scientific reports

Check for updates

OPEN

Validation and accuracy of the Hyfe cough monitoring system: a multicenter clinical study

Carlos Chaccour[1,2,3]✉, Isabel Sánchez-Olivieri[2], Sarah Siegel[4], Gina Megson[4], Kevin L. Winthrop[4], Juan Berto Botella[2], Juan P. de-Torres[2,5], Lola Jover[6], Joe Brew[6], George Kafentzis[6,7], Mindaugas Galvosas[6], Matthew Rudd[6,8] & Peter Small[6,9]

Background: The ability to passively and continuously monitor coughing for prolonged periods of time would significantly improve cough management and research. To date there is no automated clinically validated cough monitor that can be routinely used in clinical care and research. Here we describe the validation of such an automated cough monitor. Methods: This multicenter observational study compared the results of the Hyfe CoughMonitor wrist-worn device with manually counted coughs in subjects with a variety of etiologies as they went about their usual daily activities. We collected 24 h of continuous sounds from subjects while they simultaneously wore a CoughMonitor and an audio recorder. Coughs were labelled by multiple trained annotators who listened to the continuous audio recordings using validated methodology. The time stamps of these human-detected coughs were compared to those of the CoughMonitor to determine the system's overall performance using event-to-event and hourly rate correlation analyses. Results: Over the 546 h monitored, 4,454 cough events were recorded; The overall sensitivity was 90.4% (95% CI of 88.3–92.2%). The overall false positive rate was 1.03 false positives per hour (95% CI of 0.84 to 1.24). The overall correlation between manual and CoughMonitor measured hourly coughing was high (Pearson correlation coefficient of 0.99). Two case studies of long-term monitoring of patients with chronic cough are presented. Conclusion: The present analysis of cough events demonstrated that the Hyfe CoughMonitor accurately reflects them with a high sensitivity and a low false positive rate. Future studies should focus on its potential role in the management of patients with cough in clinical practice.

Registration Clinicaltrials.gov, NCT05723159.

Cough is one of the most common reasons for which patients seek medical care, it is associated with a broad range of medical conditions and greatly contributes to healthcare expenditure all over the world[1,2]. For some diseases, such as COPD and COVID-19, cough rates also help to predict adverse outcomes[3,4]. However, in an era when symptom quantification drives refinement in diagnosis and precision in therapy[5], cough is currently not measured as part of clinical practice. This is not because of lack of interest, as efforts to quantify cough date back to the 1950's[6,7].

Improvements in acoustic signal processing and machine learning techniques have fostered renewed attention to contactless fully automated cough monitoring[8–12]. Objective, prolonged cough monitoring can provide a valuable data stream for the diagnosis, prognosis, assessment of treatment response and even syndromic surveillance of respiratory diseases as well as for the development of novel therapeutics. However, there are currently no validated, unobtrusive, and fully-automated cough monitoring systems that are commonly used to monitor cough as patients go about their normal activities[13].

Validation of such systems relies upon having accurate ground truth cough data as gold standard. There is consensus that manual cough counting from audio recordings, albeit laborious, can achieve good interobserver agreement[13]. When manual labeling cough acoustic data, the cough second (a second containing at least one explosive phase of a cough) is a unit of annotation that correlates well with true cough rates and reduces the ambiguity associated with the use of individual explosive phases or cough epochs[14,15].

[1]ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain. [2]Navarra University Clinic, Pamplona, Spain. [3]CIBERINFEC, Madrid, Spain. [4]Oregon Health Science University, Portland, OR, USA. [5]IDisNa, Pamplona, Spain. [6]Hyfe, Wilmington, DE, USA. [7]Department of Computer Science, University of Crete, Heraklion, Greece. [8]University of the South, Sewanee, TN, USA. [9]University of Washington, Seattle, WA, USA. ✉email: carlos.chaccour@isglobal.org

The primary objective of the study was to assess the overall performance of the Hyfe CoughMonitor (Hyfe Inc., 2022), when used by individuals with problematic cough, in comparison to the gold standard of manual cough annotation. Additionally, we also compared the CoughMonitor's performance during the daytime versus nighttime, between individuals and as a function of cough rates.

Methods

The Hyfe CoughMonitor

The CoughMonitor App runs on an Android Smartwatch (Shenszhen Domino Times Smart 4G Watch, Model DM20). In brief, this app uses the watch's microphone to continuously capture and encrypt ambient sounds in a manner that cannot be replayed and is deleted after processing, thus ensuring that sound recordings are secure and transient. An artificial intelligence (AI) based algorithm was designed and implemented in CoughMonitor to detect cough sounds from raw audio recordings. A short description follows next. First, an acoustic event detector looks for onset acoustic events that are similar to the explosive phase of a typical cough sound. Second, once such events have been detected, they are segmented into 0.5-second chunks, transformed into an image-like representation that illustrates the acoustic energy distribution of the segmented sound over time and frequency, and passed to a convolutional neural network (CNN), a popular machine learning model originating from image processing and computer vision. Finally, the neural network – trained on *millions* of coughs and cough-like sound segments – decides whether the input audio corresponds to a cough sound or not, and in the former case, a timestamp is generated. All processing happens in-device. While charging, timestamps and cough durations are transmitted via Wi-Fi to the Hyfe cloud, where the time of each cough is converted to cough seconds. In this study, all uploading was handled by study personnel at the end of the monitoring period of each participant. The same version of the cough detection software was used throughout this study (version 1.0.0).

Continuous sounds were recorded using the same Android smartwatch (Shenszhen Domino Times Smart 4G Watch, Model DM20), but running custom software that continuously recorded all ambient sounds.

Study design

This is a multicenter observational study of individuals with problematic cough due to a variety of cough related conditions designed to assess the overall performance of the Hyfe CoughMonitor System in comparison to manually counted cough events.

Enrollment and eligibility

Individuals of both sexes who had problematic coughs consulting at two clinical sites between March 17 and Nov 7, 2023: (1) Oregon Health & Science University (OHSU) in the US, and (2) University Clinic of Navarra in Spain A third group of participants was enrolled remotely in a decentralized manner in the US through targeted outreach. Inclusion and exclusion criteria are shown in Table 1.

Sample size

Hourly cough counts are necessarily non-negative integers and follow negative binomial distributions closely[16], neither these counts nor any simple transformations thereof are normally distributed, precluding application of standard formulas for the SEs of the Pearson correlation or linear regression coefficients. A simulation based on data collected in previous studies[8,9,15,17] to estimate the sampling distribution of hourly cough rates and showed that a minimum of 18 participants contributing 20 h of monitoring each, yielding 360 paired person-hours for analysis, would result in a standard error under 0.1 in average correlations and slopes (see online data supplement for the study protocol). This target was later expanded to 23 participants in order to have over 50% of the sample recruited in the US for regulatory purposes.

Data acquisition

After obtaining informed consent, research subjects were instructed to wear two devices: (i) the Hyfe CoughMonitor on one wrist and (ii) a second identical watch running custom software that continuously recorded all ambient sounds on the other wrist.

At bedtime, subjects were instructed to charge both devices at bedside. The exact start and stop time of monitoring was recorded by study personnel. Subjects were instructed to monitor for 24 h and write down the times they went to bed, awoke and any times they had to leave the watch aside such as while showering. Participants who recorded for less than 20 h were excluded from analysis.

Inclusion criteria	Exclusion criteria
1. Age 21 years old or above.	1. Inability to accept the privacy policy and terms of use of CoughMonitor due to confidentiality or other concerns.
2. Individuals who express concern about their active cough (problematic cough).	2. Inability to avoid unusually prolonged loud environments for the duration of the 24-hour study period.
3. Anticipate that they can collect auditory recordings and keep the devices with them continuously for 24 h.	3. Need to conduct confidential conversations during the 24-hour monitoring period.
4. Residing in a domestic environment without unusually high and / or persistent background sound levels.	4. Individuals who have had significant change in antitussive therapy in the week preceding study.
5. Willing to wear a watch and audio recorder and keep them at bedside (within 3 ft from the mouth) during the night.	

Table 1. Study inclusion and exclusion criteria.

Cough annotation

To obtain the gold standard, the exact start time of each cough was manually annotated on the continuous recordings as previously described[3,15,17]. In brief, continuous audio recordings were broken into 60 s segments and presented to two trained annotators using a proprietary audio playback, visualization and data entry program. Each 60 s segment was listened to by the annotators independently and blinded to one another´s results. Every cough, throat clear, sneeze or cough-like sounds was labeled as a segment, from its beginning to its end, noting if the sound was distant or unclear. Sounds for which the two labelers disagreed on the presence of a cough, the timing of a cough's start by greater than 100 milliseconds, or the indication that the cough sound was distant or unclear were adjudicated by a third expert trained annotator. The adjudicator was aware of the discrepant annotations and listened to each 60s audio segment containing discrepancies in full. The interobserver variability of this system has been quantified as negligible with an inter-labeler Pearson´s correlation of 0.96 [15].

Event classification

The timestamps of automatically (CoughMonitor) and manually (human annotated) detected cough events were described and compared statistically. Each timestamp was converted to cough seconds, defined as a second during which at least one individual cough occurs[14]. As previously shown, cough seconds can be interchangeable units with explosive phases as units of cough[14] and a more consistent annotation metric[15]. We used the following event definitions:

- True Positive: a cough detected by the CoughMonitor matching human annotators within a 0.5 s margin.
- False Positive: a "cough" detected by the CoughMonitor but either (a) not labeled by humans or (b) labeled by humans as a different sound i.e., throat clear, sneeze or (c) not matching coughs labeled by humans within 0.5 s.

Statistical analysis

Results were analyzed on the basis of cough seconds in two complementary ways: (1) an event-to-event comparison of each cough second yielding the CoughMonitor's sensitivity, false positive rate and positive predictive value, and (2) a correlation analysis of hourly cough rates comparing human annotators and the monitor presented as a Bland-Altman plot and, in the online data supplement, as linear plots.

All statistical analyses were performed in R (R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria).

Cough event based analysis

To calculate the CoughMonitor`s sensitivity, false positive rate and positive predictive value (PPV) as measures of its success at detecting individual cough seconds, we compared (i) ground truth annotations (converted to cough seconds) with (ii) cough seconds as per the CoughMonitor per hour. When a cough second was detected by the CoughMonitor within 0.5 s of a human labeled cough, the event was considered to be matched and therefore a "true positive". Following a match, both the CoughMonitor cough second and human labeled cough second are removed from eligibility for further matching (i.e., if a human-labeled cough fell within 0.5 s of the CoughMonitor, the human-labeled cough could only match with one CoughMonitor cough), meaning that only one event is considered a true positive; the remaining, unmatched CoughMonitor event would be considered a false positive despite it being within 0.5 s of a human labeled cough. A "false positive" was considered to be a CoughMonitor detection which was not matched to a human labeled cough. By comparison with ground truth annotations, each CoughMonitor timestamp is therefore either a true positive or a false positive. Sensitivity and the false positive rate were then calculated according to the usual formulas,

$$\text{CoughMonitor Sensitivity (\%)} = \frac{Total\ number\ of\ true\ positives}{Total\ number\ of\ cough\ seconds},$$

$$\text{False positive rate (/hr)} = \frac{Total\ number\ of\ false\ positives}{Total\ number\ of\ hours\ of\ monitoring}.$$

$$\text{Positive predictive value (\%)} = \frac{Total\ number\ of\ true\ positives}{Total\ number\ of\ true\ and\ false\ positives}$$

Cough rate based analysis

To evaluate the CoughMonitor's overall performance, its hourly cough second counts were compared, person-hour by person-hour, with the gold standard hourly cough second counts obtained from trained human annotators. The agreement between these paired person-hour counts and the ideal model $y = x + \epsilon$, where ϵ denotes an integer-valued random error term with mean 0 and constant variance, was assessed with both a linear analysis and a Bland-Altman analysis.

To account for within-subject correlations, the clustered bootstrap was used to calculate all of these confidence intervals, with each cluster consisting of one individual's paired person-hour counts.

Secondary analyses

We then conducted three pre-specified secondary analyses comparing the CoughMonitor's performance (1) between day and nighttime, using a two-sample t-test (2) between each individual, using descriptive metrics and (3) as a function of cough rate by means of one-way analyses of the variance.

There was no missing data in this study.

Ethics approval and registration

The study was approved by the OHSU Institutional Review Board (number 24749), the Ethics Committee for Medical Research in Navarra (number PI_2022/101), and by WCG IRB for the distributed trial (number

20232553). All procedures were performed in accordance with relevant named guidelines and regulations. Informed consent was obtained from all participants. The study was prospectively registered in Clinicaltrials.gov (NCT05723159) on February 03, 2023.

Results

Enrollment

A total of 28 participants with problematic coughs due to a variety of etiologies were recruited. A battery misconfiguration resulted in the device turning off in the first 3 subjects who were enrolled and this misconfiguration was corrected after the 3rd participant, the detection algorithm was not changed. No subsequent technical interruptions were observed. Two participants inadvertently prematurely terminated monitoring prior to 20 h and were thus excluded from analysis (Fig. 1). There were no adverse events related to the use of CoughMonitor.

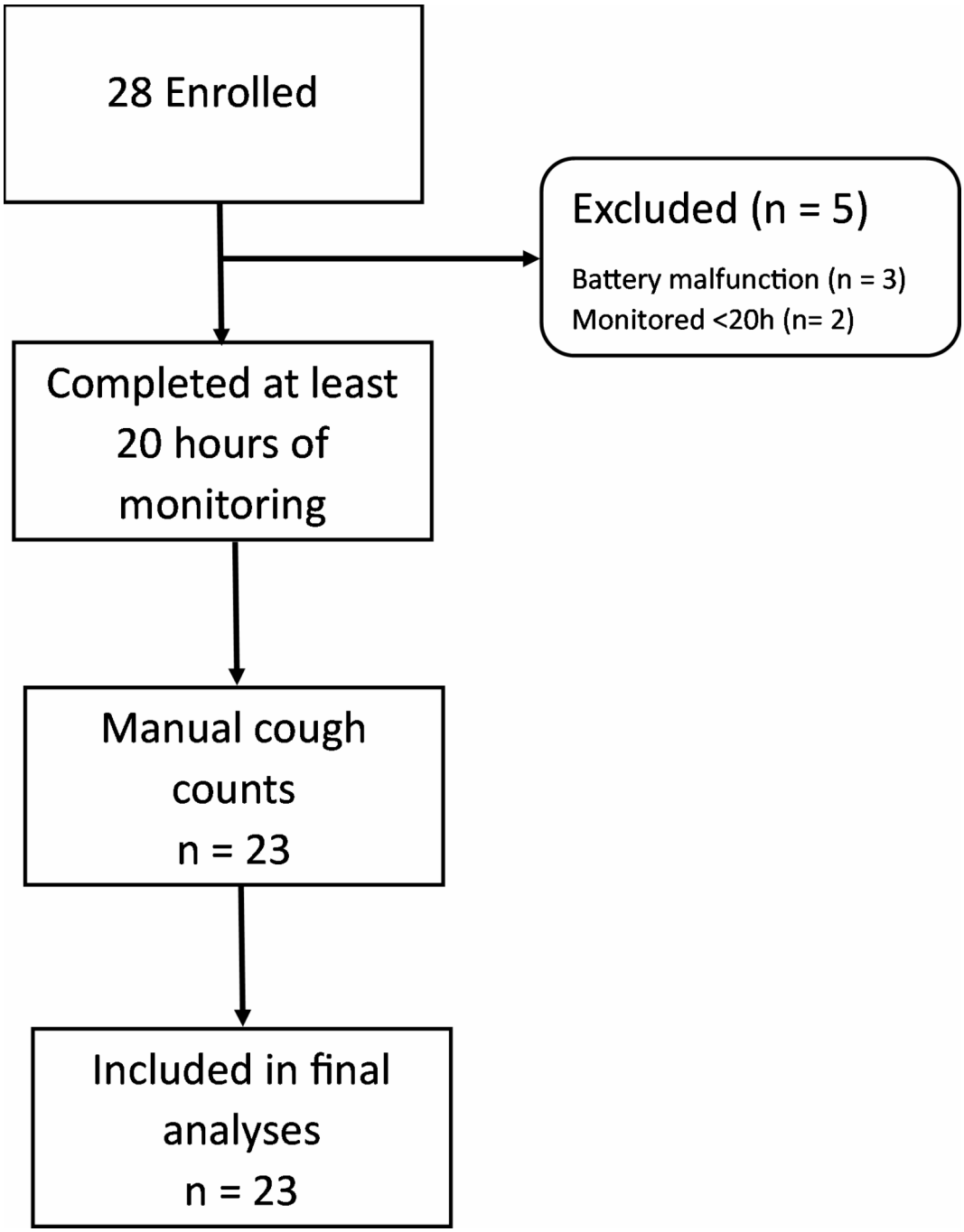

Fig. 1. Enrollment flow diagram of study.

The final analysis included data from 23 subjects (Table 2). Almost two thirds were female (15, 65%), the mean age was 52 (range 24 to 72 years), and 12 (52%) were recruited in the US (5 from Oregon Health Science University and 7 from the dispersed clinical trial). The mean monitoring time was 23.8 h (range 22.9 to 25.64) and the mean hourly cough rate annotated by humans was 8.1 (range 1.5–32). All diagnoses are provided in Table 2 (see **Table S2** in the supplementary material).

Continuous recordings

A total of 546 h of continuous audio/monitoring time containing 4,454 cough seconds was captured from all participants. The mean number of cough seconds by participant was 200 (range 36 to 821). The mean hourly cough rate was 8.15 (range 1.5 to 32).

It was apparent to annotators listening to the continuous recordings that most monitoring occurred in subjects' homes and workplaces. However, monitoring also occurred in a wide variety of acoustic environments such as grocery stores, restaurants, in vehicles, and even while listening to loud techno music.

Cough event based performance results

The overall sensitivity was 90.4% (95% CI of 88.3–92.2%). The overall false positive rate was 1.03 false positives per hour (95% CI of 0.84 to 1.24). The overall positive predictive value was 87.5% (95% CI of 81.9–91.6%). Figure 2 shows the cumulative number of cough seconds detected by the CoughMonitor and the human annotators. The sex, age and diagnosis of the participants had no impact on performance (Figure S1, S2 and S3 in the online data supplement).

The sensitivity rate for individual subjects ranged from 78.1 to 96.5% with false positivity rates ranging from 0.38 to 2.23 cough seconds per hour (Fig. 3, Figure S4).

Cough rate based performance results

CoughMonitor performance based on hourly cough rate is shown in a standard Bland-Altman plot in Fig. 4, providing a visual assessment of the agreement between the hourly cough rates as measured by the CoughMonitor and by the human annotators. We calculated the upper and lower 95% limits of agreement (LOAs), and also computed the 95% confidence interval for the upper and lower COAs using cluster bootstrap resampling techniques. As shown by the dashed horizontal lines, the overall bias (mean difference) is 0.23 (95% CI -0.039 to 0.51), the lower limit of agreement (LOA) is -3.7 (95% CI -5.2 to -3), and the upper limit of agreement is 4.8 (95% CI 4 to 6).

Linear cough rate performance results

A linear validation analysis assumes that the paired person-hour counts exhibit a linear relationship. Figure 5 below confirms this assumption visually, and the overall Pearson correlation coefficient of 0.99 (95% CI 0.962 to 0.996) quantifies the strength of this linear relationship. While having a Pearson correlation coefficient close to +1 is a necessary criterion for satisfactory performance, this is insufficient alone, as it only implies that the paired counts cluster tightly about some line; the coefficients of the regression line (the dashed black line in Fig. 5) measure how close this line is to the ideal line $y = x$ (the solid black line in Fig. 5). The slope and the intercept of the ordinary least squares (OLS) line of best fit in Fig. 5 are 0.94 (95% CI 0.91 to 0.97) and 0.74 (95% CI 0.50 to 0.99), respectively, indicating good agreement and hence satisfactory performance.

Because the results reported by the Hyfe CoughMonitor are expressed as hourly coughing rates, we calculated the correlation based on all 477 of the hours monitored. As shown in Fig. 3, the overall linear correlation was high, with a Pearson correlation coefficient of 0.99, an OLS slope of 0.94, and an OLS intercept of 0.74.

Characteristics	Value
Age, mean (range)	52 years (24–72)
Hours monitored, mean (range)	23.8 h (22.9–25.6)
Hourly cough rate, mean (range)	8.15 (1.5–32)
Female, n (%)	15 (65%)
US-based, n (%)	12 (52%)
Diagnoses, n (%)	
Bronchitis	6 (26%)
NTM pulmonary disease	5 (22%)
RCC/UCC	3 (13%)
Acute respiratory syndrome	2 (9%)
COPD	2 (9%)
COVID-19	2 (9%)
Bronchiectasis	1 (4%)
Asthma	1 (4%)
Cystic fibrosis	1 (4%)

Table 2. Demographic and cough monitoring results of study subjects. NTM: nontuberculous mycobacteria. RCC/UCC: refractory/unexplained chronic cough.

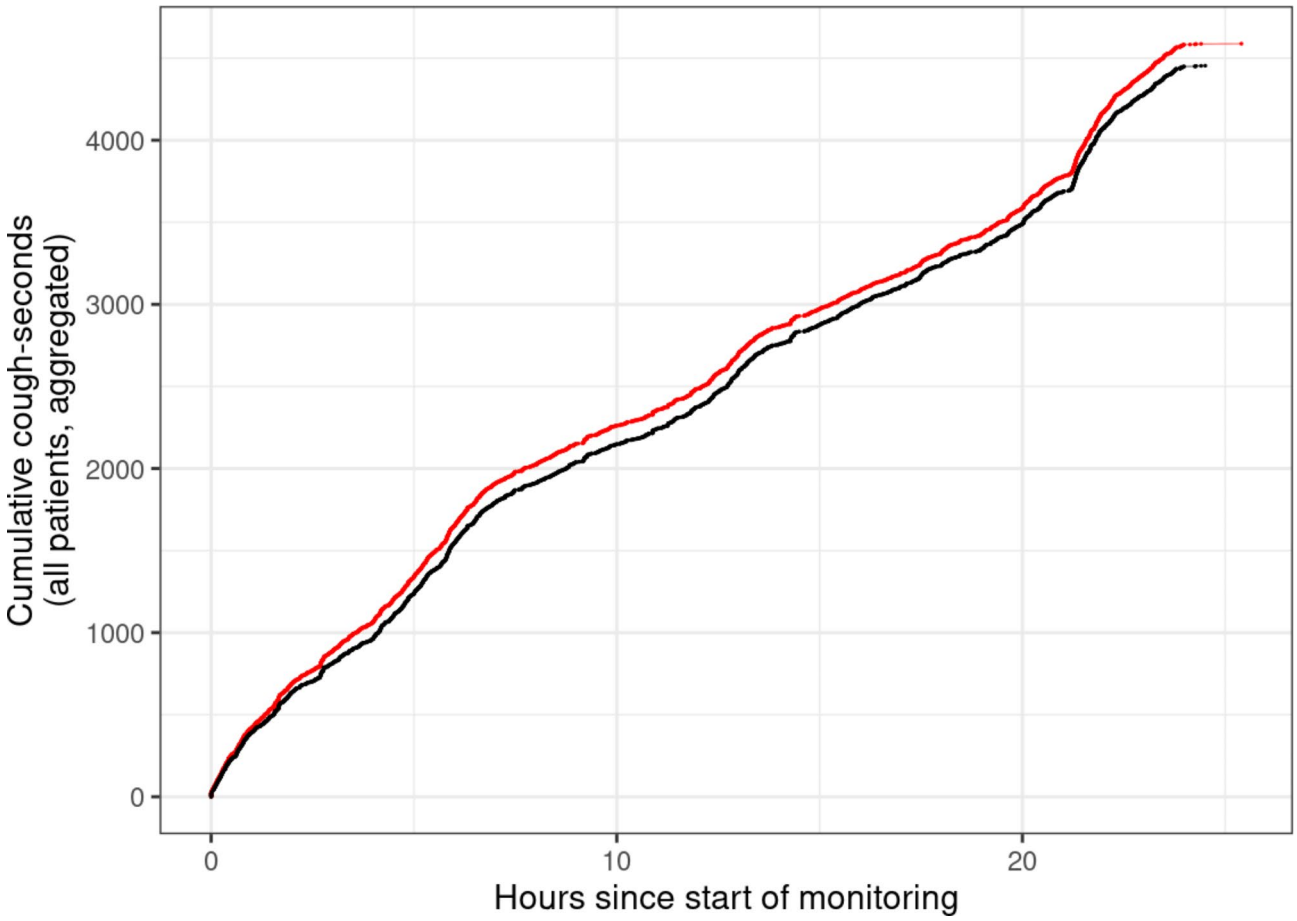

Fig. 2. Cumulative cough event-based performance results. Cumulative cough seconds from all participants over the course of the study with time. The ground truth human counts are shown in black and the Hyfe results are shown in red.

Discussion

We evaluated the Hyfe CoughMonitor in over 546 h of continuous audio containing 4,454 coughs from 23 participants with a broad range of cough-causing diagnoses. This resulted in a sensitivity of 90.4% and an overall false positive rate was 1.03 false positives per hour. Thus, the monitor is unique in its ability to continuously monitor cough frequency in a manner that is unobtrusive, fully automated, and privacy preserving as users go about their usual activities of daily living. These attributes promise to improve patients' understanding of their cough and its triggers, as well as to improve the diagnosis and management of disease. Furthermore, continuous cough monitoring addresses some of the fundamental statistical inadequacies of short-term cough monitoring that have limited clinical trials of new antitussive drugs as recently raised by the US Food and Drug Administration[18,19].

Using cough seconds as unit of analysis and applying a previously described annotation protocol, has proven accurate to establish the ground truth, showing an intra-labeler disagreement of less than two per hour monitored; (Pearson's correlation 0.98) and inter-labeler agreement with a Pearson's correlation of 0.96 [15]. In addition to showing high reproducibility, this analysis establishes the error rate of even the most rigorous human annotation with implications for their use in trials and as comparators when validating automated monitors.

Sensitivity is an essential metric of event-level performance, but detecting or failing to detect a single cough is not clinically meaningful. Hourly cough rates, on the other hand, are highly informative to patients, providers, and investigators, the Bland-Altman plots reported here show the robustness of hourly rates determined using CoughMonitor.

Finally, we illustrate the potential clinical value of unobtrusive, continuous cough monitoring for long periods of time spanning over six months in two different clinical cases.

There are several limitations to this study. First, the results were obtained with a single make and model of Android watch which could limit its generalizability to other devices. However, in the recent literature resulting from a sharp increase in the general interest in cough recognizing algorithms, comparable results have been obtained using a variety of devices that employ comparable microphones and chip sets[20,21]. Second, in this study we have made only limited efforts to distinguish between coughs from the wearer and others in close proximity. Thus, the device should not be used in environments with a high burden of non-user coughs until such an analysis has been conducted. Third, cough recognition will be influenced by the acoustic environment in which

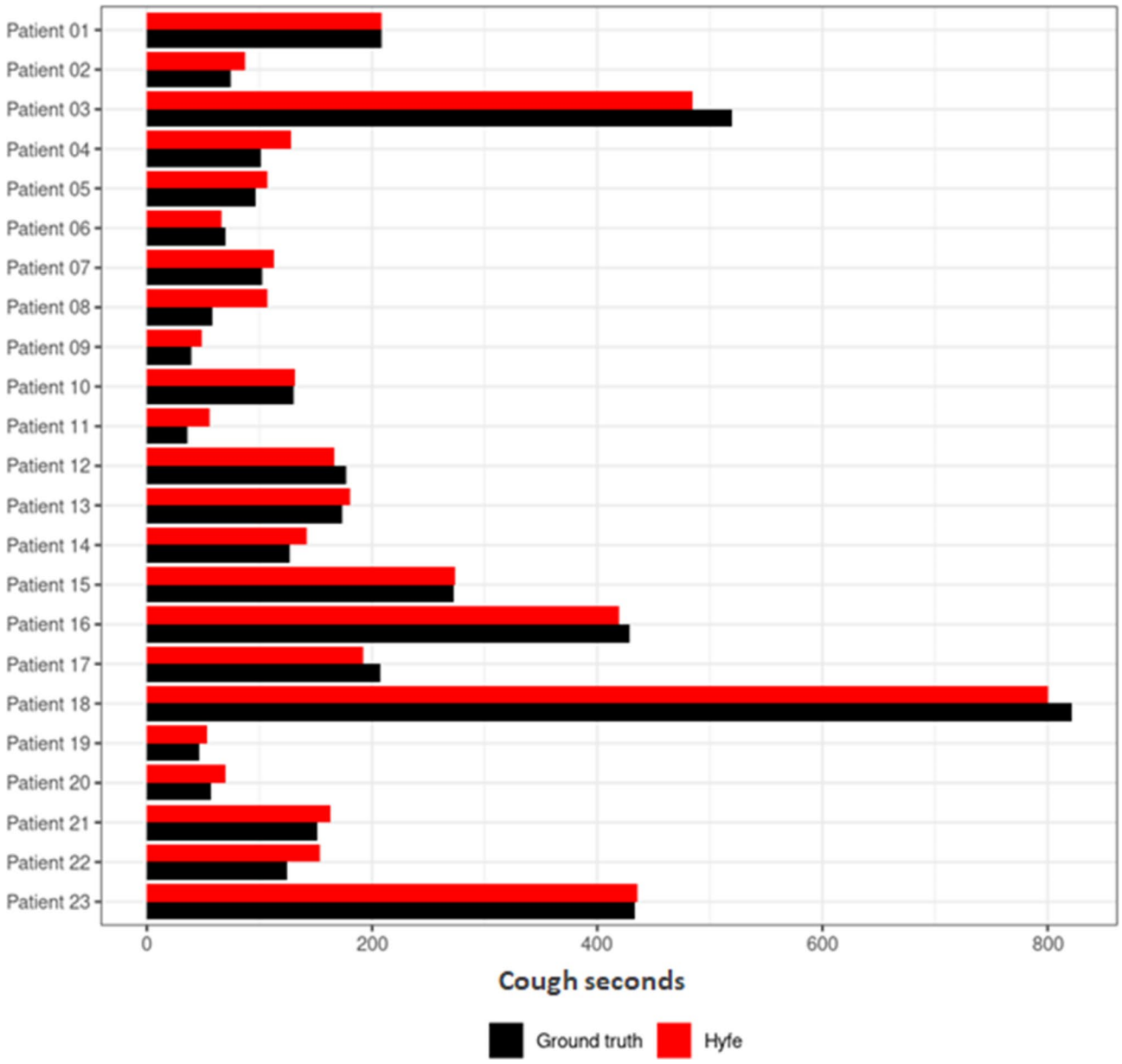

Fig. 3. Individual performance results.

the CoughMonitor is worn, as a large volume of peaks may artificially inflate the detected cough second rate, even with a low false positivity rate. Although subjects were instructed to avoid unusually loud environments, they were actually worn in a wide variety of settings such as while using public transport and listening to loud techno music. Nonetheless, users must be informed that results may be less accurate in settings with extremely loud background noise. Finally, an overall false positivity of one per hour means that the accuracy will be better among subjects with higher cough burden, this aligns well with clinical practice among subjects with chronic cough and for evaluating the efficacy of drugs for this indication but if there is intention to use the device among subjects with relatively low cough rates, appropriately powered studies in different acoustic environments should be conducted beforehand.

Conclusion

The Hyfe CoughMonitor System has an overall sensitivity for detecting a cough event above 90% and a false positivity rate of about one per hour. These results were observed in men and women with a variety of diagnoses as individuals went about their usual activities of daily living. Given the system's high accuracy, usability, and scalability, it has the potential to greatly improve clinical care, drug development and regulatory efforts, particularly among subjects with a high cough burden.

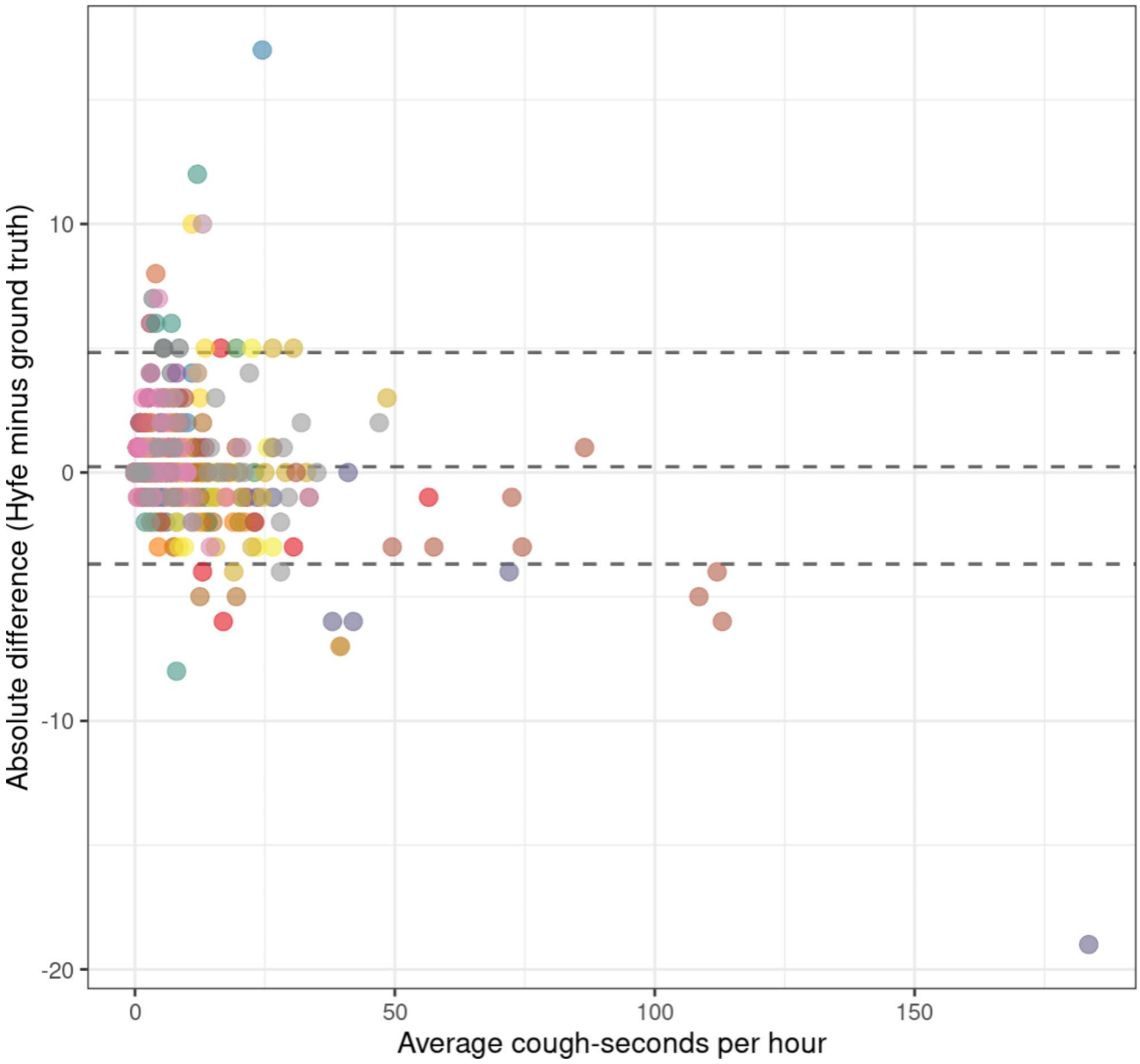

Fig. 4. Cough rate-based performance results (Bland-Altman plot). Each point is one person-hour; its x-coordinate is the average of its manual and CoughMonitor hourly cough second counts, and its y-coordinate is their difference. The dashed horizontal lines indicate the bias (mean difference) of 0.23, the lower limit of agreement of -3.7, and the upper limit of agreement of 4.8.

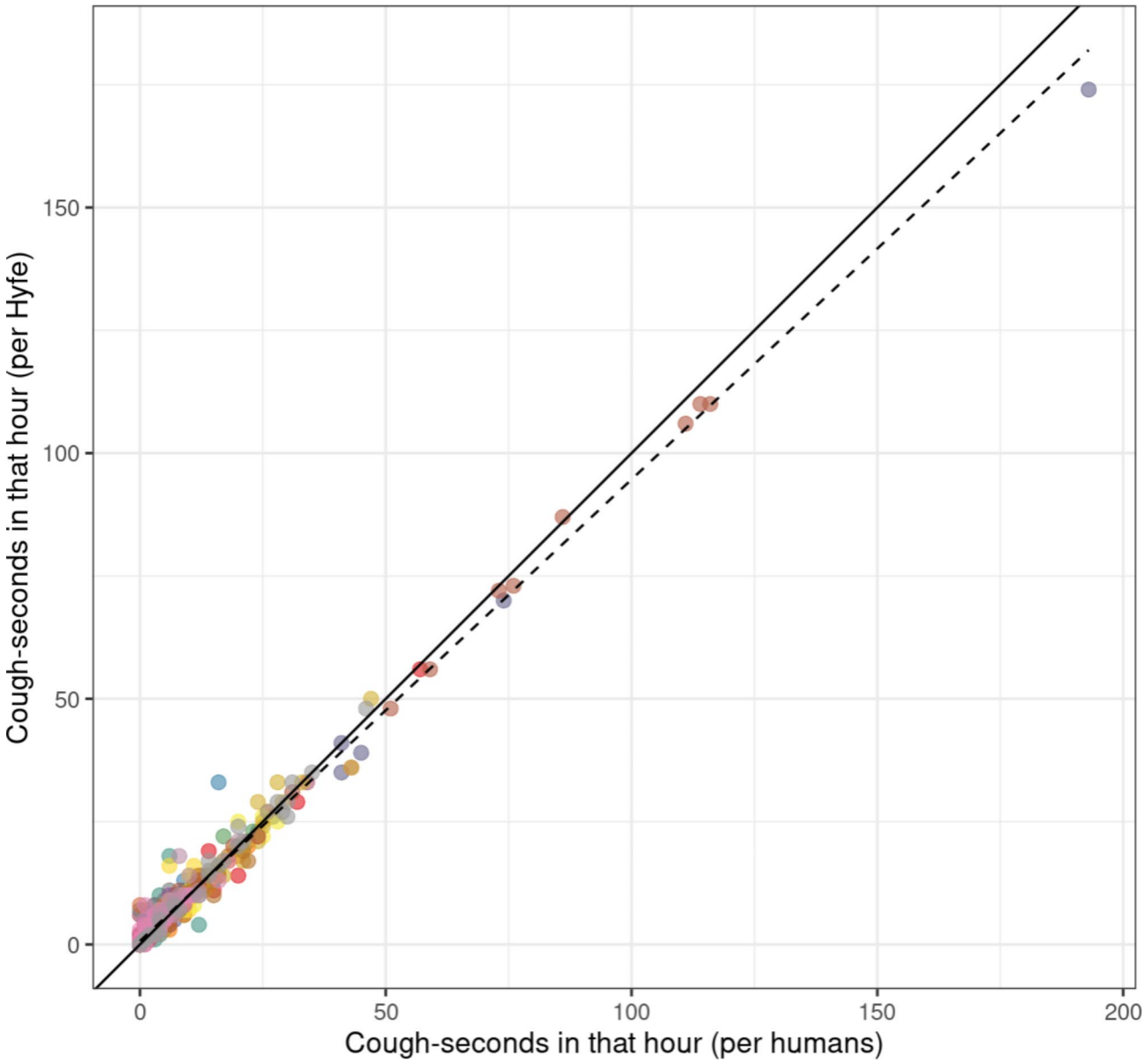

Fig. 5. Cough rate-based performance results (linear plot). Each point is one person-hour, the black dashed line is the OLS line of best fit, and the black solid line is the line of perfect agreement ($y = x$).

Data availability

Anonymized individual data and code relevant to this article will be publicly available in GitHub (https://github.com/hyfe-ai/validation/) as well as study protocol, immediately following publication, indefinitely. Further assistance with data access can be obtained from the corresponding author (CCh). CCh, ISO, MR and KW had full access to all of the data in this study and take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Received: 3 June 2024; Accepted: 2 January 2025
Published online: 06 January 2025

References

1. Cornford, C. S. Why patients consult when they cough: A comparison of consulting and non-consulting patients. *Br. J. Gen. Pract.* **48**, 1751–1754 (1998).
2. Irwin, R. S. et al. Overview of the management of cough: CHEST Guideline and Expert Panel Report. *Chest* **146**, 885–889. https://doi.org/10.1378/chest.14-1485 (2014).
3. Altshuler, E. et al. Digital cough monitoring - A potential predictive acoustic biomarker of clinical outcomes in hospitalized COVID-19 patients. *J. Biomed. Inf.* **138**, 104283. https://doi.org/10.1016/j.jbi.2023.104283 (2023).
4. Crooks, M. G. et al. Domiciliary cough monitoring for the prediction of COPD exacerbations. *Lung* **199**, 131–137. https://doi.org/10.1007/s00408-021-00435-9 (2021).

5. Gomes, B. & Ashley, E. A. Artificial intelligence in molecular medicine. *N. Engl. J. Med.* **388**, 2456–2465. https://doi.org/10.1056/NEJMra2204787 (2023).
6. Hall, J. I., Lozano, M., Estrada-Petrocelli, L., Birring, S. & Turner, R. The present and future of cough counting tools. *J. Thorac. Dis.* **12**, 5207–5223. https://doi.org/10.21037/jtd-2020-icc-003 (2020).
7. Loudon, R. G. & Spohn, S. K. Cough frequency and infectivity in patients with pulmonary tuberculosis. *Am. Rev. Respir. Dis.* **99**, 109–111. https://doi.org/10.1164/arrd.1969.99.1.109 (1969).
8. Gabaldon-Figueira, J. C. et al. Acoustic surveillance of cough for detecting respiratory disease using artificial intelligence. *ERJ Open Res.* **8**. https://doi.org/10.1183/23120541.00053-2022 (2022).
9. Gabaldon-Figueira, J. C. et al. Longitudinal passive cough monitoring and its implications for detecting changes in clinical status. *ERJ Open Res.* **8**. https://doi.org/10.1183/23120541.00001-2022 (2022).
10. Huddart, S. et al. Continuous cough monitoring: A novel digital biomarker for TB diagnosis and treatment response monitoring. *Int. J. Tuberc. Lung Dis.* **27**, 221–222. https://doi.org/10.5588/ijtld.22.0511 (2023).
11. Kang, Y. R. et al. Long-COVID severe refractory cough: Discussion of a case with 6-week longitudinal cough characterization. *Asia Pac. Allergy.* **12**, e19. https://doi.org/10.5415/apallergy.2022.12.e19 (2022).
12. Zimmer, A. J. et al. Making cough count in tuberculosis care. *Commun. Med. (Lond).* **2**, 83. https://doi.org/10.1038/s43856-022-00149-w (2022).
13. Morice, A. H. et al. ERS guidelines on the assessment of cough. *Eur. Respir. J.* **29**, 1256–1276. https://doi.org/10.1183/09031936.00101006 (2007).
14. Kelsall, A. et al. How to quantify coughing: correlations with quality of life in chronic cough. *Eur. Respir. J.* **32**, 175–179. https://doi.org/10.1183/09031936.00101307 (2008).
15. Sanchez-Olivieri, I. et al. Performance evaluation of human cough annotators: Optimal metrics and sex differences. *BMJ Open Respir. Res.* **10**. https://doi.org/10.1136/bmjresp-2023-001942 (2023).
16. Rudd, M., Song, W. J. & Small, P. M. The statistics of counting coughs: Easy as 1, 2, 3? *Lung* **200**, 531–537 (2022). https://doi.org/10.1007/s00408-022-00559-6
17. Galvosas, M. G. F. et al. C. Performance evaluation of the smartphone-based AI cough monitoring app - Hyfe Cough Tracker against solicited respiratory sounds [version 1; peer review: 1 approved with reservations, 1 not approved]. *F1000Research* **11**, 730 (2022). https://doi.org/10.12688/f1000research.122597.1
18. McGarvey, L. P. et al. Efficacy and safety of gefapixant, a P2X(3) receptor antagonist, in refractory chronic cough and unexplained chronic cough (COUGH-1 and COUGH-2): results from two double-blind, randomised, parallel-group, placebo-controlled, phase 3 trials. *Lancet* **399**, 909–923. https://doi.org/10.1016/S0140-6736(21)02348-5 (2022).
19. FDA. Pulmonary-Allergy Drugs Advisory Committee (PADAC) meeting on Gefapixant November 17. Full hearing. https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-17-2023-meeting-pulmonary-allergy-drugs-advisory-committee-11172023. Accessed Feb 2024 (2023).
20. Hegde, S. et al. Cough sounds in screening and diagnostics: A scoping review. *Laryngoscope* https://doi.org/10.1002/lary.31042 (2023).
21. Serrurier, A., Neuschaefer-Rube, C. & Rohrig, R. Past and trends in cough sound acquisition, automatic detection and automatic classification: A comparative review. *Sensor (Basel).* **22** https://doi.org/10.3390/s22082896 (2022).

Acknowledgements

The authors would like to thank the participants and study team for their efforts to make this research happen.

Author contributions

Conceptualisation: CCh, JB, PSData curation: CCh, ISO, SS, LJFormal analysis: MR, JBInvestigation: CCh, ISO, SS, GM, KW, JPDT, JB, GK, PS Methodology: CCh, MR, JB, PSSupervision: CCh, SS, PSWriting - original draft: CCh, PSWriting - review & editing: all authors contributed, reviewed and approved the last draft.

Funding

This study was funded by Hyfe Inc. The sponsor participated in study design, collection and analysis; it had no role in the decision to submit this manuscript for publication. ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the "Centro de Excelencia Severo Ochoa 2019–2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA program.

Declarations

Competing interests

MR, MG, LJ, MG, JB, GK and PS are employees of Hyfe, Inc and own equity in Hyfe Inc. CCh has received consultancy fees and owns equity in Hyfe Inc. All other authors declare no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-85341-3.

Correspondence and requests for materials should be addressed to C.C.