Artificial Intelligence in Rural Community Hospitals: How Can it Work for You?

Vinay Prasad
Professor
UC San Francisco

1

One day I was in clinic….

Visiting student rendition

2

Case Presentation

- 60-year-old sister of a 65-year-old patient with multiple myeloma

3

Case Presentation

- 60 year old sister of a 65 year old patient with multiple myeloma
- "Doctor, should I increase the amount of exercise I do to protect against myeloma?"

4

Hell of a question

5

I always think exercise is helpful, and I would do it for general health and well-being and particularly cardiovascular disease.

6

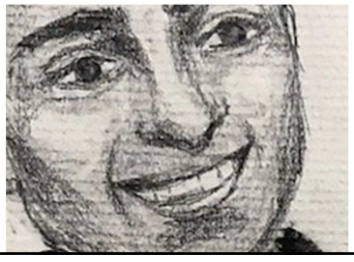
- “Didn’t you read the new study, doctor?”

7

- Didn’t you read the new study?

8

- Didn’t you read the new study?

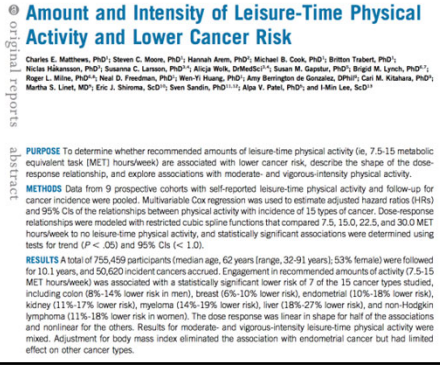


9

- Didn't you read the new study?

10

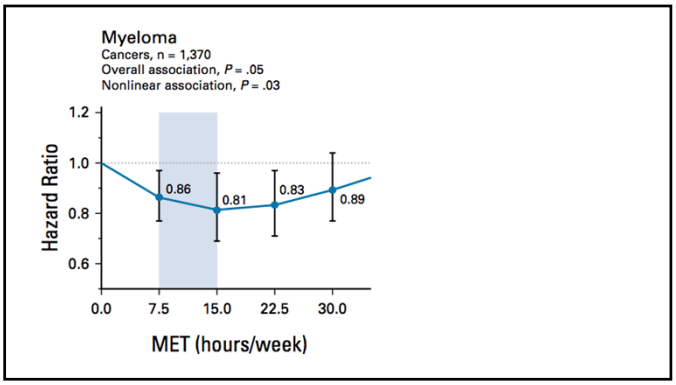
11



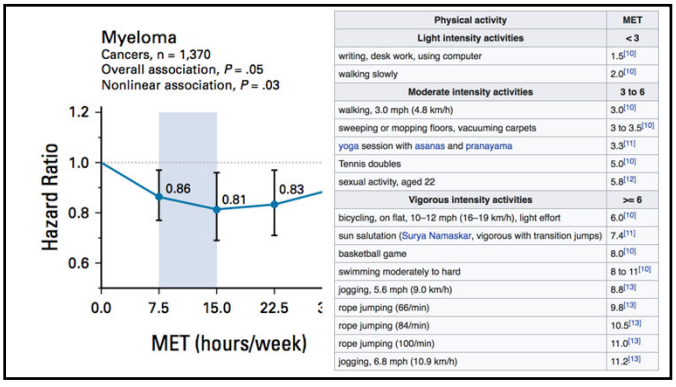
12

statistically significant lower risk of 7 of the 15 cancer types studied,

13



14

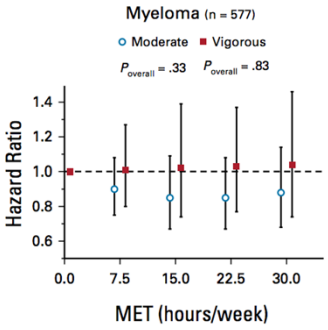


Physical activity	MET
Light intensity activities	**< 3**
writing, desk work, using computer	1.5[10]
walking slowly	2.0[10]
Moderate intensity activities	**3 to 6**
walking, 3.0 mph (4.8 km/h)	3.0[10]
sweeping or mopping floors, vacuuming carpets	3 to 3.5[10]
yoga session with asanas and pranayama	3.3[11]
Tennis doubles	5.0[10]
sexual activity, aged 22	5.8[12]
Vigorous intensity activities	**>= 6**
bicycling, on flat, 10–12 mph (16–19 km/h), light effort	6.0[10]
sun salutation (Surya Namaskar, vigorous with transition jumps)	7.4[11]
basketball game	8.0[10]
swimming moderately to hard	8 to 11[10]
jogging, 5.6 mph (9.0 km/h)	8.8[13]
rope jumping (66/min)	9.8[13]
rope jumping (84/min)	10.5[13]
rope jumping (100/min)	11.0[13]
jogging, 6.8 mph (10.9 km/h)	11.2[13]

15

Physical activity	MET
Light intensity activities	**< 3**
writing, desk work, using computer	1.5[10]
walking slowly	2.0[10]
Moderate intensity activities	**3 to 6**
walking, 3.0 mph (4.8 km/h)	3.0[10]
sweeping or mopping floors, vacuuming carpets	3 to 3.5[10]
yoga session with asanas and pranayama	3.3[11]
Tennis doubles	5.0[10]
sexual activity, aged 22	5.8[12]
bicycling, on flat, 10–12 mph (16–19 km/h), light effort	6.0[10]
sun salutation (Surya Namaskar, vigorous with transition jumps)	7.4[11]
basketball game	8.0[10]
swimming moderately to hard	8 to 11[10]
jogging, 5.6 mph (9.0 km/h)	8.8[13]
rope jumping (66/min)	9.8[13]
rope jumping (84/min)	10.5[13]
rope jumping (100/min)	11.0[13]
jogging, 6.8 mph (10.9 km/h)	11.2[13]

16



17

How did I feel?

18

19

- Implausible finding
- Potential for multiple hypothesis testing
- Confounding – being ill makes you both less likely to exercise and more likely to develop cancer
- Measurement error – self-reported is not true

20

Curiosity, I asked Gemini

21

Should I exercise more to avoid getting multiple myeloma

While exercise is incredibly beneficial for overall health, it's important to understand its relationship with multiple myeloma:

- **Exercise and Multiple Myeloma Risk:**
  - Currently, there's no definitive scientific evidence that increased exercise directly prevents multiple myeloma. The exact causes of multiple myeloma are still being researched.
  - However, maintaining a healthy lifestyle, which includes regular exercise, can contribute to overall well-being and may indirectly influence risk factors.
- **Exercise for People with Multiple Myeloma:**
  - For individuals who already have multiple myeloma, exercise is often recommended to improve quality of life, maintain bone health, and manage symptoms.
  - However, it's crucial to consult with a doctor or physical therapist before starting any exercise program. They can provide personalized recommendations based on your individual condition.

22

How did I feel?

23

- Not bad

24

25

26

AI is gaining traction in our hospitals

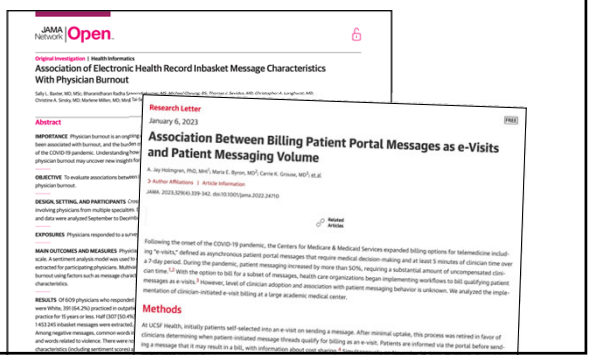
- We shouldn't fear it but embrace it.

27

AI has come of age

- Chat capabilities

28


JAMA Network Open.

Original Investigation | Health Informatics

Association of Electronic Health Record Inbasket Message Characteristics With Physician Burnout

Research Letter

January 6, 2023

Association Between Billing Patient Portal Messages as e-Visits and Patient Messaging Volume

Methods

29


Research

JAMA Internal Medicine | Original Investigation

Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

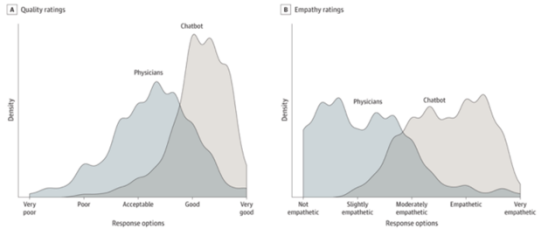
AI for Patient Messages (Ayers et al. JAMA IM 2023)

30

1. Real Messages
2. Real Physician Answers
3. A Patient's Perspective
    1. Quality of response (very poor, poor, acceptable, good, or very good)
    2. Empathy of response (not empathetic, slightly empathetic, moderately empathetic, empathetic, and very empathetic).

31

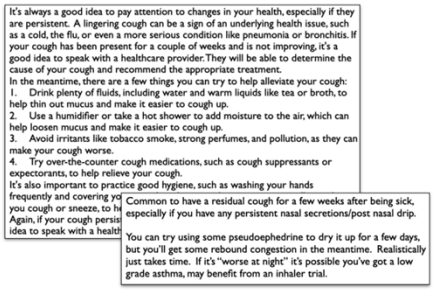
Figure. Distribution of Average Quality and Empathy Ratings for Chatbot and Physician Responses to Patient Questions



Kernel density plots are shown for the average across 3 independent licensed health care professional evaluators using principles of crowd evaluation. A, The overall quality metric is shown. B, The overall empathy metric is shown.

32

AI is not curt



AI for Patient Messages (Ayers et al. JAMA IM 2023)

33

AI has potential to answer medical questions in health care but also to provide comfort.

- Advantage is it does not fatigue, can be verbose and responsive
- Perhaps it may not always be correct, but neither are physicians

34

What about downsides?

- Is it possible we encourage anxiety by letting people ask questions all day?

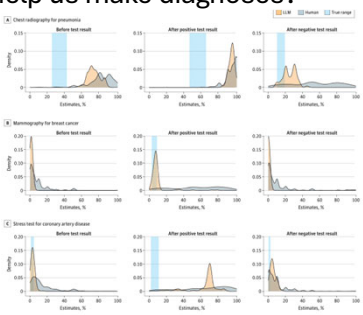
35

One additional role in health care, imo

- To translate the chart into a readable document

36

Will AI help us make diagnoses?



37

Research Letter

April 1, 2024

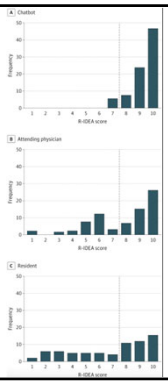
Clinical Reasoning of a Generative Artificial Intelligence Model Compared With Physicians

Stephanie Cabral, MD[1]; Daniel Restrepo, MD[2]; Zahir Kanjee, MD, MPH[1]; et al

38

	I – Interpretive Summary	the patient's main problem(s). The presence or absence of the following features is assessed: a) Key risk factors; b) Chief complaint; c) Illness time course; and d) Use of semantic qualifiers (e.g. monoarticular vs polyarticular) or unified medical concepts (e.g. volume overload, cardiovascular risk factors). NB: Some problems have an implied time course (e.g. syncope, seizure).	1 feature present	1
			2 features present	2
			3 features present	3
			4 features present	4
	D – Differential Diagnosis	Offers more than one relevant diagnostic possibility, committing to what is most likely and considering what is less likely or unlikely yet important to consider for the main chief complaint. If the chief complaint is a diagnosis or syndrome (e.g., acute on chronic systolic heart failure) then differential to rate may be around the differential for that exacerbation (e.g., medication non-compliance vs. arrhythmia).	No differential	0
			Differential is implicitly stated, given as a diagnostic category (e.g., "cardiac"), OR implicitly prioritized	1
			Differential is explicitly stated AND explicitly prioritized	2
	E – Explanation of Lead Diagnosis	Explains the reasoning behind the lead diagnosis, including the epidemiology and key features and how these compare with the patient's presentation. If objective data points are not clearly linked to the lead diagnosis or alternative diagnosis, then only designate points to lead OR alternative diagnosis and NOT both.	No explanation	0
			1 objective data point in explanation of lead diagnosis	1
			≥2 objective data points in explanation of lead diagnosis	2
	A – Alternative Diagnosis Explained	Explains the reasoning behind alternative diagnoses, including the epidemiology and key features and how these compare with	No explanation for any alternative diagnosis	0

39



40

AI in clinical medicine

- Mammography
- Colon polyps
- Knee radiographs

41

- In all these cases, AI is used to ask can I find
  - More invasive ductal adenocarcinoma
  - More polyps
  - Radiographic findings that correlate with pain

42

- In all these cases, AI is used to ask can I find
  - More invasive ductal adenocarcinoma
  - More polyps
  - Radiographic findings that correlate with pain
- AI is not being asked
  - Find lesions that should be cut out
  - Find polyps that should be removed
  - Find radiographic findings that tell me a person will benefit from TKA

43

We are asking AI

- To optimizing finding the problem
- Rather than asking it to find people who benefit from therapies
- A subtle distinction with big implications

44

Knee replacement

ARTICLES
https://doi.org/10.1038/s41591-020-01192-7

nature medicine

Check for updates

An algorithmic approach to reducing unexplained pain disparities in underserved populations

Emma Pierson[1,2], David M. Cutler[3], Jure Leskovec[4], Sendhil Mullainathan[5,6] and Ziad Obermeyer[6]

Underserved populations experience higher levels of pain. These disparities persist even after controlling for the objective severity of diseases like osteoarthritis, as graded by human physicians using medical images, raising the possibility that underserved patients' pain stems from factors external to the knee, such as stress. Here we use a deep learning approach to measure the severity of osteoarthritis, by using knee X-rays to predict patients' experienced pain. We show that this approach dramatically reduces unexplained racial disparities in pain. Relative to standard measures of severity graded by radiologists, which accounted for only 9% (95% confidence interval (CI), 3–16%) of racial disparities in pain, algorithmic predictions accounted for 43% of disparities, or 4.7× more (95% CI, 3.2–11.8×), with similar results for lower-income and less-educated patients. This suggests that much of underserved patients' pain stems from factors within the knee not reflected in standard radiographic measures of severity. We show that the algorithm's ability to reduce unexplained disparities is rooted in the racial and socioeconomic diversity of the training set. Because algorithmic severity measures better capture underserved patients' pain, and severity measures influence treatment decisions, algorithmic predictions could potentially redress disparities in access to treatments like arthroplasty.

45

Kellgren–Lawrence grading scale

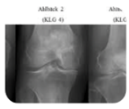
X-Ray					
OA Grade	Grade 0 (Normal)	Grade 1 (Doubtful)	Grade 2 (Mild)	Grade 3 (Moderate)	Grade 4 (Severe)
JSN	No radiographic features of OA are present	Doubtful	Possible	Definite	Marked
Osteophytes		Possible	Definite	Multiple	Large

46

The Kellgren-Lawrence (KLG) score is **a semi-quantitative method for assessing the severity of knee osteoarthritis (OA) on X-rays**:

- **0**: No evidence of OA
- **1**: Possible joint space narrowing and osteophyte formation
- **2**: Definite osteophyte formation and possible joint space narrowing
- **3**: Multiple osteophytes, definite joint space narrowing, sclerosis, and possibly bone deformity
- **4**: End-stage OA, marked by severe sclerosis, joint space narrowing, and large osteophytes

A KLG score of less than 2 is considered mild OA, a score of 2 is considered moderate OA, and a score of greater than 2 is considered severe OA. The KLG score is commonly used in clinical assessment and diagnosis of OA.



47

- Pain due to visual abnormalities in the knee is amenable to knee replacement, pain outside the knee is not.

48

Table 3 | Potential eligibility for surgery: comparing KLG and ALG-P

	Knees potentially eligible for surgery (%)		Knees in severe pain and not eligible for surgery (%)	
	Using KLG	Using ALG-P	Using KLG	Using ALG-P
Black	11% (7%, 15%)	22% (17%, 27%)	51% (45%, 57%)	40% (34%, 46%)
Lower-income	10% (8%, 12%)	13% (10%, 15%)	36% (33%, 40%)	34% (31%, 38%)
Lower-education	9% (7%, 11%)	14% (11%, 16%)	38% (35%, 42%)	33% (30%, 37%)

49

- Pain from 1 to 10 (bad)
- 10 → 1
- 8 → 6

- Who should get surgery?

50

- Pain from 1 to 10 (bad)
- 10 → 10
- 8 → 1

- Who should get surgery?

51

- Pain from 1 to 10 (bad)
- 10 → 6
- 8 → 3

- Who should get surgery?

52

More complicated

- Pain from 1 to 10 (bad)
- 10 → 9 (No surgery), but 10–> 6 (Surgery)
- 8 → 4 (No surgery), but 8 -> 3 (Surgery)

- Who should get surgery?

53

- People who have knee pain which can be *seen* on XR (whether by humans or AI) are more likely to benefit from replacing the knee than those who do not.
- But this is a big untested assumption.
- We have little idea who benefits from replacement.
- It is possible that it isn't something about what's going on in the knee, but what is going on in the entire body, and desire to feel better.
- New method finds people who do WORSE with surgery.

54

Journal of Comparative Effectiveness Research | The Evidence Base

HOME | LATEST ARTICLES | CURRENT ISSUE | ARCHIVE | ABOUT | CONTRIBUTE

OPEN ACCESS | Perspective | 27 February 2023

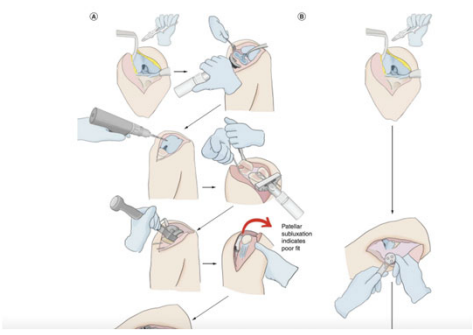
Visualizing the randomized sham-controlled trial in orthopedic research: proposed steps to conducting a total knee arthroplasty randomized controlled trial

Authors: Audrey A Tran and Vinay Prasad | AUTHOR INFO & AFFILIATIONS

Publication: Journal of Comparative Effectiveness Research • Volume 12, Number 3 • https://doi.org/10.57264/cer-2021-0275

PDF

55



56

OPEN

Deep Learning to Improve Breast Cancer Detection on Screening Mammography

Li Shen[1], Laurie R. Margolies[2], Joseph H. Rothstein[3], Eugene Fluder[4], Russell McBride[1,5] & Weiva Sieh[3]

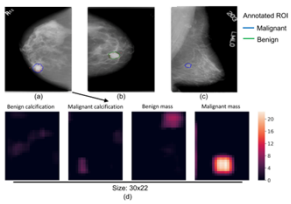


Figure 4. Saliency maps of TP (a), FP (b) and FN (c) image classifications. The outlines represent the regions of interest annotated by the radiologist, and biopsy-confirmed to contain either malignant (blue) or benign (green) tissue. The red dots represent the gradients of the input image with respect to the cancer class output. The gradients were rescaled to be within [0, 1] and a low cutoff of 0.06 was used to remove background noise. Heatmaps (d) of the four non-background classes for input image (a). The colors of the heatmaps represent the activation values after ReLU. This figure is best viewed in color.

57

Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S. Corrado, Ara Darzi, Mozziyar Etemadi, Florencia Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sunny Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, Joseph R. Ledsam, David Melnick, Hormuz Mostofi, Lily Peng, Joshua Jay Reicher, ... Shravya Shetty + Show authors

Nature **577**, 89–94 (2020) | Cite this article

58

clinical setting, we curated a large representative dataset from the UK and a large enriched dataset from the USA. We show an absolute reduction of 5.7% and 1.2% (USA and UK) in false positives and 9.4% and 2.7% in false negatives. We provide evidence of the ability of the system to generalize from the UK to the USA. In an independent study of six radiologists, the AI

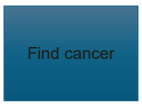
59

Mammography

60

Screening

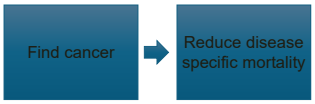
- The purpose of screening



61

Screening

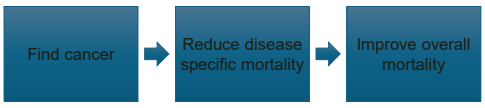
The purpose of screening



62

Screening

The purpose of screening



63

- You can't find cancer that was never going to spread.
- You can't find cancer that has spread already.
- You can't find a cancer that was going to be felt a year later and safety removed them.
- You have to find a cancer that wasn't going to felt in the future, was going to spread, but by finding it now and catching it, you prevented it from spreading.

64

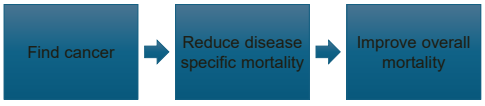
It's all about what we ask AI

- Find more cancer
- Find more lesions that we should cut out today because if we don't, we will miss them until it is too late.*
- *No one knows based on mammography, which lesions these are.

65

Screening

The purpose of screening



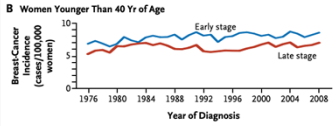
66

ORIGINAL ARTICLE

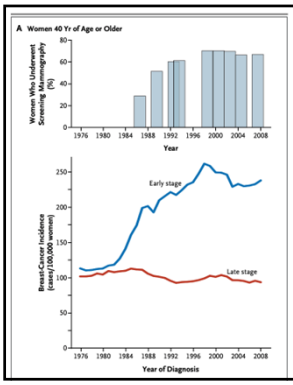
Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence

Archie Bleyer, M.D., and H. Gilbert Welch, M.D., M.P.H.

ABSTRACT



67



68

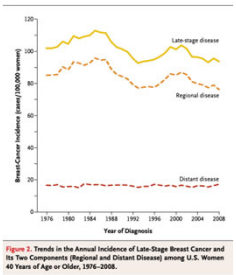


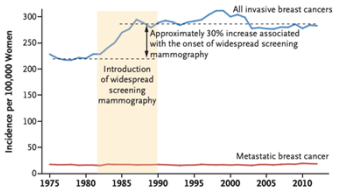
Figure 2. Trends in the Annual Incidence of Late-Stage Breast Cancer and Its Two Components (Regional and Distant Disease) among U.S. Women 40 Years of Age or Older, 1976–2008.

69

ORIGINAL ARTICLE

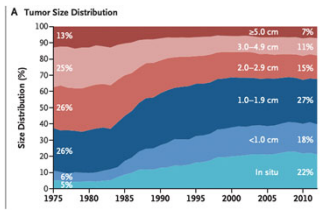
Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness

H. Gilbert Welch, M.D., M.P.H., Philip C. Prorok, Ph.D., A. James O'Malley, Ph.D., and Barnett S. Kramer, M.D., M.P.H.

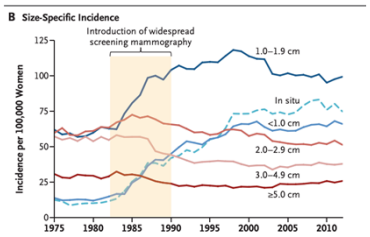


70

I see fewer big tumors in practice



71

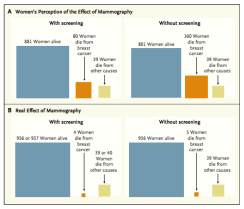


72

Cancer Screening

Abolishing Mammography Screening Programs? A View from the Swiss Medical Board

Nikola Biller-Andorno, M.D., Ph.D., and Peter Jüni, M.D.



U.S. Women's Perceptions of the Effects of Mammography Screening on Breast-Cancer Mortality as Compared with the Actual Effects.

Panel A shows the views of 50-year-old women in the United States regarding the effect of mammography every 2 years on the 10-year risk of death from breast cancer (at left), as compared with no screening (at right). The areas of the squares are proportional to the numbers of women per 1000 who would be alive (blue), die from breast cancer (orange), or die from other causes (yellow). The numbers were calculated from women's perceived relative and absolute risk reductions for breast-cancer deaths (Domenighetti et al.[4]) and U.S. mortality statistics for 2008 from the Centers for Disease Control and Prevention. Panel B shows the actual effect of mammography screening on breast-cancer deaths, with numbers calculated from breast-cancer mortality data for 2008 from

73

Penn Medicine News

News Releases | News Blog | Publications & Special Projects | Internal Newsletters | Subscribe

Home › Internal Newsletters › Presby Bulletin › Testing the Power of AI to Better Detect Colon Polyps

Testing the Power of AI to Better Detect Colon Polyps

March 01, 2023

Recently, physicians from Penn Presbyterian Medical Center (PPMC)'s division of Gastroenterology and Hepatology launched a three-year pilot program to determine if artificial intelligence (AI) technology could improve precancerous polyp detection in screening colonoscopies.

Colorectal cancer is the third-most common cancer in the world and the second-most common cause of cancer deaths. It is also among the most preventable cancers with regular screening colonoscopies, but "even if you are adherent to a standard schedule of screening and surveillance colonoscopies for a lifetime, there is still a small but real five- to nine-percent risk of developing colon cancer because there may be small polyps in the colon that are missed," said gastroenterologist **Neilanjan Nandi, MD, FACP**, who coordinates the new AI platform.

74

- Are you finding more polyps? -- current AI question
- You want to find polyps and precancer that are going to advance and cause morbidity or mortality and not
  - Ones that will be caught on the next screening
  - Ones that have already spread
  - Ones that were never going to cause harm

75

AI can blunder

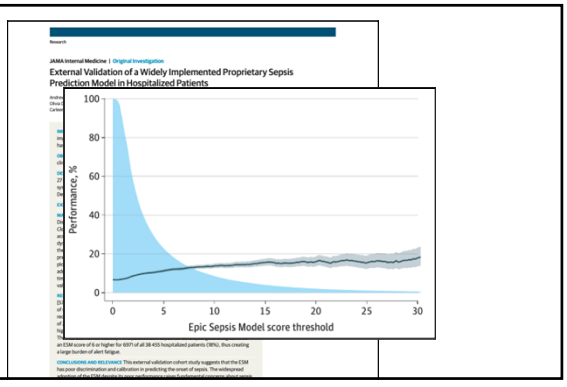
76

Research

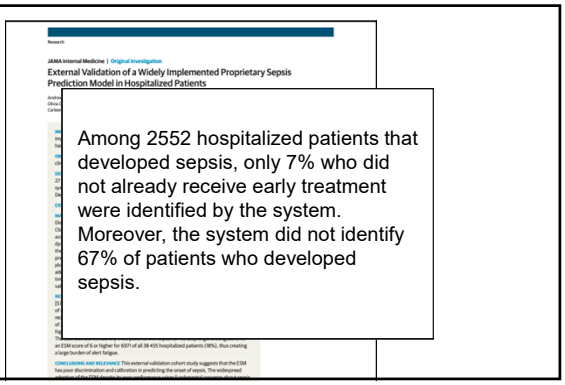
JAMA Internal Medicine | Original Investigation

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

77



78



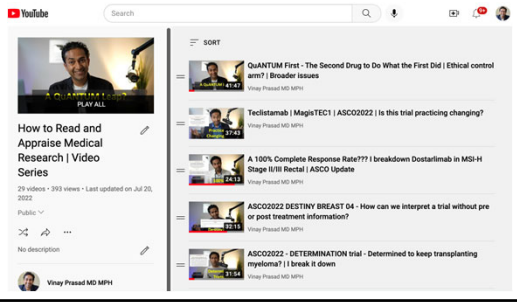
79

AI in rural hospitals

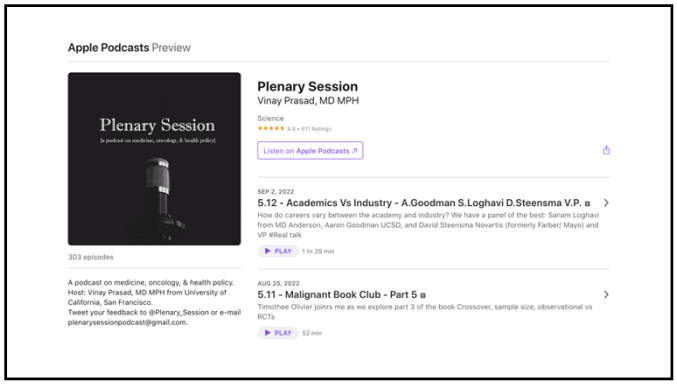
- It can help us manage the paperwork of medicine
- It can help improve our diagnoses
- It can improve quality but also backfire
- It can lead to innovation but also find things we don't want to
- It has great potential, and we should use it wisely

80

Future things to explore if you liked this talk

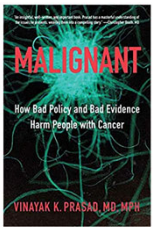


81



82

Questions



- Vinayak.prasad@ucsf.edu

83