Acoustic cough monitoring in the management of patients with known respiratory disease

NCT05042063

Date of protocol: 04 October, 2021 Version 4.0

Intended resgistry: clinicaltrials.gov

Sponsor: Clínica Universidad de Navarra/Universidad de Navarra

Principal investigator: Dr.: Carlos Chaccour

Department: Area of Infectious Diseases

Clínica Universidad de Navarra

PROTOCOL SIGNATURES SHEET

Acoustic cough monitoring in the management of patients with known respiratory disease

I have read this protocol and accept the obligation to direct this trial in accordance with all the stipulations of the protocol and with the Helsinki Declaration.

Dr. Carlos Chaccour		04/10/21
Principal investigator	Signature	Date

Dr. Juan Carlos Gabaldón		04/10/2021
Author/Trial coordinator	Signature	Date

1. GENERAL INFORMATION

1.1. Identification of trial

Protocol code: Hyfe-clinical

Title of trial: **Acoustic cough monitoring in the management of patients with known respiratory disease**

1.2. Details of sponsor

CLÍNICA UNIVERSIDAD DE NAVARRA / UNIVERSIDAD DE NAVARRA

Avenida de Pío XII, 36

31008 Pamplona

Email: ucicec@unav.es

1.3. Coordinating investigator or principal

Dr. Carlos Chaccour

Department: Area of Infectious Diseases

Clínica Universidad de Navarra

Avda. Pío XII 36

31008 Pamplona

Tel.: 948 255 400

1.4. Details of trial investigators

STUDY COORDINATOR: Juan Carlos Gabaldon

INSTITUTION: University of Navarra

PHONE: +34 674 52 12 31

EMAIL: jgabaldonfi@unav.es

CO-INVESTIGATOR: Dr. Simon Grandjean Lapierre

INSTITUTION: Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CR-CHUM)

PHONE: +001 514 743 7255

+001 514 890 8000 ext. 20935

EMAIL: simon.grandjean.lapierre@umontreal.ca

1.5. Centers where study takes place

Clínica Universidad de Navarra (CUN)

1.6. Expected duration of trial

The study will last up to five years.

2. BACKGROUND AND JUSTIFICATION OF THE STUDY

Respiratory diseases are a leading cause of death in both the developed and underdeveloped world, with conditions such as tuberculosis (TB), Chronic Obstructive Pulmonary Disease (COPD), lung cancer and pneumonia collectively causing over 10 million deaths globally every year (Mathers et al. 2009). Furthermore, as of April 2021, the COVID-19 pandemic has affected 133 million people and killed 2.9 million (John Hopkins 2021).

Cough is a cardinal symptom of all these conditions, and it is also related to increased transmission of communicable respiratory diseases, as pathogens typically spread through droplets and aerosols expelled while coughing.

Cough is generally considered to be a conspicuous, easily identifiable symptom, and changes in its frequency and time pattern can provide valuable information to clinicians on aspects such as disease severity (Hall, 2020), progression and even origin. Nonetheless, objective collection of these data is considerably difficult, and greatly depends on subjective, self-reported information provided by patients (Jaeger, Szidon, Doucette, 1995).

Although several alternative electronic recording cough mechanisms exist, these are typically dependent on the use of external systems using microphones to detect noises exceeding pre-defined thresholds, associated with cough's acoustic signature. (Matos et al, 2007, Barton et al. 2012, McGuinness et al. 2012). These systems have several flaws, including their high cost, difficult scalability, continuous recording (which compromises patient's privacy), and incapacity to integrate cough data with other variables such as time and location of individual cough episodes. (Hall et al. 2020).

We propose using a smartphone-based artificial intelligence (AI) system to detect putative cough sounds, which are then relayed to a remote server and analyzed by an AI machine learning model able to classify sounds as coughs or not coughs. The system also records the time and approximate GPS location of the user when a cough was detected. Furthermore, despite having access to the mobile phone's microphone, the app only records putative cough sounds, protecting the user's privacy.

Our team is currently evaluating the efficacy of community-level surveillance using this tool for the early detection of respiratory infection outbreaks in a parallel study conducted in Navarra, Spain.

At an individual level, continuous cough monitoring has also been long-considered a good proxy of response to treatment in patients with chronic pathologies. But the process still depends on either subjective report by patients, or manual counts of cough episodes recorded by specialized instruments. (Decalmer et al, 2007) We now intend to evaluate the efficacy of a mobile acoustic tool to monitor treatment progression and facilitate etiologic diagnosis in patients with chronic cough of unknown origin at the Clínica Universidad de Navarra.

Furthermore, these systems could also be refined to facilitate the identification of patterns associated to specific diseases. Similar strategies have been already used for the screening of asthma (Porter et al, 2019, Sharan et al, 2018), tuberculosis (Prahar et al, 2021) and COVID-19 (Laguarta, Hueto, Subirana, 2020), but not in the context of chronic cough, whose clinical diagnosis remains particularly challenging. In this study, we will (i) evaluate the value of AI-enabled cough monitoring systems for screening, diagnosis, and management of respiratory diseases. We will also simultaneously (ii) collect data from patients with a known diagnosis to enrich existing, public disease-annotated cough datasets for the further training of cough-recognition AI systems, and (iii) assess the technical performance of available digital cough monitoring applications.

2.1. Investigational diagnostic intervention

We will evaluate the potential role of the mobile app Hyfe Cough Tracker™ as a tool to screen for, diagnose and support clinical management of respiratory diseases. For example, among other potential applications, we will use digital cough monitoring to monitor response to

treatment in patients with a diagnosis of chronic cough of unknown origin, and compare it with self-perceived improvement, while enriching a dataset of disease-specific annotated coughs. Hyfe Cough Tracker™ has been developed by a multidisciplinary team including medical doctors, data scientists and software developers, for mobiles using Android and IOS operative systems. It runs in the background of the mobile, having constant access to the microphone. However, it only records explosive sound snippets of 0.5 seconds or less. These records are then relayed to a remote server, where they are classified by the convolutional network as "coughs" or "not coughs". Sounds labeled as coughs are then registered both in the app, and in a dashboard accessible by researchers. Hyfe does not record any sound longer than 0.5 seconds, therefore, it will not register conversations or background sound, protecting the privacy of participants. The algorithm can be used on the recordings from a wearable recording system incorporating a lapel mic, which can be used to monitor cough continuously in the absence of a smartphone (Hyfe Air).

2.2. Diseases under study

For the current study we will focus on patients who present to the pneumonology, check-ups and infectious diseases outpatient clinic at the Clínica Universidad de Navarra with cough. This includes patients with cough of unknown origin and those with a previously diagnosed respiratory disease. As a symptom, cough: (i) represents an important reason for consultation at the clinic, (ii) represents a diagnostic challenge, and (iii) is an indicator of the efficacy of specific treatments for underlying respiratory conditions.

Examples of these conditions include, but are not limited to:

- **Chronic cough:** Typically considered to be any cough that lasts for 8 weeks or longer. This condition is thought to affect around 10% of the world adult population and can be caused by a myriad of different conditions, including: allergies, sinus infection, asthma, chronic obstructive pulmonary disease (COPD), cancer, gastroesophageal reflux disease (GERD), among others. (Song et al, 2016).
  Its diagnostic process is typically lengthy, and largely based on empiric observations of response to different treatment schemes, which is in turn dependent on subjective reports from patients. Designing a method to objectively monitor this process will likely speed up diagnosis and lead to improved clinical outcomes and reduced costs in the healthcare sector.
- **Tuberculosis:** *Mycobacterium tuberculosis* is the leading infectious cause of death in the world, with most cases occurring in low-income countries. Global prevalence is unknown, but 25-47%

of the total population of India, China, the Philippines, and Pakistan (the five countries with the highest burden) are estimated to be infected (Ragonnet, Trauer, Geard, 2019). Chronic cough is a symptom of pulmonary tuberculosis, and a major driver for transmission of the disease (Turner, 2019).

- **Non-tuberculous mycobacterial infection (NTM):** Infections caused by mycobacteria other than the *M. tuberculosis* complex. NTMs are a common cause of chronic cough among patients with an underlying anatomical airway abnormality, and prevalence has steadily increased in the last decades, following improvements in diagnostic methods and increased incidence of iatrogenic immunosuppression for various rheumatologic and oncologic conditions, for example. Treatment schemes are often long and associated to poor patient adherence. Mortality rates among people with NTM are more than twice as high as those compared to the general population. (Wassilew, Hoffman, Andrejak, 2016. Park, Kan, Han, et al, 2019).
- **Asthma:** Over 339 million people was estimated to have asthma in 2016, making it once of the most important non-communicable diseases in the world. Cough, wheezing, and shortness of breath are typical symptoms. Although relatively simple, treatment with inhaled corticosteroids requires regular follow ups and monitoring based on subjective, self-reported indicators. (WHO, 2018).

2.3. Evaluation of benefit/risk ratio for the participants in the clinical trial

Privacy concerns have been addressed by the research team. All data and metadata collected by Hyfe are secured and stored in safe, GDPR-compliant servers. Only anonymized data is shared between the sponsor and co-investigators. There are no anticipated benefits for taking part in this study. Cough records will be used to enrich Hyfe's database and help evaluate its potential as a screening diagnosis tool for respiratory diseases.

2.4. References from literature

- Barton A, Gaydecki P, Holt K, et al. 2012. Data reduction for cough studies using distribution of audio frequency content. Cough 8:12
- Boulet, Louis-Philippe, et al. "Tools for assessing outcomes in studies of chronic cough: CHEST guideline and expert panel report." Chest 147.3 (2015): 804-814.
- Bujang, Mohamad Adam, and Tassha Hilda Adnan. "Requirements for minimum sample size for sensitivity and specificity analysis." *Journal of clinical and diagnostic research: JCDR* 10.10 (2016): YE01.

- Decalmer, Samantha Clare, et al. "Chronic cough: how do cough reflex sensitivity and subjective assessments correlate with objective cough counts during ambulatory monitoring?." Thorax 62.4 (2007): 329-334.
- Gabaldon-Figueira, Juan Carlos, et al. "Digital acoustic surveillance for early detection of respiratory disease outbreaks in Spain: a protocol for an observational study." BMJ open 11.7 (2021): e051278.
- Hall, J. I., Lozano, M., Estrada-Petrocelli, L., Birring, S., & Turner, R. (2020). The present and future of cough counting tools. Journal of Thoracic Disease, 12(9), 5207.
- Jaeger, R. J., J. P. Szidon, and P. J. Doucette. "Instrumentation for measuring frequency of cough." Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society. Vol. 2. IEEE, 1995.
- John Hopkins. 2021. COVID-19 dashboard. Accessed 3 January 2021.
- J. Laguarta, F. Hueto and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 275-281, 2020, doi: 10.1109/OJEMB.2020.3026928.
- Matos, Sergio, et al. "An automated system for 24-h monitoring of cough frequency: the leicester cough monitor." IEEE Transactions on Biomedical Engineering 54.8 (2007): 1472-1479.
- McGuinness K, Holt K, Dockry R, et al. 2012. Validation of the VitaloJAK 24 hour ambulatory cough monitor. Thorax 67:A159.
- Pahar, M., Klopper, M., Reeve, B., Theron, G., Warren, R., & Niesler, T. Automatic Cough Classification for Tuberculosis Screening in a Real-World Environment. (2021) *arXiv preprint arXiv:2103.13300*.
- Park, S.C., Kang, M.J., Han, C.H. et al. Prevalence, incidence, and mortality of nontuberculous mycobacterial infection in Korea: a nationwide population-based study. BMC Pulm Med 19, 140 (2019)
- Porter P, Abeyratne U, Swarnkar V, et al. A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children. Respir Res. 2019;20(1):81.

- Ragonnet, R., Trauer, J.M., Geard, N. et al. Profiling Mycobacterium tuberculosis transmission and the resulting disease burden in the five highest tuberculosis burden countries. BMC Med 17, 208 (2019)
- Sharan RV, Abeyratne UR, Swarnkar VR, Claxton S, Hukins C, Porter P. Predicting spirometry readings using cough sound features and regression. Physiol Meas. 2018;39(9):095001
- Song WJ, Chang YS, Faruqi S, Kang MK, Kim JY, Kang MG, Kim S, Jo EJ, Lee SE, Kim MH, Plevkova J, Park HW, Cho SH, Morice AH. Defining Chronic Cough: A Systematic Review of the Epidemiological Literature. Allergy Asthma Immunol Res. 2016 Mar;8(2)
- Turner RD. Cough in pulmonary tuberculosis: Existing knowledge and general insights. Pulm Pharmacol Ther. 2019 Apr;55:89-94.
- Wassilew N, Hoffmann H, Andrejak C, Lange C. Pulmonary Disease Caused by Non-Tuberculous Mycobacteria. Respiration. 2016;91(5):386-402. doi: 10.1159/000445906.
- World Health Organization , Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva; 2018.

3. OBJECTIVES OF THE TRIAL

Primary objective:

1. To evaluate the value of AI-enabled cough monitoring systems for screening, diagnosis, and management of respiratory diseases in clinical practice.

Secondary objectives:

1. To assess the clinical performance of questionnaire-based clinical cough assessment by comparing cough records generated by Hyfe Cough Tracker™ to self-reported cough scores.
2. To generate databases of cough sounds annotated with demographic and medical data.
3. To determine the sensitivity and specificity of Hyfe Cough Tracker to detect and classify coughs in a controlled environment.
4. To determine the sensitivity and specificity of Hyfe Cough Tracker and Hyfe Air in a real-life setting.

Research questions to be addressed by this project include: (i) Is there a difference in the self-reported cough frequency of participants and that registered with acoustic surveillance systems? And (ii) can acoustic cough patterns facilitate the etiologic diagnosis of patients with chronic cough?

4. DESIGN OF TRIAL

4.1. Global design

This is a single center study to be carried out at the two campuses of the Clínica Universidad de Navarra, in the cities of Pamplona, and Madrid, Spain.

For the cough-monitoring process, we will use, an Artificial-Intelligence system (AI) that detects and records explosive putative cough sounds and identifies human cough based on acoustic characteristics (see below).

Potential participants either attending the outpatient clinic or hospitalized with a complaint of cough will be invited by their treating physician, or a member of the research team and included in the study by part of the research team. A researcher will instruct participants on how to install and use Hyfe Cough Tracker in their smartphones. If necessary, the researcher might provide participants with a smartphone configured to exclusively run Hyfe Cough Tracker™. Researchers who enroll participants will also be responsible of obtaining informed consent forms personally.

Once turned on, the app will monitor cough. Participants provided with a study phone will be instructed to run the app continuously and keep the study smartphone close to them for the period of monitoring.

Depending on the objective, participants will either be submitted to diagnostic procedures deemed necessary by their attending physician or will receive the treatment recommended by their attending physicians and will be monitored for a variable period.

Monitoring period

The monitoring period will vary for outpatient and inpatient participants. Inpatients will be monitored from the moment of first consultation/admission to the hospital until their discharge. Outpatients will be monitored for an initial period of one month, subject to modification

depending on the characteristics of specific diseases under study, and prior obtention of informed consent.

Assessment of correlation with self-reported questionnaires

Participants will be asked to complete a daily online, standardized 100mm cough visual analogue scale (VAS), as described by Boulet et al in 2015. Participants will also be instructed to fill a validated Spanish version of the Leicester Cough Questionnaire (LCQ) (Muñoz et al, 2016) at the time of recruitment, and again at the end of the monitoring period.

These records will be compared with those registered by the cough surveillance system. We will evaluate the correlation between the mean weekly VAS score and the registered cough frequency.

Generation of annotated cough datasets

Participants with an etiologic diagnosis at the time of enrollment, or in which it is reached before the end of the monitoring period, will be retrospectively included in disease-annotated datasets, to assess differences in the acoustic patterns of specific diseases with the neural network created by Hyfe.

Validation of Hyfe's technical performance

A subgroup of participants and/or a group recruited outside the clinical setting will provide consent to participate in this sub-study by marking the sub-study box in the consent form.

Participants will remain for approximately 30 minutes in a quiet room. A smartphone connected to a lapel microphone will be placed at approximately 50 cm from the participant's mouth, another smartphone of the same model, not connected to an external microphone will be placed next to the first one. Each participant will also wear a lapel microphone connected to an audio recorder.

Participants will be instructed via a computer monitor to generate a series of voluntary sounds including but not limited to: reading a text, coughing, counting, sneezing and similar common sounds. There will be silent pauses between the elicited sounds. All sounds throughout this period will be recorded. Any accidental conversation will be deleted. The final audio will be stored in a database, coded with the participant´s ID, sex, and age.

The cough counts registered by the two smartphones will be compared to the consensus evaluation of two medically trained researchers, who will be instructed to determine whether individual sounds were coughs or not.

Finally, the experiment will be repeated, playing back the records obtained from participants using a desktop speaker located at 50 cm of the two smartphones, to compare the application's performance detecting live vs recorded coughs. Audio will be played at decibels matching the average intensity of human coughs (Doherty et al, 1997).

A second sub-group of at least 24 participants will be instructed to use Hyfe continuously for a period between 6 and 24 hours, while they record themselves using a MP3 recorder connected to a lapel microphone. For this, participants will be provided with a smartphone exclusively destined to run Hyfe Cough Tracker, while simultaneously wearing the Hyfe Air wearable system. MP3 records will be processed and manually reviewed by two specially trained healthcare providers to count the number of coughs recorded. This data will be used to evaluate the performance of both devices compared to human listeners in a real-life scenario.

A patient (or their family or legal representative) may interrupt their participation in the study at any time and for whatever reason. The principal investigator may also withdraw a patient from the study if he considers that it is in the best interest of the patient.

5. SELECTION OF STUDY SUBJECTS

5.1. Inclusion criteria

All the patients should comply with the following criteria for inclusion:
For objectives 1 and 2:

1. Outpatient or inpatients at the Clinical Universidad de Navarra with a complaint of cough.
2. The patient or his/her legal representative, have given consent to participate in the study.

For objectives 3 and 4:

1. Being 18 years or older.
2. Providing consent for the specific the sub-study.

5.2. Exclusion criteria

Patients who present any of the following criteria for exclusion cannot be included in the clinical trial:

1. Inability to accept the privacy policy and terms of use of Hyfe.
2. Lack of access to a Wi-Fi network at the site of residence (for objectives 1 and 2).
3. Unwillingness to regularly use the cough-surveillance system throughout the monitoring period.

The investigator shall try to ensure that the participant completes the required monitoring period. The data of the withdrawn patients shall be collected up to the moment of withdrawal. Losses to follow up will not be replaced.
The withdrawal shall be documented in the patient's clinical history and informed to the researchers via a form to be filled online by the patient.

6. DESCRIPTION OF THE TOOL STUDIED

Hyfe is an AI-enabled mobile app that records short snippets (<0.5 seconds) of putative cough explosive sounds and then classifies them as cough or non-cough using a convolutional neural network (CNN) model. Briefly, the acoustic characteristics of recorded sounds are converted into an image file, which is then processed by an algorithm trained to identify graphical differences in images. This creates an adjustable prediction score, with values above it, resulting in a sound being classified as “cough”, and those below being classified as “non-cough” (Figure 1).

This allows the discrimination of cough-like sounds from the rest. Cough sounds are then matched with time and GPS coordinates.

The preliminary analytical performance of the system, understood as its sensitivity and specificity differentiating cough episodes from similar recorded sounds, has been validated in an ongoing study in the province of Navarra, Spain (Gabaldon-Figueira et al, 2021). Between November 2020 and January 2021, Hyfe™ registered nearly 700,000 putative cough sounds, from 62 recruited participants. Of these, 119,876 were manually classified by human observers, revealing a sensitivity of 96.34%, and a specificity of 96.54% with a cough-positivity score threshold set at 0.85 (Figure 2). However, since these records exclusively include explosive sounds detected by the application’s algorithm (ignoring coughs that might have

been catalogued as non-explosive sounds, and therefore not furtherly analyzed), this analytical performance must be validated in a real-life scenario, comparing it to continuous records obtained from participants.

Hyfe Air is a wearable device with an incorporated wireless lapel microphone. The device´s recordings can be run through the same cough-detection algorithm described above, while its results are directly stored in a remote database and are not displayed to participants. While the algorithm is the same, differences in the characteristics of the incorporated microphone compared to those in smartphones suggest that the device's performance might differ from those previously reported, requiring an independent validation process.

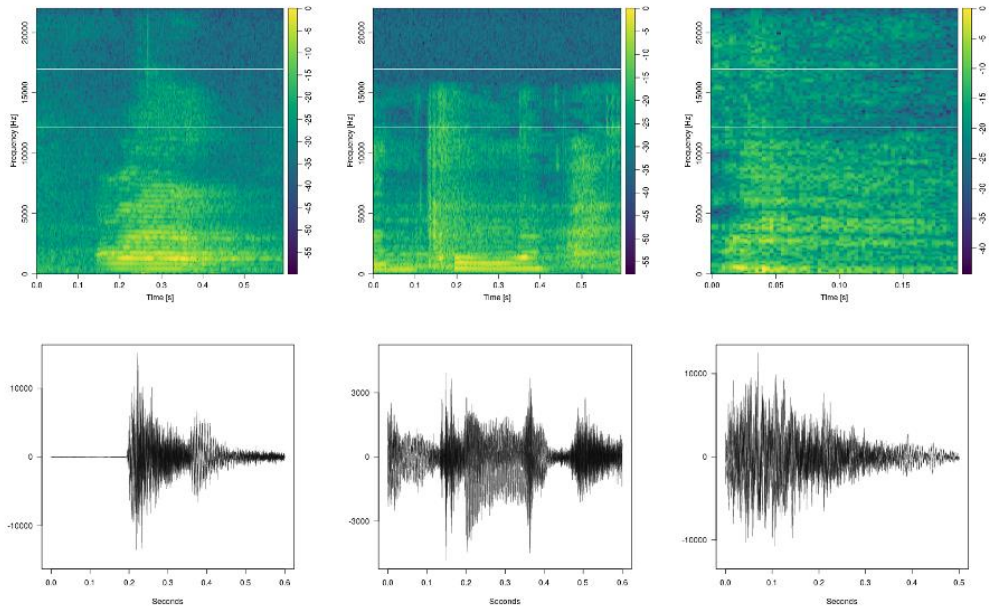

Figure 1: Audio recordings (bottom) are converted to image files (top) and compared with existing datasets using a CNN. Data not published.

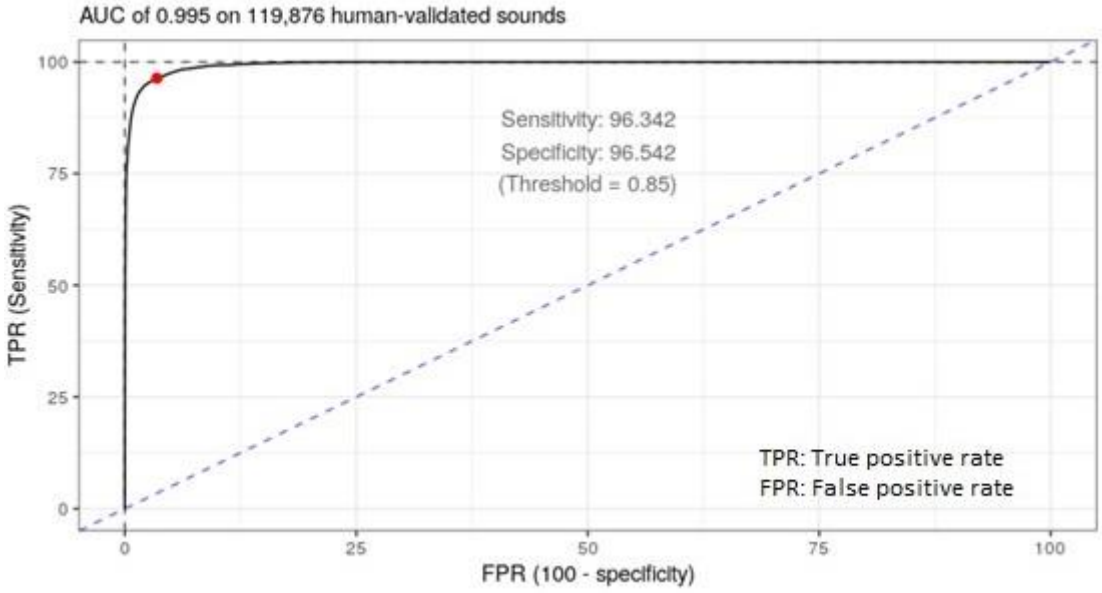

Figure 2: Analytical performance of Hyfe as evaluated in cough records collected between November 2020 and January 2021. Taken from Gabaldon-Figueira et al (2021).

6.1. Drugs or treatments that are permitted (including rescue medication) and not permitted before and/or during the trial

No medications or specific treatments will be prescribed to participants as part of this study. Participants shall continue with any pre-established treatment, or those newly prescribed by their treating doctor during the study period. All treatment prescribed will be recorded in the patients file and study database.

7. EVALUATION OF RESPONSE

7.1. Study plan and procedures

7.1.1. Obtaining consent

A written informed consent must be obtained before carrying out any of the specific procedures of the study. Part of the process of the informed consent consists of clearly explaining the purposes, methods, objectives, and risks of the study to the patient, their family or legal representative. A copy of the informed consent, also signed by the investigator, shall be kept with the clinical documentation. Another copy shall be given to the patient.

The physician shall record the patient's consent to participate in the study in his/her clinical documentation.

Once consent is obtained, the treating physician will be asked to fill a medical information questionnaire (see attached) which will include information on the patient's demographics, diagnosis and past relevant medical history.

In the case of minors, the informed consent will be signed by parents or legal guardians. All adolescents aged 12-17 will provide informed assent.

7.2. Efficacy variables

For the primary objective 1, efficacy will be measured by determining Hyfe's capacity to predict a diagnosis, or reflect clinical evolution of patients, based on cough signatures. For this, the acoustic characteristics of coughs (mean frequency, volume, amplitude) registered from participants with an etiologic diagnose, will be described, to identify patterns associated to specific conditions. Additionally, the Area Under the Curve, sensitivity, and specificity of the AI system will be retrospectively evaluated with the annotated cough registries.

For the secondary objective 1, the efficacy variable will be the daily score in the online VAS, and the coughs per recorded hour.

For the secondary objective 2, there is no clear efficacy variable. We will retrospectively include anonymous coughs registered in patients with an etiologic diagnosis and annotated with data in online datasets from where they can be used to train this and similar AI systems in the future.

For the secondary objectives 3 and 4, efficacy variables will include the calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of Hyfe, compared to two trained human observers.

7.3. Schedule for evaluation, recording and analysis of the efficacy parameter

Participant recruitment is set to start in Q3 2021. Cough registries of patients with an etiologic diagnosis will start to be included in the annotated datasets as soon as the diagnosis is reached.

8. SAFETY APPRAISAL

This study is considered safe. There are no anticipated risks or benefits from participants taking part in this study. All participants can withdraw from the study at any point, prior notification to researchers.

Privacy concerns will be addressed by clarifying that no sound beyond 0.5 seconds of every cough is ever stored. All measures taken to protect patient's privacy and confidentiality are detailed in the "confidentiality" section.

9. STATISTICS

9.1. Description of the statistical methods to be used, including the schedule of all the planned intermediate analyses

For completion of the primary objective

To assess the accuracy of the cough surveillance system to detect cough patterns associated to specific diseases, we will use participants in which an etiologic diagnose is reached before completion of the study. We will search for digital footprints of singular detected cough events, and/or the pattern of cough in time for participants with a given diagnosis.

These records will be used to train a previously developed convolutional neural network to perform respiratory disease cough classification. We will initially use a random split-sample approach (70% derivation & 30% validation) to assess the performance of the newly generated cough classification tool (sensitivity, specificity, positive and negative predictive values, AUC). Performance will then be validated by retrospectively analyzing these participants' coughs with the newly trained AI-network.

The cough classification tool will be supplemented with patient-specific information to assess the incremental value of cough-based multi-variable prediction.

For secondary objective 1

The daily VAS score of participants will be compared to the cough frequency registered by the cough surveillance system. These data will be used to fit a linear regression model to compare self-reported VAS scores to daily cough frequency and calculate a correlation coefficient (r). The

average weekly VAS score and cough frequency on week 1, will be compared to that at study withdrawal using a paired t-test, and represented using before-after plots. P values below 0.05 will be considered significant. A similar approach will be used to compare results of the LCQ at the beginning vs the end of the study. Before-after plots will be constructed to represent these differences. This approach is also expected to provide information on the system's capacity to reflect clinical evolution of participants (Primary objective).

For completion of secondary objective 2

Cough registries of participants with an etiologic diagnosis of respiratory disease will be included in annotated datasets. There are no specific statistical analysis required for completion of this objective.

For completion of secondary objectives 3 and 4

The validation parameters for the application will be calculated in the following way: Sensitivity = True positives (TP)/True positives (TP) + False negatives (FN), Specificity = True negatives (TN)/True negatives (TN)+ False positives (FP). PPV= TP/TP+FP, NPV= TN/TN+FN.

Any sound identified as a cough by both human observers will be considered a positive result for cough. TP will be defined as those sounds identified as a cough by Hyfe, and both human observers. Similarly, a sound identified as non-cough by both observers will be considered a negative. TN will be defined as those sounds not recorded as coughs by neither Hyfe, nor any human observer. Sounds identified as coughs by Hyfe, but not by both human observers will be considered FP, and those identified as non-cough by Hyfe, but not by both human observers will be classified as FN. Sounds for which no consensus is reached by both human observers will be excluded from the analysis. Time stamps from each individual sound will be used to identify them, and determine which ones were not picked up by Hyfe's detection algorithm.

9.2. Sample size calculations

Primary objective

For the primary objective, sample size calculations are complicated due to the variability of cough frequency between different patients with different conditions. To evaluate the system's capacity to detect and classify coughs based on their etiology, we will create and prospectively

increase a split-sample (derivation / validation) dataset and test performance until it plateaus (see data analysis). Based on previous digital cough diagnostic studies performed by the co-investigators, we anticipate needing 15,500 individual cough data points and 200 cough time series by groups for every individual diagnosis. This number of coughs might be rapidly reached by participants diagnosed with a cause of chronic cough but will likely take substantial time for other acute conditions. Thus, the suggested duration of 5 years.

To evaluate the capacity of the system to reflect changes in cough frequency associated to clinical evolution, sample size calculations are more straightforward and detailed below.

Secondary objective 1:

We estimate needing 28 participants. This accounts for an anticipated effect size in the outcome (VAS score) of 0.35, with one predictor (cough frequency registered by Hyfe), and accounting for a loss of follow up of 10% of the participants. The significance level is set at 0.05, with a statistical power of 80%.

Secondary objective 2:
We aim to include all participants with an etiologic diagnosis of cough into the annotated datasets, so no specific sample size is required for completion of this objective.

Secondary objective 3:
We aim to determine the sensitivity and specificity of Hyfe to detect and classify coughs appropriately. For this objective, all coughs will be voluntary, so the cough prevalence of coughs in our sample will be 100%. Including a minimum of 24 patients with 12 voluntary coughs each (288 coughs in total) will be required to achieve a power of 90% at 5% significance level to detect a difference of 5% in the value of sensitivity from 95% in the human observer to 90% by Hyfe. This minimum sample size is also sufficient to detect with 70% power a difference in specificity of 10% from 90% in the human observer to 80% in Hyfe at 5% significance. (Bujang and Adnan, 2016).
Secondary objective 4

Given the similarity with secondary objective 3, the same sample size calculation method was used. Twenty-four participants (12 males and 12 females, preferentially from different age groups) will be recruited for this exploratory analysis.

9.4. Criteria for premature termination of the study

The study might be early terminated if the following criteria are met:

- It is evident that the inclusion of patients is unsatisfactory in terms of quality and/or quantity.
- The data recording is vague and/or incomplete.
- A major risk to the patients' privacy or confidentiality is discovered.

9.5. Procedure for reporting all the deviations from the original statistical plan

All the deviations from the original statistical analysis plan shall be included in the final report of the clinical trial.

10. ETHICAL ASPECTS

10.1. Good Clinical Practice

The study shall be carried out in accordance with the International Conference on Harmonization (ICH) regarding good clinical practice and the corresponding regulatory requirements. The investigator shall be completely familiar with the correct use of the investigational tool as described in the protocol. The essential clinical documents shall be kept to demonstrate the validity of the study and the integrity of the collected data. The main files should be established when the study starts, shall be kept over the course of the study and stored in accordance with the relevant legislation.

10.2. Ethical considerations

The study shall be carried out in accordance with the ethical principles of the latest revision of the Helsinki Declaration and with legislation currently in force. The Navarra Ethical Committee for Research with Medications (CEIm) shall examine all the documentation relating to the study to protect the patients' rights, safety and wellbeing. The study shall only be carried out at the centers for which approval has been obtained from the CEIm. The protocol shall be submitted

to the CEIm, along with the informed consent form, advertising (if applicable), the written information given to patients, updates linked to safety, annual progress reports and any changes made to the above documents.

This study may only start after a favorable decision in writing is received from the CEIm.

Recording of sounds poses specific ethical concerns regarding privacy of data. We propose those are addressable at several independent levels. Device level – While prolonged sound recordings can be used to identify participants (e.g. Amazon's Alexa device), short snippets (< 0.5 seconds) cannot. Conversations or acoustic environments are not recorded. Study level – At all points, participants can opt out of the study and apps can be turned off or removed from phones. Our consent process will explicitly describe exactly what is and is not recorded. The use of unique PIN IDs on phones will ensure that only the investigators can link cough data to personal identifiers. Data sharing level Data transfer agreements will be designed and implemented to ensure codified digital cough data and medical metadata are shared by the investigator in non-identifiable ways.
Datasets shared with other research partners apart from the sponsor will include participant identifiers, but not any other Identifiable information (such as names, IDs, phone numbers, or addresses). Datasets including this kind of information information will be kept under lock, and in password-protected PCs exclusively by the sponsor. Hyfe solely collects non-identifiable data from users. These data will be stored in servers compliant with the European General Data Protection Regulation (GDPR).

10.3. Information for the patient and informed consent

Once the entire study is explained, a written informed consent shall be obtained from the patient, their tutor or legal representative before their participation in the study may be made effective. Adolescents aged 12-17 will sign an informed assent. The method used to obtain and document the informed consent and its contents should comply with the International Conference on Harmonization (ICH) regarding good clinical practice and with all the relevant regulatory requirements.

The investigator (or person delegated by him) shall sign and date the Consent/assent Forms. The investigator shall file the original forms in the Investigator's File in the center.

The patient shall receive the Informed Consent/assent Form and shall be informed that participation in the study is voluntary and may be withdrawn at any time without prejudice to the prior medical care. Neither the Patient Information Sheet, nor the Informed Consent/assent can be modified without agreement from the CEIm and the sponsor.

A copy of the Informed Consent/assent Form should be kept by the patient.

The Consent/assent Form includes information about the need to inspect the Clinical Histories and to enable basic data to be obtained.

10.4. Confidentiality of patients

With a view to respecting patients' privacy, the patients shall be identified with an assigned patient number in all the case report forms, accountability records, reports and communiqués of the study. The investigator shall provide inspectors and any possible auditors or collaborators appointed by the sponsor and the regulatory authorities with access to original records of the patients so that they can verify the data in the case report forms and audit the data collection process. Confidentiality shall be maintained and the patient's identity shall not be made public, to the extent permitted by relevant legislation and regulations (Law 14/1999 and its subsequent regulation Royal Decree 1720/2007).

10.5. Compensation for the patient

The investigational procedures deriving from the study shall be financed by the sponsor. The sponsor shall not bear the costs of the habitual care of the patient, these are any procedures that would be practiced or treatments that would be received independently of their participation in the study.

10.6. Compliance with protocol

The investigator shall carry out the study in accordance with the protocol provided by the sponsor and once approval or a favorable decision is obtained from the CEIm and the relevant regulatory authorities. The protocol should not be changed without the consent of the investigator and the sponsor. Any relevant changes to the protocol require approval or a favorable decision in writing from the CEIm prior to its implementation unless the modification is necessary to prevent immediate risks to patients. The sponsor shall present all the changes

made in the protocol to the regulatory authorities in accordance with legislation currently in force.

When an immediate deviation of the protocol is required to prevent immediate risks to patients, the investigator shall contact the sponsor, if the circumstances so permit, to consider the measures to be adopted. Any deviation from the protocol should be documented in detail in the CRF and the original documentation.

11. PRACTICAL CONSIDERATIONS

11.1. Direct access to source data/documents

The sponsor shall guarantee in the protocol or other written agreement that the investigator or the institution shall permit direct access to the source data or documents for monitoring, auditing, revision by the CEIm, and for inspection of the trial by the medical authorities.

11.2. Responsibilities of all the participants in the trial

Investigator

The investigator should agree with this protocol and have an in-depth knowledge of the properties of the products used in the clinical trial.

The investigator should give the information sheet to the patient and collaborate with him/her to help them understand the explanation provided in the document. It is important for him/her to inform patients that their participation in the study is totally voluntary and that it does not affect the doctor/patient relationship, and to assure them that all the persons involved in the study shall respect the confidential nature of any information relating to the patient.

The Principal Investigator or one of his/her collaborators shall be responsible for correctly collecting, recording, and reporting the data and shall ensure that any serious or unexpected adverse events shall be reported within 24 hours.

It is the duty of the investigator to regularly inform the CEIm of the progress of the study and he/she shall be jointly responsible with the sponsor in preparing the final report.

Sponsor

The sponsor of the study is the natural person or legal entity that has an interest in completing it, signs the applications for authorization sent to the CEIm and/or the Spanish Agency of Medicines and Health Care Products (AEMPS) and is responsible for same, including its execution, commencement and completion. The sponsor shall also be responsible for ensuring compliance with the relevant legal standards.

The sponsor takes on the obligations of a sponsor contained in legislation currently in force, providing all the resources and collaborators required to fulfill said responsibility with full guarantees.

The sponsor shall provide the investigator with an Investigator's File. This file shall be used for all the relevant documents related to the study. The investigator shall be responsible for updating the Investigator's File, checking that all the required documents are included during and after the study. The file shall be inspected during the monitoring visits and shall be kept by the investigator after the study.

11.4. Audit/Inspections

The regulatory authorities, the CEIm and the sponsor or an appointed representative may ask for access to all the original documents, case report forms of the patients and other documentation of the study in order to carry out an audit or inspection at the center. The investigator should guarantee direct access to these documents and collaborate at all times in said activities.

12. DATA MANAGEMENT

12.1. Case Report Forms (CRF)

The CRFs shall be completed by the investigation team, transcribing the data from the original documents that form the patient's clinical history. The original CRF shall be sent to the sponsor and a copy shall remain at the center. Indelible black ink shall be used to record all the details in the CRF required by the sponsor, along with any other significant details. The investigator shall sign and date the CRF to guarantee its authenticity and accuracy. The CRF may be completed by any authorized person, whose signature is recognized. Any changes shall be made in a clearly

visible manner and shall have the initials of the person who made the correction and the date when they were made. The completed CRF shall be checked during the monitoring visit.

Any outcome outside the range of normality shall be duly commented.

12.2. Documents to be kept by the investigator

The investigator shall keep all the original documents of the clinical history and a copy of all the CRFs and the identification list of the subjects for 25 years or when the sponsor indicates otherwise. This documentation shall not be destroyed without written consent from the trial sponsor.

The investigator shall keep all the records of the study in accordance with the good clinical practices of the International Conference on Harmonization (ICH) and the corresponding regulatory requirements.

12.3. Study archive

Clínica Universidad de Navarra shall keep a Principal Archive of the Study for the lifetime of the product.

12.5. Final report

The investigation team shall draw up a report suitable for presentation to the relevant authorities.

In addition to this, aggregated datasets wit anonymized registries will be securely shared between the sponsor of the study (Clínica Universidad de Navarra), and co-investigators (Dr. Simon Grandjean Lapierre), who represents the collaborating institution (CR-CHUM)

13. PUBLICATION OF RESULTS OF TRIAL AND USE OF INFORMATION

The results of this clinical trial shall be published in scientific journals and shall mention the CEIm, which approved the study.

Basic regulations of the trial

The Principal Investigator and/or the Hospital are obliged not to use, transmit to third parties, divulge and/or publish the results obtained in this trial without prior written consent from the

Clínica Universidad de Navarra/ Universidad de Navarra. They should in any case respect the following conditions:

The results of this study may not be published until completion of the primary objective, or before, if both parties agree to do so.

a) The results of the trial and any proposal for publication shall be sent to the sponsor, at least 30 days before it is sent for publication. The sponsor should reply in writing within said period, granting said authorization or giving reasons for denying it. If no reply is received in this period, it shall be assumed that publication has been authorized.

b) The sponsor shall not cite the names of the investigators without their authorization, except when referring to work that is already published.

c) The sponsor shall permit publication of the data obtained in this trial in journals of recognized scientific prestige and the divulgation of its contents in seminars and conferences in the professional medical sphere, as long as the conditions in paragraphs a) and b) of this section are respected and if the final draft of the article may be reviewed within a maximum of thirty days. In any case, the legitimate interests of the trial sponsor shall be protected, such as in obtaining optimal protection of patients, coordination in the presentation of documents to medical authorities or other studies in progress in the same field, protection of confidential data and information, etc.

d) These regulations are understood to be applicable to information obtained in uncompleted trials or studies that were suspended before termination.