When to Stop Getting Tested: The Theory of Diagnostic Tests

Extended Abstract

Anson Kahng
University of Rochester
Rochester, NY, USA
anson.kahng@rochester.edu

Joseph Saber
University of Rochester
Rochester, NY, USA
jsaber@ur.rochester.edu

ABSTRACT

We present and analyze a simple model for medical care in which patients may take relatively low-cost medical tests in order to update their beliefs about the probability that they have a particular disease. At any point, patients may leave the testing domain and either risk the effects of the disease or seek expensive treatment. We explore the question of when patients should stop getting tested. In particular, we study two variants of this game in which patients may take an infinite or finite number of tests. We numerically solve for optimal strategies in both the finite and infinite cases, and analytically derive optimal behavior for well-behaved subcases in the infinite setting. We also experimentally explore variants of our basic medical testing model.

KEYWORDS

Algorithmic game theory; repeated games; diagnostic testing

ACM Reference Format:
Anson Kahng and Joseph Saber. 2025. When to Stop Getting Tested: The Theory of Diagnostic Tests: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Modern medicine is advancing at an unprecedented pace, and recent developments demonstrate significant promise to both catch and combat diseases more effectively. However, medicine is only as helpful as patients' abilities to seek and accept care, ideally as early as possible. Especially for patients who live in remote, low-income areas of the world with few medical professionals, the first major hurdle is seeking out a specific diagnosis for an ailment.

There have been significant pushes to streamline and democratize medical testing, and advancements in computer science and artificial intelligence have enabled the use of machine learning algorithms as diagnostic screening tools, supercharging the reach of decentralized medical testing. One particularly promising application is a website, parktest.net, that screens for Parkinson's disease based on audio and visual responses to basic commands [1–6].

However, although these tests drastically improve access to diagnostic screening tools, there is still one major consideration left unaddressed: After receiving results from the test, what should patients now do with this information?

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1.1 Model

We propose a formal model of patient care based on a one-player repeated game. Patients have access to a set of diagnostic tests, and they know the specificity (true positive rate), sensitivity (true negative rate), and cost of each test. In each stage game, a patient has three choices: (1) *test*: take a test, (2) *risk*: exit the system and risk the effects of disease, or (3) *treat*: seek expensive treatment. We assume two knowledge models for patients: in one model, patients know their initial probability of having the disease, and in the other, patients are unaware of this initial value. In both models, we ask: *When should a patient stop getting tested and either seek treatment or exit the system?*

Throughout, we assume that patients are capable of making perfect Bayesian updates given the results of diagnostic tests. Let us discuss the reasonableness of this assumption. While our results hold for patients who are capable of performing perfect Bayesian updates, we do not necessarily require rationality on the part of users. Indeed, we may think of the model as a decision support tool that itself performs Bayesian updates and, based on these updates, advises patients to seek treatment, take additional tests, or do exit the system. Importantly, even in this reframing, it is still up to the patient to accept or reject the mechanism's advice, i.e., they do not completely relinquish their decision-making power to an algorithm.

1.2 Our Contributions

We propose a simple model for medical testing in which patients play a repeated game before making a decision to either leave the system (i.e., take the risk) or seek treatment. We establish many structural lemmas about optimal behavior in this model, particularly for the infinite setting in which patients receive independent results from tests. In the infinite setting, we identify two subcases in which the optimal solution exactly converges in finite time, which we term the *simple* and *semi-simple* settings. We also present experimental results in the finite setting where patients do not receive independent results from tests. Lastly, we explore some extensions of the model that account for smoothed risk and treatment curves, thresholds for treatment, and disease progression over time.

2 EMPIRICAL RESULTS

Our experimental results in the finite setting show a marked improvement over real-world baselines such as "believe the negative," "believe the positive," or taking one test.

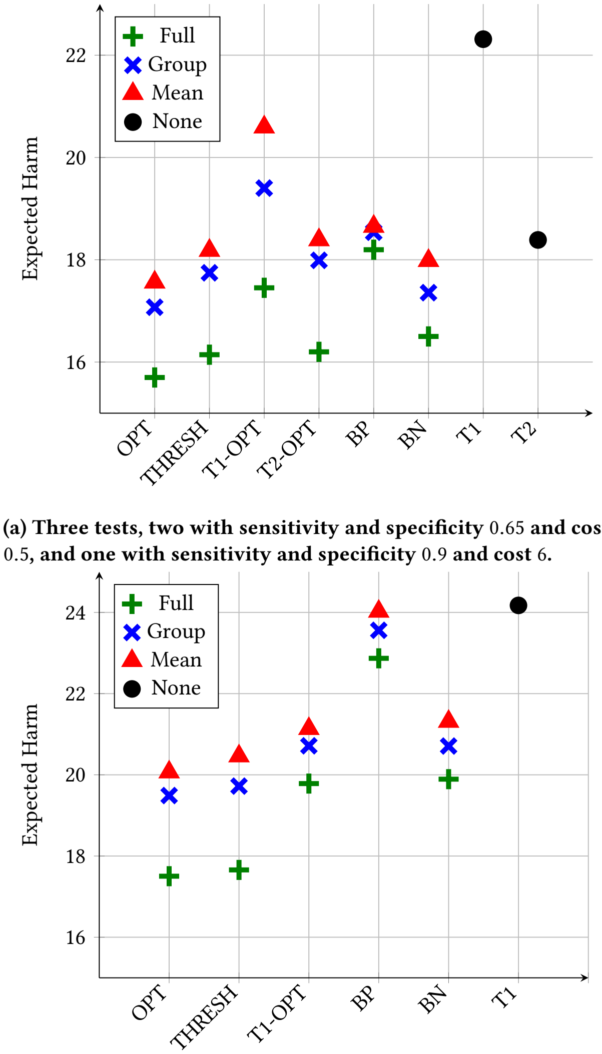
In our simulations, we generate a population of 1,000,000 patients whose probability of disease is drawn from one of two distributions, one with a lower mean and variance representing a low-risk population, and one with a higher mean and variance representing a subpopulation with a higher risk of having the disease. The high-mean distribution is drawn from one sixth of the time, and the low-mean distribution is drawn from otherwise. Additionally, to model factors that affect how much the treatment costs for each patient and how much the disease would affect them, we give each patient different payoffs for *risk* and *treat* drawn from a distribution.

A patient may choose to take any test, risk, or treat according to their strategy, but they may not take the same test more than once. We compare the performance of the optimal (finite) strategy (**OPT**) and a simpler probability threshold strategy (**THRESH**) against a set of baselines. OPT enumerates all possible decision trees and executes the one with the least expected harm. We also examine the performance of OPT restricted to only one of the three tests, labeled **T1-OPT** and **T2-OPT** when restricted to the weak or strong test respectively (tests with identical sensitivity, specificity, and cost are combined). THRESH takes a lower bound probability and an upper bound probability and takes a random remaining test as long as the current probability of disease is in the middle. It will risk if its probability drops below the lower bound and test if it reaches above the upper bound. Finding the optimal thresholds for this strategy is analytically difficult, so we choose as our lower bound the smallest probability at which taking any of the tests is useful assuming it is the only test to be taken, and our upper bound is similarly the greatest probability at which taking any test is useful.

In practice, a patient may not have full knowledge of all the factors affecting their probability of disease. Therefore, we compare the performance of these strategies with different levels of knowledge, which we call **full**, **group**, and **mean**. In the full knowledge scenario, patients know their exact probability of disease. In the group scenario, patients only know whether they were drawn from the low or high risk population, and thus the mean of their subpopulation becomes their expected probability of disease. In the mean scenario, patients only know the entire population's mean probability of disease, and so this becomes their probability.

We compare our strategies with several baselines. One of our baselines is the option of simply taking one of the tests, risking if it comes back negative, and treating if it comes back positive. These are listed as **T1** and **T2** for the weak and strong test. The "Believe Positive" (**BP**) strategy takes a predetermined sequence of tests and seeks treatment if any come back positive, and only risks if all are negative. "Believe Negative" (**BN**) is similar but reversed. We consider these as part of our "no knowledge" scenario as these strategies do not depend on the patients' individual risk cost R, treatment cost T, and probability of disease. Because the performance of BP and BN is greatly affected by which set of tests is taken and in what order, we assume each patient chooses the optimal subset and sequence of tests despite the fact that in the no knowledge scenario they would not have access to the information that lets them make this choice. This has the effect of making these baselines as strong as possible.

Our results are presented in Figure 1. In our first scenario in Figure 1a, patients have access to two different weak but inexpensive tests, and a strong but expensive test. The second scenario in Figure 1b consists of three different weak and inexpensive tests. In both scenarios, OPT with full knowledge outperforms all other strategies, while the threshold strategy with full knowledge is not far behind. As expected, performance degrades with less information. BN performs quite well as a baseline while BP performs somewhat poorly. This is because the population overall has a rather low probability of disease, so it makes more sense to perform a riskier strategy. In the scenario with only weak tests, neither baseline performs as well. The improvements from using multiple tests is greater in the second scenario with multiple weak tests.



(a) Three tests, two with sensitivity and specificity 0.65 and cost 0.5, and one with sensitivity and specificity 0.9 and cost 6.

(b) Three tests: sensitivity and specificity 0.65, cost 0.5.

Figure 1: Expected harm of various strategies under different information conditions. Black dots are baselines.

3 DISCUSSION

We view our model as a first step toward a more comprehensive and useful theory of diagnostic testing in a new age of widely-available, low-cost distributed medical testing made possible by advances in technology. Please see the full version of the paper for our theoretical contributions (including useful structural lemmas for optimal solutions and closed-form solutions in special cases) and proofs, as well as additional experimental results.

REFERENCES

[1] Mohammad Rafayet Ali, Javier Hernandez, E Ray Dorsey, Ehsan Hoque, and Daniel McDuff. 2020. Spatio-temporal attention and magnification for classification of Parkinson's disease from videos collected via the internet. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 207–214.

[2] Mohammad Rafayet Ali, Taylan Sen, Qianyi Li, Raina Langevin, Taylor Myers, E Ray Dorsey, Saloni Sharma, and Ehsan Hoque. 2021. Analyzing head pose in remotely collected videos of people with parkinson's disease. *ACM Transactions on Computing for Healthcare* 2, 4 (2021), 1–13.

[3] Kurtis G Haut, Adira Blumenthal, Sarah Atterbury, Xiaofei Zhoul, Wasifur Rahman, Emanuela Natali, M Rafayet Ali, and Ehsan Hoque. 2022. Assistive Video Filters for People with Parkinson's Disease to Remove Tremors and Adjust Voice. In *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.

[4] Md Saiful Islam, Wasifur Rahman, Abdelrahman Abdelkader, Sangwu Lee, Phillip T Yang, Jennifer Lynn Purks, Jamie Lynn Adams, Ruth B Schneider, Earl Ray Dorsey, and Ehsan Hoque. 2023. Using AI to measure Parkinson's disease severity at home. *npj Digital Medicine* 6, 1 (2023), 156.

[5] Raina Langevin, Mohammad Rafayet Ali, Taylan Sen, Christopher Snyder, Taylor Myers, E Ray Dorsey, and Mohammed Ehsan Hoque. 2019. The PARK framework for automated analysis of Parkinson's disease characteristics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–22.

[6] Wasifur Rahman, Sangwu Lee, Md Saiful Islam, Victor Nikhil Antony, Harshil Ratnu, Mohammad Rafayet Ali, Abdullah Al Mamun, Ellen Wagner, Stella Jensen-Roberts, Emma Waddell, et al. 2021. Detecting parkinson disease using a web-based speech task: Observational study. *Journal of medical Internet research* 23, 10 (2021), e26305.