Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-22-1492

Reviewer A

This study has 2 components i) description of performance of the AI algorithm and ii) correlation with clinical outcomes. The principle behind the need for easier and longer cough monitoring is important in the field of chronic cough, and this manuscript provides some additional data, and competition to existing technologies, which is a step in the right direction.

I have some major and minor comments for the authors to review:

Comment 1: The authors have used validation throughout the manuscript, however, i am very uncomfortable with this term as this is a very small study, of asthmatics, who are young, in a quiet room, without loud noises, sleeping alone and recorded over a very short time period daily and at night. The term "performance characteristics" is probably more appropriate.
Reply 1: We appreciate the reviewer's comment. As suggested by the reviewer, we deleted "validation" (line 69 and 155) or substituted it with "performance characteristics" (line 271).

Changes in the text
(Abstract – Line 69-71) In this prospective observational validation study, this AI-based cough algorithm was applied among real-world patients with an acute exacerbation of asthma.
(Introduction – Line 155-158) In this prospective validation study, the objective was to evaluate whether cough counts measured by the CoughyTM AI algorithm are … during their treatment with systemic steroids and bronchodilator.
(Results – Line 277) Validation Performance characteristics of AI-based Cough Counting Algorithm

Comment 2: The authors have performed pearson correlation coefficients which is not the optimum method to validate. Clinical validation metrics are more complicated/detailed and require showing accuracy, sensitivity, specificity, (2 x 2 tables etc). The authors have not shown this but described something in the discussion only with one sentence.

Reply 2: Thank you for raising this important issue. The confusion matrix is as below and we added the following information to the results.

		Cough counted by human	
		Cough	Non-cough
Cough counted by AI algorithm	Cough	20,379	5,449
	Non-cough	4,369	625,789

Sensitivity: 82.3%
Specificity: 99.1%
Accuracy: 98.5%

Changes in the text
(Results – Line 282-285) From the audio data total 655,986 cough-like sounds were extracted and classified by human and AI algorithm. The cough count results were compared and it showed 82.3% of sensitivity, 99.1% of specificity and 98.5% of accuracy.

Comment 3: Cough counts are most commonly non-normally distributed, but the authors have used pearson throughout including for clinical data compared to cough data. the authors need to demonstrate normality of data before using pearson. Spearman rank correlations may be more appropriate.
Reply 3: Thank you for pointing this out and we agree with the reviewer's opinion. We replaced Pearson correlation by Spearman rank correlation as you recommend. The results were also revised accordingly.

Changes in the text
(Abstracts – Line 81-88) Cough counts by AI were strongly correlated with manual cough counts during sleep time ($r = 0.908$, $p < 0.001$) and awake time ($r = 0.847$, $p < 0.001$). Sleep time cough counts were moderately to strongly correlated with CS-VAS ($r = 0.339$, $p < 0.001$), the frequency of waking up ($r = 0.462$, $p < 0.001$), and salbutamol use at night ($r = 0.243$, $p < 0.001$). Small-to-moderate correlations were found between awake time cough counts and CS-VAS ($r = 0.313$, $p < 0.001$), the degree of activity limitation ($r = 0.169$, $p = 0.005$), and salbutamol use at awake time ($r = 0.276$, $p < 0.001$). Neither awake time nor sleep time cough counts were significantly correlated with PEFR.
(Methods – Line 230-232) To assess the validity of the CoughyTM AI-based algorithm, Pearson's Spearman's rank correlations (r) were calculated between sleep and awake cough counts derived from the AI algorithm and the trained human counters.
(Results – Line 286-295) Sleep time and awake time cough counts measured collectively during 7-day study period by trained human cough counters and the AI algorithm were strongly correlated to each other ($r = 0.981$ 0.908, $p < 0.001$ and $r = 0.956$ 0.847, $p < 0.001$, respectively; Fig. 3A and Fig. 3B). Interestingly, sleep time and awake time cough counts per hour for each subject were strongly correlated to each other, irrespective of being counted by humans ($r = 0.772$ 0.686, $p < 0.001$; Fig. 3C) or AI ($r = 0.796$ 0.648, $p < 0.001$; Fig. 3D).
In addition, small or moderate to strong correlations were found between sleep time cough counts measured by AI and CS-VAS ($r = 0.472$ 0.339, $p < 0.001$; Fig. 4A), the frequencies of wake-up ($r = 0.568$ 0.462, $p < 0.001$; Fig. 4B), and salbutamol use at night ($r = 0.319$ 0.243, $p < 0.001$; Fig. 4C), but not with A.M. PEFR ($r = -0.003$ 0.007, $p = 0.969$ 0.901; Fig. 4D).

(Results – Line 295-300) Small to moderate correlations were evidenced between awake time cough counts measured by AI and with CS-VAS (r =0.423 0.313, p <0.001; Fig. 5A), the degree of activity limitation (r =0.216 0.169, p =0.006 0.005; Fig. 5B), and the frequency of salbutamol use at awake time (r =0.361 0.276, p <0.001; Fig. 5C), but not with P.M. PEFR (r = -0.132085, p =0.093 0.115; Fig. 5D).

Comment 4: The information in the introduction needs updated as vitaloJAK has been used in the majority of clinical trials of novel anti-tussive, not just gefapixant, these include TRPV1 antagonists, Nav blockers, NK-1 antagonists, BLU-5937, sivopixant. furthermore, the vitaloJAK is a two-step process involving an algorithm to initially cut-out non-cough sounds and then manual tagging is performed on the shorter cut-down file. the authors should reference the most recent vitaloJAK validation research letter in the ERJ. similarly, the LCM is a different process, which relies only on an audio sensor and not accelerometer. the details of their system need greater clarity in the introduction.
Reply 4: We appreciate the reviewer's comment. As suggested by the reviewer, we additionally mentioned as below:

Changes in the text
(Introduction – Line 130-133) The VitaloJAK™ has also been widely used in clinical trials of various cough medications such as TRPV1 antagonists, sodium channel blockers, neurokinin-1 receptor antagonists, and selective P2X3 receptor antagonists including BLU-5937, sivopixant, and eliapixant (8-13).
(Introduction – Line 133-139) The VitloJAK™ has a two-step process involving an algorithm to initially filter out non-cough sounds, which results in a shorter file for analysis. Recently, the performance of this VitaloJAK™ filtering algorithm was assessed on 143 patients with refractory chronic cough, showing high sensitivity and reliability (14). Although the VitaloJAK™ is only approved system for monitoring cough for clinical trials of newly developed medications, this cough monitoring system is dependent on manual counts by trained human cough counters, which is labor intensive to quantify the data.

Comment 5: Whilst in principle running an AI algorithm on an iphone app sounds like a good idea. I do not think the practicality of wearing an iphone around your arm is really feasible or practical in the vast majority of subjects. Hence, I do not believe this device will be generalisable to the broad population. The majority will not agree to carry a phone like the one described, walk around in public in quite places, or sleep alone.
Reply 5: Thank you for your comment and we agree with the reviewer's concern. It may be too early to mention the general usability of this application because of such limitation. However, we believe this may be a small start, and further research is needed. We added the following to the discussion section.

Changes in the text

(Discussion – Line 393-395) We now plan to evaluate whether light-weight devices like smartwatches can serve as a substitute for smartphone-based cough recordings to improve patient acceptability and usability.

Comment 6: The data presented has been generated by an individualised 2 hr daytime and 5hr night time - there is currently no data available to demonstrate that this method would be a reliable method. The authors need to conduct a proper validation with 24hr cough monitoring/counting simultaneously to demonstrate correlations and relationship with 24-hr cough frequency.
Reply 6: We appreciate the reviewer's comment and we agree it is one of the important study limitations, which should be described additionally. We explained this issue more concretely in the discussion section as follows:

Changes in the text
(Discussion – Line 396-400) There are several limitations with this study. First, we did not conduct 24-hour cough monitoring. We designed this study to record cough sounds for 2 hours during the day based on each patient's own selection, and this 2-hours of recording may not be sufficient to represent the entire period of awake time. Furthermore, the recording time was not controlled by study protocol but selected by each patient, which may weaken the power of our study.

Comment 7: From this data it is unknown if the phone device missed any small intensity coughs as there was no concurrent validity?
Reply 7: Thank you for raising this issue. In the case of very low intensity coughs, the possibility of missing cannot be ruled out, but smartphone mics are known to be sensitive enough to detect even breathing sounds (Faezipour, M, and Abuzneid, A. Smartphone-based self-testing of COVID-19 using breathing sounds. Telemedicine and e-Health 2020; 26:1202-1205). The most widely used cough monitors, VitaloJAKTM and LCM, also use audio recordings to detect coughs. Nonetheless, to address this issue about very small intensity coughs, we are planning to conduct another future study with non-audio approach to analyze the limitation of audio for capturing coughs.

Comment 8: There needs to be more information about whether the AI algorithm is a live algorithm or whether the audio recording is recorded and then analysed sequentially.
Reply 8: We appreciate the reviewer's comment. As suggested by the reviewer, we added some information to make it clearer.

Changes in the text
(Method – Line 204-206) Recorded sound files from the smartphone application were collected and uploaded hourly to a secure server. The sound files were sequentially analyzed by both the AI algorithm and trained human cough counters and compared.

Comment 9: Please provide information about data privacy, storage.
Reply 9: We appreciate the reviewer's comment. We added the followings:

Changes in the text
(Method – Line 206-208) To protect participant privacy, the cough analysts could only listen to the 0.5-sec cough-like audio segments and additional 1 second around the segments if clarification is needed.

Comment 10: Please provide information about how the AI algorithm was developed in the first place. what data was used?
Reply 10: We appreciate the reviewer's comment. We additionally described this in the discussion section.

Changes in the text
(Discussion – Line 343-347) The core of AI algorithm consists of two phases: detecting cough-like sounds and classifying events into cough or other sounds. At the cough-like sound detection, the time stamps of cough-like events are saved, and an 0.5 seconds length audio is extracted from each time stamp. Next, the audios are classified into cough sounds or non-cough sounds by deep learning model based on 2D-CNN (Convolutional Neural Network).
(Discussion – Line 347-353) To establish this AI cough recognition system, cough data from 130 patients with cough lasting 3 weeks or more were obtained at allergy or pulmonology clinics in Korea. The cough data was divided into training set and test set consisting of 122 subjects (30,786 coughs) and 8 subjects (2,214 coughs), respectively, it was utilized 30,786 coughs from 122 subjects from a training set and used 2,214 coughs from 8 subjects as a test set, and the AI algorithm achieved sensitivity of 92% and specificity of 96% in ambulatory setting.

Comment 11: Please provide information about battery requirements needed to run the algorithm.
Reply 11: We appreciate the reviewer's comment. The algorithm is run on the server so there is no requirement of additional battery for running algorithm. For the multi-day recording the participants were guided to charge the smartphone during every sleep time. We added the followings:

Changes in the text
(Method – Line 199-200) The smartphone was charged during sleep time to record for consecutive days.

Comment 12: Please provide patient report usability data.
Reply 12: Thank you for raising this issue. Since this study was not conducted to obtain medical device certification, we did not collect separate usability data and the IRB also approved our study without requiring patient report usability data. However, most of the study subjects completed the test without any problems or complaints.

Comment 13: Figure 1 cough counts seem highly variable - please provide as box and

whisker as mean/SD or median/IQR or geometric means
Reply 13: Thanks for raising this issue. As you recommended, we revised Figure 1 and Figure 2 as box and whisker as median/IQR. Please refer to revised Figure 1 and Figure 2.

Comment 14: The cough counts in figure 1 in the day/night seem unusual. suggests cough frequency at night is about 20/hr and in the daytime is also 50/hr on day 1. i accept that this during an exacerbation, but even so, this seems a bit high whilst patients are asleep. this is unusual as all studies have previously reported massive reduction whilst asleep. even on day 7, the cough frequency at night (over 5hrs) is about 10 coughs/hr with wide variability, and day-time is similar. this does not seem to fit the pattern of diurnal cough variation between awake/asleep from all other datasets. please explain.
Reply 14: We appreciate the reviewer's comment. Your comment is appropriate. Here is our opinion. The hospitals in which this study was conducted are tertiary referral hospitals that mainly see asthma patients with relatively high severity. In addition, when each researcher enrolled subjects in this study, acute exacerbated asthmatic patients with particularly complaints of cough as the main symptom were given priority, who were considered suitable candidates for the study aim. Another aspect is that coughing in asthma exacerbations has the characteristics of a "paroxysmal and spasmodic dry cough".
The above reasons may explain the high number of coughs during the day and the lack of suppression of coughing to zero level during sleep. Although there was no 'massive' reduction in night time cough, it seems to be consistent with the decrease in coughing at night, as previously known, in that coughing during sleep was also considerably reduced compared to daytime in our study. We hope this has adequately answered your question.

Comment 15: You need to provide data on what time of day patients actually recorded the daytime and asleep coughs. previous data demonstrates that asthmatic patients cough quite a lot when they first wake up the morning:
https://www.jaci-inpractice.org/article/S2213-2198(18)30849-3/pdf
Reply 15: We appreciate the reviewer's comment. As suggested by the reviewer, we added about this information on the result section.

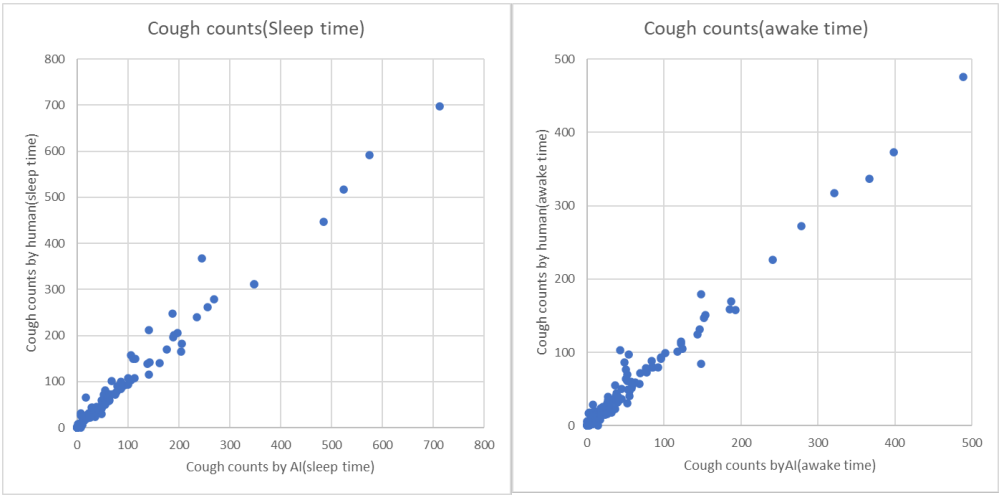
Changes in the text
(Results – Line 279-280) The median start time of recording was A.M. 08:59 [IQR, 07:21-09:55] for awake time and A.M. 00:01 [22:40-00:57] for sleep time.

Comment 16: Figure 3 C and D, please re-check x-axis labels - seem incorrect. unmatched.
Reply 16: Thank you for pointing this out. We meant that sleep time and awake time cough counts per hour from each subject were strongly correlated to each other, irrespective of being counted by humans or the AI algorithm as described in test. We

previously described this in result section as below.

In the text
(Result – Line 280-291) Interestingly, sleep time and awake time cough counts per hour for each subject were strongly correlated to each other, irrespective of being counted by humans (r =0.772 0.686, p <0.001; Fig. 3C) or AI (r =0.796 0.648, p <0.001; Fig. 3D).

Comment 17: Figure 3 correlations suggest inaccuracies greatest at higher cough frequencies, please can you explain this.
Reply 17: Thank you for pointing this out. The inaccurate points came from a single subject who coughs continuously in extremely short period of time. The AI algorithm counts two coughs as one if the time difference between two coughs is shorter than 0.25 seconds which prevent duplicated count. This anomaly was observed only in one subject. Omitting cough counts from this subject, we found more significant correlation between cough counts measured by trained human cough counters and those by the AI algorithm at sleep time (γ =0.981, p <0.001) and awake time (γ =0.953, p <0.001). Please refer to the following scatter dot plots. We think that it is a point need to be improved in the algorithm.



Comment 18: Figure 4 and 5 correlations are highly unconvincing due to the high variability. please check for normality and replot either using log transformed axis or use spearman rank.
Reply 18: We appreciate the reviewer's comment. As suggested by the reviewer, we revised Figure 4 and 5 with Spearman rank correlation.

Reviewer B

It is a prospective observational study perform to validate an artificial intelligence-based cough count system in patients with asthma exacerbation. The subject of the

study is very interesting and up to date as there are only few tools dedicated to monitor cough. A smartphone-based application for cough counting seems to be promising solution useful in clinical settings.
The manuscript is well written. Methods were described properly. The results were presented both in text and in tables and figure. Limitations of the study were mentioned. The conclusions are justified.
However, there are few shortcomings which, in my opinion, need explanation or correction:
Comment 1: Introduction:
Authors state that both available cough monitoring systems (VitaloJAK and LCM) are dependent on manual counts by people, which is labor intensive, but these two types of monitors differ significantly as far as operator input time is concerned, what should be mentioned.
Reply 1: We appreciate the reviewer's comment. As suggested by the reviewer, we additionally mentioned as below:

Changes in the text
(Introduction – Line 139-141) It needs manual labor time of 87 minutes (15) and 5 minutes (16) per 24-h recording for the VitaloJAKTM and LCM, respectively.

Comment 2: Methods:
Exclusion criteria: Were patients with upper airway diseases excluded from the study?
Reply 2: Thank you for pointing this out. Since the majority of asthmatics has allergic rhinitis and/or chronic rhinosinusitis, we did not exclude patients with rhinitis or sinusitis. We added some comment to make it clearer.

Changes in the text
(Methods – Line 175-181) Subject were excluded from the study if they met any of the following criteria: having a disease or condition other than asthma that may cause cough or dyspnea, such as pneumonia, tuberculosis, interstitial lung disease, heart failure, renal failure, pulmonary arterial hypertension, suspected gastroesophageal reflux (GERD) symptoms (e.g. heart burn, regurgitation), proven GERD by gastric endoscopy in the past year, and upper airway disease other than allergic rhinitis and sinusitis which are common comorbid diseases in asthma patients.

Comment 3: Methods and results:
Is interobserver agreement for two trained experts (who analyzed number of cough episodes) available?
Reply 3: We appreciate the reviewer's comment. Currently, we can only use the final results of the labeling of coughs by two trained experts. We agree that the need for the future study to validate the inter-observer agreement for two trained experts.

Comment 4: Statistics: For correlation analysis when ordinal variable are analyzed (such as wake up or salbutamol use) Kendall's tau-b or Spearmans's coefficient (rho)

would more appropriate statistic.
Reply 4: Thank you for pointing this out and we agree with the reviewer's opinion. We replaced Pearson correlation by Spearman rank correlation as you recommend. The results were also revised accordingly.

Changes in the text
(Abstracts – Line 81-88) Cough counts by AI were strongly correlated with manual cough counts during sleep time (r =0.908, p <0.001) and awake time (r =0.847, p <0.001). Sleep time cough counts were moderately to strongly correlated with CS-VAS (r =0.339, p <0.001), the frequency of waking up (r =0.462, p <0.001), and salbutamol use at night (r = 0.243, p < 0.001). Small-to-moderate correlations were found between awake time cough counts and CS-VAS (r =0.313, p <0.001), the degree of activity limitation (r =0.169, p =0.005), and salbutamol use at awake time (r =0.276, p <0.001). Neither awake time nor sleep time cough counts were significantly correlated with PEFR.
(Methods – Line 230-232) To assess the validity of the CoughyTM AI-based algorithm, Pearson's Spearman's rank correlations (r)- were calculated between sleep and awake cough counts derived from the AI algorithm and the trained human counters.
(Results – Line 286-295) Sleep time and awake time cough counts measured collectively during 7-day study period by trained human cough counters and the AI algorithm were strongly correlated to each other (r =0.981 0.908, p <0.001 and r =0.956 0.847, p <0.001, respectively; Fig. 3A and Fig. 3B). Interestingly, sleep time and awake time cough counts per hour for each subject were strongly correlated to each other, irrespective of being counted by humans (r =0.772 0.686, p <0.001; Fig. 3C) or AI (r =0.796 0.648, p <0.001; Fig. 3D).
In addition, small or moderate to strong correlations were found between sleep time cough counts measured by AI and CS-VAS (r =0.472 0.339, p <0.001; Fig. 4A), the frequencies of wake-up (r =0.568 0.462, p <0.001; Fig. 4B), and salbutamol use at night (r =0.319 0.243, p <0.001; Fig. 4C), but not with A.M. PEFR (r = -0.0030.007, p = 0.9690.901; Fig. 4D).
(Results – Line 300-304) Small to moderate correlations were evidenced between awake time cough counts measured by AI and with CS-VAS (r =0.423 0.313, p <0.001; Fig. 5A), the degree of activity limitation (r =0.216 0.169, p =0.006 0.005; Fig. 5B), and the frequency of salbutamol use at awake time (r =0.361 0.276, p <0.001; Fig. 5C), but not with P.M. PEFR (r = -0.132085, p =0.093 0.115; Fig. 5D).

Reviewer C

In the present study, the authors validated and reported the outcome of CoughyTM in patients experiencing asthma exacerbation. The topic is clinically relevant in terms of adopting AI in cough counting tools and its use in specific clinical situation. Here are several issues that need to be addressed.
Comment 1: It would be informative to provide other symptoms of participants except for cough, such as wheezing, dyspnea, and sputum. I also suggest the authors to clarify

if there is any potential impact of additional breathing sounds on CoughyTM to detect coughs.

Reply 1: We appreciate the reviewer's comment. Severity of asthma symptom including cough, wheezing, dyspnea, and chest tightness was graded from 0 to 4 according to the Amgen asthma symptom diary (ASD), and their correlations with cough counts measured by AI and human were additionally evaluated. We additionally mentioned as below:

Please refer to revised Fig. 2, Fig. 4, and Fig. 5 for detailed results. Correlations between cough counts and asthma-related symptom as well as FEV1 and FVC were also analyzed and described as follows.

For the issue about breathing sound records, we think that our mobile phone device setup was designed to acquiring cough sounds which includes abrupt sound change. Thus, breathing sounds with relatively smooth and lower magnitude were not clearly recorded. Nevertheless, technical approach of this study could be applied to clarifying the characteristics of breathing sounds with appropriately acquired data.

Changes in the text

(Methods – Line 216-224) Subjects were also asked to complete a number of patient-reported outcome (PRO) measures daily on the CoughyTM app, including: 1) cough symptom visual analogue scale (CS-VAS), which is a measure of overall daily cough symptom severity ranging from 0 cm (not at all) to 10 cm (extremely), 2) frequencies of wake-up during sleep time, 3) salbutamol use (during awake time and sleep time), 4) degree of activity limitation (0, not at all; 1, a little; 2, moderate; 3, quite a bit; 4, extremely) at awake time, and 5) degrees of asthma-related symptoms including cough, wheezing, dyspnea, and chest tightness at sleep time and awake time (0, not at all; 1, mild; 2, moderate; 3, severe; 4, very severe) (23).

(Results – Line 272-275) The degrees of cough and other symptoms at sleep time decreased significantly since Day 4 and Day 5, respectively (Supplementary table 4). The degrees of cough or wheezing and other symptoms at awake time decreased significantly even earlier since Day 2 and Day 3, respectively.

(Results – Line 295-300) Small to moderate correlations were observed between sleep time cough counts measured by AI and degrees of asthma-related symptoms including cough ($r = 0.444$, $p < 0.001$; Fig. 4E), wheezing ($r = 0.299$, $p < 0.001$; Fig. 4F), dyspnea ($r = 0.318$, $p < 0.001$; Fig. 4G), chest tightness ($r = 0.314$, $p < 0.001$; Fig. 4H) at sleep time, and FEV1 ($r = 0.481$, $p = 0.032$; Fig. 4I), but not with FVC ($r = 0.171$, $p = 0.447$; Fig. 4J).

(Results – Line 304-308) Moderate to large correlations were seen between awake time cough counts measured by AI and degrees of asthma-related symptoms including cough ($r = 0.423$, $p < 0.001$; Fig. 5E), wheezing ($r = 0.544$, $p = 0.005$; Fig. 5F), dyspnea ($r = 0.320$, $p < 0.001$; Fig. 5G), but not chest tightness ($r = 0.318$, $p < 0.001$; Fig. 5H) at awake time, FEV1 ($r = 0.376$, $p = 0.205$; Fig. 5I), or FVC ($r = 0.431$, $p = 0.142$; Fig. 5J).

(Discussion – Line 317-320) Further, cough counts using AI algorithm were also moderately to strongly correlated with cough severity as measured by the CS-VAS and the frequencies of wake-up, and rescue medication use, as well as the degree of activity

limitation and asthma-related symptom other than cough.

Comment 2: There is limited information on CoughyTM. Please introduce CoughyTM and applied AI algorithm particularly regarding its mechanism as the reference 13 is available only in the abstract form.
Reply 2: Thank you for pointing this out and your request is very reasonable. We added additional information about CoughTM and its AI algorithm in the discussion section as below.

Changes in the text
(Discussion – Line 343-347) The core of AI algorithm consists of two phases: detecting cough-like sounds and classifying events into cough or other sounds. At the cough-like sound detection, the time stamps of cough-like events are saved, and an 0.5 seconds length audio is extracted from each time stamp. Next, the audios are classified into cough sounds or non-cough sounds by deep learning model based on 2D-CNN (Convolutional Neural Network).
(Discussion – Line 347-353) To establish this AI cough recognition system, cough data from 130 patients with cough lasting 3 weeks or more were obtained at allergy or pulmonology clinics in Korea. The cough data was divided into training set and test set consisting of 122 subjects (30,786 coughs) and 8 subjects (2,214coughs), respectively it was utilized 30,786 coughs from 122 subjects from a training set and used 2,214 coughs from 8 subjects as a test set, and the AI algorithm achieved sensitivity of 92% and specificity of 96% in ambulatory setting.

Comment 3: The change of clinical outcomes (cough count, VAS, frequency of salbutamol use, activity limitation) is simply described (e,g. decreased, improved) in the manuscript. However, their change needs to be statistically analyzed by comparing the data from day1 and day7 in along with the absolute values. Moreover, the statistical significance is recommended to be depicted on the figures.
Reply 3: We appreciate the reviewer's comment. As suggested by the reviewer, we additionally analyzed the change of clinical outcomes by comparing the data of day 2-7 with those of day 1. These results are presented in Supplementary table 3 and the statistical significance is also depicted on Fig. 2.

Changes in the text
(Results – Line 266-272) The overall CS-VAS also decreased significantly from Day 3 to Day 7 (Fig. 2A) (Supplementary table 3). The frequencies of wake-up decreased significantly on Day 5 and Day 7, whereas salbutamol use at sleep time and awake time did not change significantly (Fig. 2B). The degree of activity limitation decreased significantly from Day 3 to Day 7 (Fig, 2C), and morning and evening PEFR improved significantly since Day 3 and Day 2, respectively (Fig. 2D).

Comment 4: I recommend the authors to analyze the correlation between the change of cough count and FEV1 which was evaluated before and after the treatment (albeit no

association with PEFR). Whether it is correlated or not, it would be informative to highlight the importance of cough as an independent domain in asthma.
Reply 4: We appreciate the reviewer's valuable suggestion. Correlation between changes in cough counts and those in FEV1 and FVC was additionally analyzed, which is added in Fig. 4I, Fig. 4J, Fig. 5I, and Fig. 5J. The description was also added in results section as previously described in the answer 1.

Changes in the text
(Results – Line 295-300) Small to moderate correlations were observed between sleep time cough counts measured by AI and degrees of asthma-related symptoms including cough ($r = 0.444$, $p < 0.001$; Fig. 4E), wheezing ($r = 0.299$, $p < 0.001$; Fig. 4F), dyspnea ($r = 0.318$, $p < 0.001$; Fig. 4G), chest tightness ($r = 0.314$, $p < 0.001$; Fig. 4H) at sleep time, and FEV1 ($r = 0.481$, $p = 0.032$; Fig. 4I), but not with FVC ($r = 0.171$, $p = 0.447$; Fig. 4J).

(Results – Line 304-308) Moderate to large correlations were seen between awake time cough counts measured by AI and degrees of asthma-related symptoms including cough ($r = 0.423$, $p < 0.001$; Fig. 5E), wheezing ($r = 0.544$, $p = 0.005$; Fig. 5F), dyspnea ($r = 0.320$, $p < 0.001$; Fig. 5G), but not chest tightness ($r = 0.318$, $p < 0.001$; Fig. 5H) at awake time, FEV1 ($r = 0.376$, $p = 0.205$; Fig. 5I), or FVC ($r = 0.431$, $p = 0.142$; Fig. 5J).

Comment 5: Please clarify whether all the participants recovered from asthma exacerbation after 7-day treatment. Although the authors obtained daily reported outcomes, there is no objective criteria or validated PROs to assess the clinical improvement of the patients. In addition, the absolute change of CS-VAS, salbutamol use, degree of activity limitation, and PEFR seems minimal, which is hard to assess its clinical relevance.
Reply 5: Thank you for pointing this out. As previously described in answer 1, we additionally evaluated ASD and their correlations with cough counts. All symptoms at sleep time and awake time were alleviated significantly after 7-day treatment. We added the followings in the results section and also provided Supplementary table 4.

Changes in the text
(Results – Line 272-275) The degrees of cough and other symptoms at sleep time decreased significantly since Day 4 and Day 5, respectively (Supplementary table 4). The degrees of cough or wheezing and other symptoms at awake time decreased significantly even earlier since Day 2 and Day 3, respectively.
(Discussion – Line 317-320) Further, cough counts using AI algorithm were also moderately to strongly correlated with cough severity as measured by the CS-VAS and the frequencies of wake-up, and rescue medication use, as well as the degree of activity limitation and asthma-related symptom other than cough.

Comment 6: It was noted that the mean value of cough counts was still around 30 during awake time and CS-VAS of 4 on day 7. The authors' interpretation for this finding is

recommended to add in the discussion section.
Reply 6: We appreciate the reviewer's comment. The mean values of cough counts measured by AI and human at sleep time and awake time decreased from 51-61 on Day 1 to 8-12 on Day 7. Meanwhile, the mean values of CS-VAS decreased from 5 on Day 1 to 4 on Day 7 (Supplementary table 1). The treatment duration of 7 days may not be sufficient for completely alleviating cough.

Reviewer D

Comment 1: A key limitation is that the recording time was not controlled by study protocol but selected by each patient. This seriously weakens the reliability of cough count data and also the validity of correlation analyses with clinical indicators.
Reply 1: We appreciate the reviewer's comment. We agree it is one of the important study limitations, which should be described additionally. We described this issue more concretely in the discussion section as follows:

Changes in the text
(Discussion – Line 396-400) First, we did not conduct 24-hour cough monitoring. We designed this study to record cough sounds for 2 hours during the day based on each patient's own selection, and this 2-hours of recording may not be sufficient to represent the entire period of awake time. Furthermore, the recording time was not controlled by study protocol but selected by each patient, which may weaken the power of our study.

Comment 2: In Abstract, 696 hours were collected from 24 patients. Then it is estimated that each patient recorded about 29 hours. However, this is about 60% of the planned recordings (a total of 49 hours per subject in Methods). First, the actual recorded hours (mean with SD or median with IQR) should be described in Results. Second, the low adherence should be discussed.
Reply 2: Thank you for pointing this out. There was a typo and 1417.6 hours were collected from 24 patients (median [IQR]: 58.0 [56.2-63.0] hours). We are sorry for this mistake and appreciate the reviewer's valuable comment. In this study, the participants could continue recording after time limits, which were 2 hours and 5 hours for awake and sleep, respectively. It was automatically scheduled by the software to secure the minimum recording time. We corrected and added the following to the results:

Changes in the text
(Results – Line 278-279) A total of 1417.6 hours audio data was collected including 353.7 hours and 1063.9 hours for awake and sleep, respectively.
(Results – Line 280-282) The median recording hours per subject were 58.0 [IQR, 56.2-63.0] hrs, 14.3 [IQR, 14.1-15.0] hrs and 43.8 [IQR, 41.4-47.6] hrs for total, awake and sleep time.

Total recording time

Total	**1417.6**	**hours**

Awake	353.7	hours
Sleep	1063.9	hours
Hours of recording by subject(median[IQR])		
Total	**58.0[56.2-63.0]**	**hours**
Awake	14.3[14.1-15.0]	hours
Sleep	43.8[41.4-47.6]	hours

Comment 3: Further information should be described on the study population; e.g. their asthma severity and medication status.
Reply 3: We appreciate the reviewer's comment. As suggested by the reviewer, we additionally mentioned as below and also described medication status at the bottom of Table 1.

Changes in the text
(Results – Line 253-256) All subjects had received low dose inhaled corticosteroid and long-acting beta2 agonist for their asthma. Among them, 5 (16%), 1 (4.7%), 3 (14.3%) had also received leukotriene receptor antagonist, anticholinergics, and oral corticosteroid (prednisolone 5mg/day) for their asthma, respectively. Three (14.3%) additionally had taken antihistamine for their comorbid rhinosinusitis.

Comment 4: The symptom profile of asthma exacerbations of patients at study recruitment should be described in Results. Methods described that "Subjects were eligible to participate if they predominantly presented with cough and needed short-term systemic steroids and a short-acting bronchodilator for symptom control"; however, "predominant cough" is not typical for classical asthma exacerbation.
Reply 4: We appreciate the reviewer's comment. Severity of asthma symptom including cough, wheezing, dyspnea, and chest tightness was graded from 0 to 4 according to the Amgen asthma symptom diary (ASD), and their correlations with cough counts measured by AI and human were additionally evaluated. We additionally mentioned as below:
Please refer to revised Fig. 2, Fig. 4, and Fig. 5 for detailed results. Correlations between cough counts and asthma-related symptom as well as FEV1 and FVC were also analyzed and described as follows.

Changes in the text
(Methods – Line 216-224) Subjects were also asked to complete a number of patient-reported outcome (PRO) measures daily on the CoughyTM app, including: 1) cough symptom visual analogue scale (CS-VAS), which is a measure of overall daily cough symptom severity ranging from 0 cm (not at all) to 10 cm (extremely), 2) frequencies of wake-up during sleep time, 3) salbutamol use (during awake time and sleep time), 4) degree of activity limitation (0, not at all; 1, a little; 2, moderate; 3, quite a bit; 4, extremely) at awake time, and 5) degrees of asthma-related symptoms including cough, wheezing, dyspnea, and chest tightness at sleep time and awake time (0, not at all; 1,

mild; 2, moderate; 3, severe; 4, very severe) (23).

(Results – Line 295-300) Small to moderate correlations were observed between sleep time cough counts measured by AI and degrees of asthma-related symptoms including cough ($r = 0.444$, $p < 0.001$; Fig. 4E), wheezing ($r = 0.299$, $p < 0.001$; Fig. 4F), dyspnea ($r = 0.318$, $p < 0.001$; Fig. 4G), chest tightness ($r = 0.314$, $p < 0.001$; Fig. 4H) at sleep time, and FEV1 ($r = 0.481$, $p = 0.032$; Fig. 4I), but not with FVC ($r = 0.171$, $p = 0.447$; Fig. 4J).

(Results – Line 304-308) Moderate to large correlations were seen between awake time cough counts measured by AI and degrees of asthma-related symptoms including cough ($r = 0.423$, $p < 0.001$; Fig. 5E), wheezing ($r = 0.544$, $p = 0.005$; Fig. 5F), dyspnea ($r = 0.320$, $p < 0.001$; Fig. 5G), but not chest tightness ($r = 0.318$, $p < 0.001$; Fig. 5H) at awake time, FEV1 ($r = 0.376$, $p = 0.205$; Fig. 5I), or FVC ($r = 0.431$, $p = 0.142$; Fig. 5J).

(Discussion – Line 317-320) Further, cough counts using AI algorithm were also moderately to strongly correlated with cough severity as measured by the CS-VAS and the frequencies of wake-up, and rescue medication use, as well as the degree of activity limitation and asthma-related symptom other than cough.

Comment 5: Severity of asthma exacerbations should be described in Results.
Reply 5: Thank you for pointing this out. As previously described in answer 4, we additionally evaluated severity of asthma using ASD and their correlations with cough counts. All symptoms at sleep time and awake time were alleviated significantly after 7-day treatment. We added the followings in the results section and also provided as supplementary table 4.

Changes in the text
(Results – Line 272-275) The degrees of cough and other symptoms at sleep time decreased significantly since Day 4 and Day 5, respectively (Supplementary table 4). The degrees of cough or wheezing and other symptoms at awake time decreased significantly even earlier since Day 2 and Day 3, respectively.

(Discussion – Line 317-320) Further, cough counts using AI algorithm were also moderately to strongly correlated with cough severity as measured by the CS-VAS and the frequencies of wake-up, and rescue medication use, as well as the degree of activity limitation and asthma-related symptom other than cough.

Comment 6: Only 24 patients were recruited. It looks small for a validation study.
Reply 6: Thank you for pointing this out and we agree with the reviewer's opinion. It is one of the study limitation, and we additionally mentioned in the discussion section.

Changes in the text
(Discussion – Line 403-405) In addition, only 24 patients with asthma were included

in the study and further work is needed to generalize the findings to other respiratory disease asthma.

Comment 7: Table 1 should describe baseline cough severity score and day 1 cough count.
Reply 7: We appreciate the reviewer's comment. As suggested by the reviewer, we provided Cough count and cough severity score through the study period (including Day 1) in Supplementary table 1 and 3, respectively.

Changes in the text
(Results – Line 259-262) Across the study period Compared to Day 1, awake time and sleep time cough counts measured by human cough counters and the AI algorithm and human cough counters decreased significantly from Day 5 to Day 7, while awake time cough counts measured by the AI algorithm and human cough counters deceased earlier on Day 2 (Fig. 1A) (Supplementary table 1). Sleep time cough counts per hour measured by the AI algorithm and human cough counters decreased significantly on Day 5 and Day 7, whereas awake time cough counts measured by the AI algorithm and human cough counters deceased earlier on Day 3 (Fig. 1B) (Supplementary table 2).
(Results – Line 266-267) The overall CS-VAS also decreased significantly from Day 3 to Day 7 (Fig. 2A) (Supplementary table 3).

Comment 8: Figure 1 should indicate if the data were hourly cough count and present in Box and Whisker plots.
Reply 8: We appreciate the reviewer's comment. As you recommended, we revised Fig. 1 and 2 as box and whisker as median/IQR, and also added hourly cough count in Fig. 1B.

Comment 9: In Methods or Discussion, it should be described how missing data was handled.
Reply 9: We appreciate the reviewer's comment. We added the followings:

Changes in the text
(Methods – Line 228) Missing data were omitted, and the remaining data were analyzed.

Comment 10: Reference: There are more papers that attempted to continuously monitor cough counts using smartphone-based applications. They should be included and discussed.
Reply 10: Thank you for pointing this out. We conducted an additional literature search and added this information to the discussion section, as recommended by reviewer.

Changes in the text
(Discussion – Line 324-329) In particular, a smartphone-based cough monitoring system is emerging, which includes continuous sound collection with a smartphone application, subsequent signal processing, and noise removal, and finally cough sound

identification through machine learning (28,29). There have been studies conducted for detection cough among 7 children hospitalized for respiratory disease, COVID-19 screening, and early recognition of exacerbation of chronic obstructive lung disease and chronic heart failure (30-33).