Hyfe

WHITEPAPER:

The Primacy of Observation Duration Over Event Accuracy in Cough Monitoring

Simulation evidence supporting continuous, automated monitoring as a superior approach for clinical research.

The Hyfe Team

October 2025

hyfe.com

Abstract

Accurately quantifying cough frequency is essential in many contexts: understanding respiratory disease burden, monitoring treatment response, and conducting clinical research. While the “gold standard” of monitoring cough through human annotation may achieve near-perfect accuracy, the practice is limited to very short recording durations due to logistical, financial, and privacy concerns. This paper argues that extended monitoring with an automated system - such as Hyfe’s wearable cough monitor - yields more accurate estimates of an individual’s true cough rate than short-term (typically 24 hours) recordings from a perfect device, even if that automated system does not achieve perfect accuracy. Using simulated data and theoretical considerations, we show that the dominant source of error in cough quantification arises not from device misclassification, but from the sampling error inherent in brief observation windows.

Hyfe

Table of Contents

Introduction

The intuitive assumption in measurement science is that higher device accuracy always leads to better data. Yet, in practice, perfect [1] measurement systems are often limited by cost, comfort, privacy, and practicality. In cough monitoring, the most accurate systems can be cumbersome, obtrusive, and costly, with a maximum practical use period of just a day.

By contrast, automated solutions like Hyfe's Cough Monitor are less precise on a per-event basis but can be worn comfortably and continuously for weeks or even months. This enables continuous data collection across the natural variability of daily cough frequency. The key question, therefore, is not simply *which monitor is more accurate*, but *which approach better estimates an individual's underlying cough rate* over clinically relevant timeframes.

.

Theoretical Framework

Sources of Error

Two fundamental sources of error affect cough rate estimation:

1. **Classification error** – arising from missed detections (false negatives) and incorrect detections (false positives).
2. **Sampling error** – arising from observing only a short portion of a highly variable time series.

A "perfect" monitor minimizes classification error but suffers from sampling error when used briefly (ie, only 24 hours). "Imperfect" long-term monitors

[1] There is no "perfect" measurement of cough, as even the "gold standard" of human annotation is known to have a non-zero rate of inter and intra labeler discrepancies (ie, one expert annotator thinks a sound is a cough; another thinks it's a throat clear).

reduce sampling error substantially, generally at the cost of higher rates of classification error associated with automation.

The Human Observer Thought Experiment

Imagine a perfect monitor: a human observer watching a subject for one hour, counting 27 coughs perfectly. Extrapolating to a week yields 4,536 coughs. Yet, if the subject's cough frequency fluctuates daily, the true weekly total will differ. The observation was perfectly accurate for one hour—but a poor estimate of the week.

This demonstrates that in real-world behavior monitoring, **duration often outweighs per-event accuracy,** especially for a biological phenomenon as highly variable as cough rate.

Simulation Design

We simulated a cohort of 1,000 patients, each with an underlying *true* daily cough rate. The simulation was based on empirically observed cough rates (both absolute magnitude and variability) amongst patients with chronic coughs. The average hourly cough rate in the cohort was 32 per hour with an IQR of 25.5 to 36.6. The "truth" was defined as each individual's 30-day mean cough count (30 being considered sufficient to capture the patient's true cough rate).

We then compared two monitoring strategies:

- **Perfect monitor (human listener)**: 100% sensitivity and 0% false positives.
- **Imperfect monitor (Hyfe)**: Average 85% sensitivity and approximately average two false positives per hour (false positives were injected into the simulated data using the same distribution as the empirically observed false positive distribution amongst patients in Hyfe's

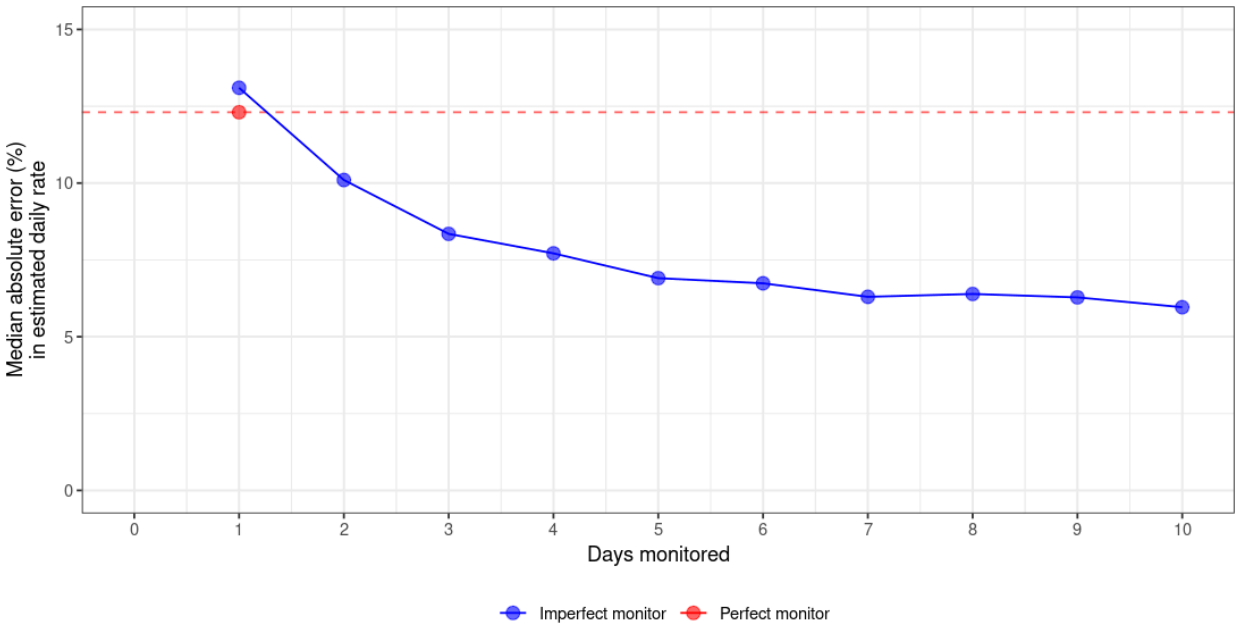
validation trial).

Each system was used to estimate the 30-day average from shorter monitoring durations ranging from one to thirty days. The resulting estimates were compared to the true rate, and the **absolute median percentage error (MAPE)** was calculated for each configuration.

Results

The Cost of Short Monitoring Windows

Using a perfect monitor (human labeling) for 30 days would hypothetically yield zero error but is operationally impossible –patients wouldn't agree to such an intrusion into their privacy, nor would study sponsors be able to absorb the cost of 720 person-hours of human annotations for each participant. Reducing the window saves money and makes studies operationally feasible, but introduces increasing error. Using our simulated data, the estimates generated by one day of "perfect" monitoring show ~13% median absolute error in estimating the true 30-day rate (below chart, red line). An imperfect monitor (below chart, blue line) is less accurate than the perfect monitor if only used for one day; but if used for multiple days, the median absolute error rate reduces such that it more closely approximates the truth than the perfect monitor.



Imperfect but Persistent

The imperfect monitor introduces classification error but benefits from extended observation. Over 20 days, Hyfe's estimated error stabilizes around 6%. Despite its per-event imperfections, prolonged monitoring cuts the total error in half compared to a one-day perfect monitor. Most of this improvement occurs within the first week of continuous use.

Discussion

The results highlight a crucial insight: **duration is a more powerful determinant of accuracy than per-event precision in highly variable physiological signals like cough**. Continuous, unobtrusive monitoring captures the temporal variability that short snapshots miss.

The implications are significant for clinical trials, disease monitoring, and digital biomarker validation. A practical, wearable, automated system—even if imperfect—offers truer estimates of patient behavior over time.

Conclusions

Hyfe's approach to cough monitoring—continuous, comfortable, and scalable—achieves *more accurate estimates* of real-world cough frequency than "perfect" but short-term systems. **One week of Hyfe data** halves the error of one day of perfect data.

In short, **we do not need perfect accuracy to measure cough optimally—only enough accuracy sustained for a long enough time**.