



Vocal Recognition of Depression?

Michael E. Johansen, MD, MS, Associate Editor

Ann Fam Med 2025;23:89. <https://doi.org/10.1370/afm.240615>

The *Annals of Family Medicine* encourages readers to develop a learning community to improve health and health care through enhanced primary care. With the *Annals Journal Club*, we **encourage** diverse participants—particularly among students, trainees, residents, and interns—to think critically about and discuss important issues affecting primary care, and even consider how their discussions might inform their practice.

HOW IT WORKS

The *Annals* provides discussion tips and questions related to one original research article in each issue. We welcome you to post a summary of your conversation to our **eLetters section**, a forum for readers to share their responses to *Annals* articles. Further information and links to **previous Annals Journal Club features** can be found on our website.

CURRENT SELECTION

Mazur A, Costantino H, Tom P, Wilson MP, Thompson RG. Evaluation of an AI-based voice biomarker tool to detect signals consistent with moderate to severe depression. *Ann Fam Med*. 2025;23:60-65. doi: [10.1370/afm.240091](https://doi.org/10.1370/afm.240091)

DISCUSSION TIPS

Depression is an under-treated and under-recognized condition. Many patients who are depressed don't seek care, while others are seen in primary care and go unrecognized and untreated. Additional options for screening for depression in primary care could be potentially valuable, especially if this didn't increase the administrative burden on the primary care system. This study investigates a novel method of using voice recordings and a machine learning system to identify patients with a PHQ-9 of >9.

Discussion Questions

- What question is asked by this study and why does it matter?

- How does this study advance beyond previous research and clinical practice on this topic?
- How strong is the study design for answering the question?
- What is the gold standard for a major depressive disorder diagnosis? How is this different than the diagnosis used in this study?
- What are sensitivity, specificity, positive predictive value, and negative predictive value? How are these calculated?
 - What are false negative and false positives? How are these related to sensitivity and specificity?
 - What is spectrum bias? How could this be applicable to this study?
- To what degree can the findings be accounted for by:
 1. How and where people were recruited from for the study?
 2. Who opted to participate in the study?
 3. How data were collected for the study?
 4. How data were determined to be included in the study data set?
 5. Decision on where to make cut points in the data?
 6. How "further evaluation recommended" was dealt with in the calculation of sensitivity and specificity?
- What are the main study findings?
- What is the difference in validation on a portion of the initial sampled population vs a completely different population? Why is this distinction important?
- How comparable is the study sample to patients in your practice? What is your judgment about the transportability of the findings?
- What contextual factors are important for interpreting the findings?
- How might this study change your practice? Policy? Research?
- Who are the constituencies for the findings, and how might they be engaged in interpreting or using the findings?
- What are the next steps in interpreting or applying the findings?
- What researchable questions remain?