



OPEN ACCESS

EDITED BY

Stefano Severi,
University of Bologna, Italy

REVIEWED BY

Laszlo Balkanyi,
University of Pannonia, Hungary

*CORRESPONDENCE

Noam Keidar,
✉ noamkeidar@campus.technion.ac.il
Yael Yaniv,
✉ yaely@bm.technion.ac.il

RECEIVED 07 February 2025

ACCEPTED 31 March 2025

PUBLISHED 09 April 2025

CITATION

Keidar N and Yaniv Y (2025) The prophet's rite of passage – pitfalls in evaluating real-time prediction in medicine.
Front. Physiol. 16:1569008.
doi: 10.3389/fphys.2025.1569008

COPYRIGHT

© 2025 Keidar and Yaniv. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The prophet's rite of passage – pitfalls in evaluating real-time prediction in medicine

Noam Keidar* and Yael Yaniv*

Laboratory of Bioelectric and Bioenergetic Systems, Faculty of Biomedical Engineering, Technion-IIT, Haifa, Israel

KEYWORDS

prediction, atrial fibrillation, epilepsy, clinic, real-time prediction model

Introduction

The future has always captivated human imagination, with efforts to assess disease prognosis dating back to ancient Egyptian times: “*If the heart trembles, has little power and sinks, the disease is advancing ...and death is near ...*” (Papyrus Ebers, circa 1550 BC). However, the risks of relying on predictions were also acknowledged in antiquity: “...The prophecy has been taken from the prophets and given to the fools and babies instead ...” (Babylonian Talmud: Baba Bathra 12b). Recent advancements in medicine have significantly enhanced prognostic accuracy. The availability of comfortable and reliable wearable devices capable of measuring cardiac, neuronal, and other physiological signals, combined with sophisticated machine learning algorithms designed to interpret these signals in real time (Davoodi et al., 2024; Elul et al., 2024; Fira et al., 2024; Kerr et al., 2024), marks the dawn of a new era in preventive medicine—the age of real-time prediction.

Prediction vs. risk factor

Scanning the existing literature reveals considerable semantic confusion surrounding the terms prediction, risk factor identification, detection, and diagnosis, which are often used interchangeably. To clarify, we propose the following distinctions:

Prediction

Definition: “Event X is likely to occur within time interval T for individual P”

Relation: Future-oriented; depends on statistical/machine learning models and longitudinal data.

Example: “This patient has a 30% chance of developing heart failure within the next 5 years.”

Risk factor

Definition: “Individual P is at higher risk for event X due to factor F”

Relation: Identifies modifiable/non-modifiable contributors to an event.

Example: “Hypertension is a risk factor for stroke.”

Detection

Definition: “Individual P is currently experiencing event X.”

Relation: Real-time or near real-time identification of an ongoing event.

Example: “A wearable ECG detects atrial fibrillation in real time.”

Diagnosis

Definition: “Individual P has condition D.”

Relation: Typically involves clinical assessment, imaging, or lab tests.

Example: “The patient has been diagnosed with myocardial infarction based on ECG and biomarker analysis.”

In summary, we can define these conceptual relations:

Prediction → Risk Factor Identification (Prediction models often incorporate risk factors.)

Risk Factor Identification → Prediction (Identified risk factors improve predictive accuracy.)

Detection → Diagnosis (Detected anomalies may trigger diagnostic confirmation.)

Diagnosis → Prediction (Diagnosis may inform predictions about disease progression.)

New era of prediction

Traditionally, preventive medicine has focused on risk factor identification and mitigation, screening tests, patient and caregiver education, and the deployment of response systems for life-threatening events. However, the advent of real-time prediction is transforming this discipline. Imagine a world where non-invasive wearable devices issue alerts before the onset of various conditions, such as life-threatening cardiac arrhythmias, heart failure decompensation, or stroke. These capabilities could pave the way for precisely timed, real preventive therapies, enabling clinicians to achieve optimal outcomes with fewer interventions while preventing serious medical events and their potentially life-threatening consequences.

Such real-time prediction systems are already being developed by research groups and companies, showing promising results (Fira et al., 2024; Kerr et al., 2024). Yet, the true clinical value and significance of these advancements can be challenging to define and fully comprehend.

Alarm goes off

The essence of a real-time prediction is the alarm, an actionable output presented to patients and/or caregivers at the right time, to convey a clear message and facilitate the most appropriate intervention. If the alarm does not warrant a specific action, not accurately timed or has an ambiguous meaning, its clinical utility might be questionable. While these requirements may seem obvious, the usual tools used for evaluation of prediction systems are limited due to unique challenges posed by the continuity of time. To validate that a prediction system has a positive effect on human health, the clinical setting, the intended use, and the continuum of time must be considered, as demonstrated in the following examples.

Unlike a classic predictive test that can be either “positive” or “negative” but not both, an alarm at a certain time may be both or neither. Imagine a system predicting sudden cardiac arrests and activating Emergency Medical Services (EMS), enabling immediate intervention on onset. An alarm 20 s before the event is useful, but considering response times of even the best EMS, no added benefit is gained. Such alerts are indeed true positive as the event

was accurately predicted, but false negative in the sense that it does not support a timely response. Similarly, an alarm set off 5 min after onset is technically false, but not “as false” as an alarm calling the EMS for no reason at all.

Another challenge when testing a prediction system is consideration of the distribution of false or true alarms over time. For example, 30 false alarms within the same hour will mean a single unnecessary EMS activation, but when issued every few days over 4 months, will repeatedly unnecessarily activate EMS, and result in patient anxiety and stress as well as complacency, possibly leading to missed true alarms. On the other hand, even a single timed alarm is enough for intervention, making any “false negative” that relates to the same event meaningless.

How to report real time prediction results?

To illustrate the complexity of the problem, consider the example of predicting out-of-hospital sudden cardiac arrest. The incidence of the predicted event is approximately 55 cases per 100,000 person-years (Berdowski et al., 2010). But many studies showing event prediction results calculate sensitivity, specificity, and accuracy for patient groups with even representation for positive and negative intervals, introducing a selection bias. In our example, even if we consider an entire day before the event as a relevant prediction window, 99% specificity in a 1:1 positive-negative sample population is equivalent to 0.0003% specificity.

Researchers from different scientific fields approach this problem in various ways. Earthquake prediction research commonly uses a predictive ratio (Kagan and Knopoff, 1987) that compares the predicted probability in space and time against a null hypothesis assuming the events are random (Poisson stochastic process). This method is unbiased and rigorous, but sheds little light on the expected clinical utility if adopted to medicine. Arrhythmia predictions for example, usually use samples of positive and negative intervals (e.g., the paroxysmal atrial fibrillation prediction challenge (Moody et al., 2000) uses a 1:1 positive-negative ratio), which provides clear clinical trial metrics (sensitivity, specificity, etc.) but introduces a selection bias. Epileptic seizure prediction studies usually show metrics on a sample of intervals, adding the false alarm rate per hour (Daoud and Bayoumi, 2019), a measure useful for estimation of false alarm burden on patients and healthcare systems, but sensitive to the alarm time-distribution problem described above.

Authors solicited opinion regarding real-time prediction system results

We are of the opinion that introduction of real-time prediction systems to clinical practice requires a usage-centered approach. This would first require definition of the action that patients and caregivers should take once an alarm is raised, e.g., “activate EMS, then observe the patient for 6 h after the alarm”. Once the action is clear, a true positive can be appropriately defined, i.e., for each event, define if the alarm issued at a relevant time for the action is efficient. For example, assuming an 8-min notice is needed to significantly improve patient survival, all events with at least one

alarm between 6 h and 8 min before onset would be considered true positive. Thereafter, sensitivity can be defined as the rate of true positives vs. events. A false positive would be a needless action, e.g., an EMS activation. As the patient is observed for 6 h, this is the minimal allowed gap between two activations; two false alarms 3 h apart would be considered a single false positive alarm. These definitions of true and false positives can then be applied to meaningfully determine the positive predictive value. Alarms not achieving the goal but not causing an irrelevant action, e.g., an alarm 20 s before or during an event, are neither positive nor negative.

Conclusion

We hope that clear definitions, awareness of pitfalls and the proposed approach will help set a benchmark for clinically relevant real-time prediction systems, creating the prophet rite of passage, and retrieving prophecy from the fools.

Author contributions

NK: Investigation, Validation, Writing – original draft, Writing – review and editing. YY: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported

by the Israel Ministry of Science and Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Berdowski, J., Berg, R. A., Tijssen, J. G. P., and Koster, R. W. (2010). Global incidences of out-of-hospital cardiac arrest and survival rates: systematic review of 67 prospective studies. *Resuscitation* 81 (11), 1479–1487. doi:10.1016/j.resuscitation.2010.08.006
- Daoud, H., and Bayoumi, M. A. (2019). Efficient epileptic seizure prediction based on deep learning. *IEEE Trans. Biomed. Circuits Syst.* 13 (5), 804–813. doi:10.1109/TBCAS.2019.2929053
- Davoodi, M., Aspis, N., Drori, Y., Weiser-Bitoun, I., and Yaniv, Y. (2024). LieRHRV system for remote lie detection using heart rate variability parameters. *Sci. Rep.* 14 (1), 30749. doi:10.1038/s41598-024-80480-5
- Elul, Y., Rozenberg, E., Boyarski, A., Yaniv, Y., Schuster, A., and Bronstein, A. M. (2024). Data-driven modeling of interrelated dynamical systems. *Commun. Phys.* 7 (1), 141. doi:10.1038/s42005-024-01626-5
- Fira, M., Costin, H. N., and Goraș, L. (2024). Ventricular fibrillation prediction and detection: a comprehensive review of modern techniques. *Appl. Sci.* 14, 11167. doi:10.3390/app142311167
- Kagan, Y. Y., and Knopoff, L. (1987). Statistical short-term earthquake prediction. *Science* 236 (4808), 1563–1567. doi:10.1126/science.236.4808.1563
- Kerr, W. T., McFarlane, K. N., and Figueiredo, P. G. (2024). The present and future of seizure detection, prediction, and forecasting with machine learning, including the future impact on clinical trials. *Front. Neurol.* 15, 1425490. doi:10.3389/fneur.2024.1425490
- Moody, G., Goldberger, A., McClennen, S., and Swiryn, S. (2000). "Predicting the onset of paroxysmal atrial fibrillation: the Computers in Cardiology Challenge 2001," in *Computers in Cardiology 2001*, Rotterdam, Netherlands, 23–26 September 2001 (IEEE), 215–220.