈٹس/تکنیکی API رہنمائی) دی، مگر مصنفین کہتے ہیں OpenAI choose کرنے کا فیصلہ offer سے پہلے تھا، اور کمپنی ٹرائل ڈیزائن/ڈیٹا جمع آوری/تجزیہ/اشاعت فیصلہ میں شامل نہیں تھی۔

22 اپریل-16 جولائی 2025، 9,691 مریض۔ بنیادی نتیجہ جان بوجھ کر سخت/مریض-مرکز: ملاقات کے 14 دن کے اندر ماہر-ماہرین سے جانچا گیا مرکب علاج ناکامی blinded—معالج پینل نے حل طلب/بگرتی ہوئی بیماری جیسے خراب نتائج judge کیے۔ ٹرائل پیش رفت (Pan-African registered) کلینیکل ٹرائلز Registry (202502499779176)۔

یہ ڈیزائن انتخاب نقطہ ہے۔ اے آئی معالج کے notes بدلتی ہے دکھانا آسان؛ مریض پر اثر دکھانا مشکل اور معنی خیز۔

انہوں نے کیا پایا

بنیادی نتیجہ بہتر کرنا نہیں ہوا۔ علاج ناکامی اے آئی گروہ 102/4,693 (2%۔2)، کنٹرول 94/4,654 (0%۔2)۔ percentages crude۔ اے آئی میں fractionally زیادہ؛ adjustment کے بعد نقطہ تخمینہ فائدہ کی طرف: odds adjusted تناسب 77.0 (95% CI 55.0–08.1)، P=0.13—شماراتی طور پر اہم نہیں۔ arithmetic flip raw/adjusted غلطی نہیں، cluster فرق adjust ہوتی ہیں۔ CI نہیں-اثر شامل کرتی ہے، فائدہ دعویٰ نہیں؛ مطلق اثر۔ tiny کلینیکل-تنازع رہنما۔

حدود کے اندر نہیں حفاظت اشارہ۔ ٹول-متعلقہ سنجیدہ adverse واقعات نہیں؛ آزاد جائزہ نہیں حفاظت اشارہ۔ مگر ٹرائل نایاب severe harms شناخت کرنے کے لیے powered نہیں، prespecified صورتی/noninferiority حفاظت فریم ورک نہیں؛ واقعہ-uncommon حفاظت ثابت کرنا نہیں۔

دستاویزات بہتر۔ 2,000 encounters blinded-reviewed میں تشخیص، علاج plan، مجموعی تکمیل سب domains اے آئی گروہ بہتر۔

نسخہ نویسی بمشکل - changed درست اینٹی بایوٹک استعمال سمیت اہم فرق نہیں (aOR 86.0, 95% CI 48.0-55.1)؛ اینٹی بایوٹک نسخہ نویسی شرح تقریباً غیر تبدیل۔

مریض کو فرق محسوس نہیں ہوا۔ 826 respondents satisfaction میں essentially satisfaction یکساں، consultation اوقات مشابہ۔

لاگت ہلکا کمی کا اشارہ۔ adjusted تجزیہ میں اینٹی بایوٹک-متعلقہ لاگت اے آئی گروہ کم-ممکن cheaper انتخاب، fewer prescriptions نہیں-اور فی-مریض بچت ٹول چلانے کی لاگت سے greater دکھائی دی۔ اشارہ دینے والا، مکمل معاشی جائزہ نہیں۔

مصنفین کا ایک-لائن صاف خلاصہ: حدود کے اندر LLM assistance محفوظ رہی، 14-دن علاج ناکامی کم کرنا نہیں ہوا، any فائدہ غالباً محدود۔

“نہیں اہم فرق” کا مطلب “کام نہیں کرتی” نہیں

بنیادی نتیجے کو دونوں سمتوں میں ضرورت سے زیادہ معنی دیا جا سکتا ہے۔

پہلی، ٹرائل مشاہدہ شدہ اثر سے بڑا اثر پکڑنے کے لیے تیار تھا۔ خراب بنیادی-نگہداشت نتائج ~2% نایاب؛ چھوٹا فائدہ شور سے distinguish کرنے کے لیے huge N- بعد از وقوع طاقت کے مطابق مشاہدہ شدہ-جسم اثر شناخت کرنے کو <100,000 مریض ترتیب ٹرائل چاہیے۔ غیر-اہم نتیجہ چھوٹا حقیقی فائدہ قاعدہ out نہیں کرتا؛ مطالعہ واضح کرنا نہیں کر سکی۔

دوسری، موازنہ جزوی طور پر blurred ترتیب غلطی سے کچھ کنٹرول معالجین کو مختصر وقت کے لیے اے آئی رسائی؛ مشترک نیٹ ورک معالجین عادات/knowledge حصہ۔ دونوں بازو مزید alike → حقیقی فرق صفر کی طرف۔ host bias نیٹ ورک already نسبتاً بلند معیارات، بہتری headroom کم۔

یہ rescue breakthrough نہیں کرتا۔ calibrated مطالعہ: hardest اختتامی معیار پر weeks two میں واضح طور پر ثابت مریض فائدہ نہیں، جبکہ عمل/دستاویزات بہتر کرنا اور حفاظت اشارہ غائب۔

یہ کیا ثابت نہیں کرتا

- اے آئی نے مریض نتائج بہتر کرنا کیے-نہیں، بنیادی اختتامی معیار غیر-اہم۔
- اے آئی -useless نہیں، نقطہ تخمینہ، favourable دستاویزات بہتر کرنا، لاگت کمی کا اشارہ؛ چھوٹا فائدہ حل طلب۔
- نایاب harms کے لیے حفاظت ثابت کرنا-نہیں، نہیں اشارہ ≠ نایاب-واقعہ۔ certification۔
- “اے آئی سے بہتر ہے ڈاکٹرز” یا جگہ لیتا ہے معالجین-نہیں، فیصلہ تائید؛ معالج اختیار مکمل۔
- خودکار عمومی اطلاق-نہیں، واحد نجی شہری Kenyan نیٹ ورک؛ دیہی/نیم شہری/بلند-آمدنی مختلف ہو سکتے۔
- لاگت savings—establish نہیں، اشارہ دینے والا اشارہ۔

شواہد کتنے مضبوط ہیں؟

مرکزی دعویٰ—مختصر-اصطلاح مریض harm کمی ثابت شدہ نہیں—ڈیزائن کے طور پر مضبوط، نتیجہ humble - appropriately preregistered prospective کلسٹر بے ترتیب کنٹرول شدہ آزمائش + blinded مریض-سطح مرکب بینچ مارک/vignette مطالعات سے far مزید - informative

ثانوی دستاویزات/نسخہ نویسی/satisfaction/لاگت نتائج اچھا مگر ثانوی، آلودگی/واحد-نیٹ ورک حدود - apply حفاظت reassuring مگر حد میں۔

بہترین موقف “کام کرتا ہے” یا “failed” نہیں: حقیقی دنیا ٹرائل میں حفاظت اشارہ غائب + عمل بہتر، مگر 2-week مریض فائدہ demonstrate نہیں، چھوٹا فائدہ شناخت کرنے کو larger much مطالعہ چاہیے۔

یہ کیوں اہم ہے

طبی اے آئی بحث بینچ مارک-بھاری ہے، حقیقی pragmatic تصادفی مریض-نتیجہ ٹرائلز نایاب۔ یہ انہی میں سے ہے، اور truth unglamorous دیتی ہے: ٹیسٹ pass کرنا مریض مدد کے برابر نہیں۔

ٹول معالجین کے لیے genuinely مفید ہو سکتی ہے۔ بہتر، cheaper notes، drugs، دوسرا جوڑا of اور—eyes مشکل نتیجہ 14 دن میں نہ حرکت کرے۔ دونوں حقائق together رکھنا بالغ صحت نظام کی کام ہے۔

of burden ثبوت: reset: کلینیکل اے آئی improves نگہداشت دعویٰ کے لیے leaderboard نہیں، مریض نتائج ٹرائل چاہیے—اور محدود فوائد کے لیے بہت بڑا ٹرائل۔

صاف خلاصہ

Kenyan 16 بنیادی-نگہداشت facilities کے کلسٹر بے ترتیب کنٹرول شدہ آزمائش میں “اے آئی” (GPT-4o Consult) پر مبنی کو کلینیکل officers کے EMR میں add کیا گیا۔ 9,691 مریض میں 14-دن ماہر-ماہرین سے جانچا گیا علاج ناکامی 2%، 2 اے آئی بمقابلہ 0%، 2 کنٹرول؛ adjusted OR 77.0، 95% CI 0.13، 55.0، 0.13—P=0.13، 08.1—55.0 CI 95%، 77.0 OR adjusted؛ حفاظت اشارہ نہیں، دستاویزات تمام domains بہتر، نسخہ نویسی/satisfaction تقریباً غیر تبدیل، اینٹی بایوٹک لاگت somewhat کم اشارہ۔ مصنفین: حدود کے اندر محفوظ، علاج ناکامی کم کرنا نہیں، ممکن فائدہ محدود؛ مشاہدہ شدہ اثر شناخت کرنے کو +100k~ مریض پیمانہ چاہیے۔ یہ مریض فائدہ ثابت نہیں کرتا، uselessness بھی نہیں—عمل مدد ہوئی، مریض نتیجہ demonstrably نہیں۔

بلا مبالغہ جانچ

مقالے کا دکھانا ہے: حقیقی دنیا تصادفی ٹرائل میں LLM تائید نے حفاظت اشارہ نہیں دکھایا، دستاویزات بہتر، نسخہ نویسی تقریباً غیر تبدیل، سخت 14-دن علاج-ناکامی مرکب significantly کم کرنا نہیں۔

قرین قیاس مگر ثابت شدہ نہیں: چھوٹا درست مریض فائدہ، underpowered مکمل لاگت پر، savings بہتر دستاویزات eventually بہتر نگہداشت۔

نہیں دکھاتا: مریض نتائج بہتر کرنا، نایاب-واقعہ حفاظت، certification جگہ لیتا ہے/ outperforms معالجین، تمام حالات، generalize economics- established

حدود: نایاب ~2% نتیجہ، کنٹرول آلودگی/ config غلطی + مشترک عادات صفر-bias، ایک نجی شہری نیٹ ورک/ بلند بنیادی موازنہ معیارات، نہیں صوری حفاظت/ noninferiority فریم ورک، 14-دن افق۔

اعتماد: اس آزمائش کے اندر حفاظتی اشارے اور دستاویزات میں بہتری پر مناسب اعتماد، 14 دن کے اندر مریضوں کے واضح فائدے پر بلند اعتماد نہیں، اور "کام کرتا ہے" یا "ناکام ہے" جیسے دو ٹوک فیصلے پر کم اعتماد۔ مناسب موقف: محتاط کلینیکل شواہد، بینچ مارک کی مبالغہ آرائی نہیں۔

ماخذ

(2026) Medicine Nature al., et Agweyu

<https://doi.org/10.1038/s41591-026-04503-6>

202502499779176 registration Registry, Trials Clinical Pan-African

<https://pactr.samrc.ac.za/>

(2026) trial the on release press PATH / EurekAlert

<https://www.eurekalert.org/news-releases/1133583>