

# Validation of the Cough Phenotype TBQ among Elderly Finnish Subjects

Heikki O Koskela (✉ [heikki.koskela@kuh.fi](mailto:heikki.koskela@kuh.fi))

University of Eastern Finland

Johanna T Kaulamo

University of Eastern Finland

Tuomas A Selander

Kuopio University Hospital

Anne M Lähti

Kuopio University Hospital

---

## Research Article

**Keywords:** Cough, chronic cough, phenotype, asthma, chronic rhinosinusitis, gastroesophageal reflux disease, cluster analysis

**Posted Date:** May 13th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1637749/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Phenotypes are defined as the observable properties of disorders that are produced by the interactions of the genotype and the environment. The approaches to phenotype disorders have evolved from subjective expert opinion to data-driven methodologies. A previous cluster analysis among working-age subjects with cough revealed a phenotype TBQ (Triggers, Background disorders, Quality of life impairment), which included 38% of the subjects with cough. The present study was carried out to validate this phenotyping among elderly, retired subjects with cough.

## Methods

This was an observational, cross-sectional study conducted via email among the members of the Finnish Pensioners` Federation (N = 26 205, 23.6% responded). The final analysis included 1109 subjects with current cough (mean age 72.9 years (range 64–90 years) with 67.7% females). All filled in a comprehensive 86-item questionnaire including the Leicester Cough Questionnaire. Phenotypes were identified utilizing K-means partitional clustering.

## Results

Two clusters were identified. The cluster A included 75.2% and cluster B 24.8% of the subjects. The three most important variables to separate the clusters were the number of the cough triggers (mean 2.47 (SD 2.34) vs. 7.08 (3.16), respectively,  $p < 0.001$ ), Leicester Cough Questionnaire physical domain (5.38 (0.68) vs. 4.21 (0.81), respectively,  $p < 0.001$ ), and the number of the cough background disorders (0.82 (0.78) vs. 1.99 (0.89), respectively,  $p < 0.001$ ).

## Conclusion

The phenotype TBQ could be identified also among elderly, retired subjects with cough, thus validating the previous phenotyping among working-age subjects. The main underlying pathophysiological feature separating the phenotype TBQ from the rest of the coughing subjects is probably hypersensitivity of the cough reflex arc.

## Background

Phenotypes are defined as the observable properties of disorders that are produced by the interactions of the genotype and the environment [1]. They represent groups with similar clinical characteristics, prognosis and/or therapeutic needs and therefore, can be utilized in the clinical management of the disorders [2].

Cough is usually classified according to the length of the current episode, to acute (< 3 weeks), subacute (3–8 weeks), or chronic (> 8 weeks) cough. This classification is utilized in the current European, British, North American, Indian, Japanese, Korean, Australian, German, and Chinese cough management guidelines for adults [3–11]. Given the high prevalence of cough [12], this classification directs the management of cough in hundreds of millions of people in all continents. However, the classification of cough by the length of the episode is single-dimensional and mainly based on an expert opinion.

In recent years, approaches to phenotype disorders have evolved from subjective expert opinion to more data-driven methodologies like cluster analysis [13]. These methods aim to explore the unknown nature of data through an unsupervised separation of a dataset with little or no ground truth, into a discrete set of “natural,” hidden data structures [14]. This is in contrast to the traditional methods based on human observation and testing of hypotheses using prior knowledge.

We recently performed a cluster analysis in 975 working-age subjects with current cough [15]. Two clusters were identified. The smaller one, involving 367 subjects, was especially characterized by several triggers of cough, many cough background disorders, and poor cough-related quality of life. We proposed an acronym TBQ (Triggers, Background disorders, Quality of life impairment) for this phenotype. Compared to the subjects of the larger cluster, those with the TBQ phenotype showed a significantly higher tendency for cough prolongation in the follow-up survey 12 months later [15]. The population of that analysis was rather homogenous; all were public service employees at their working age, living in two towns. The present study was carried out to validate that analysis in a different population, namely in aged, retired subjects living all over Finland, and under different conditions, namely in the middle of the covid-19 pandemic.

## Methods

### Study design, setting, and population

This was an observational, cross-sectional study conducted via email among the members of the Finnish Pensioners` Federation, with the residential area of the whole Finland. The 26 205 members (mean age 72.7 years, 63.5% female), who had an email address, were sent an invitation to participate along with information about the study. The questionnaires were sent in April 2021 and one reminder email was sent two weeks later. The responses were directly recorded in an electronic datasheet.

The study was approved by the Ethics Committee of Kuopio University Hospital (289/2015). Permission to conduct the study was obtained from the Finnish Pensioners` Federation. Patients were not involved in the design or conduct of this study. The decision of the subject to reply was considered as an informed consent.

### The questionnaire

The questionnaire was almost identical with the one used in our previous cluster analysis among working-age subjects [15]. For the present study, new questions were added about symptoms of obstructive sleep apnea (OSA), Covid-19 infection and vaccination, symptoms of flu at the beginning of the current cough episode, and recurrence of cough episodes. The list of potential cough triggers was completed by speaking, laughing, and deep inspiration. There were 86 questions altogether, and many of them had several alternatives: Questions about social background, lifestyle, general health, doctors' diagnoses and visits, and medications. Appropriate symptom questions for asthma, chronic rhinosinusitis, gastro-oesophageal reflux disease (GORD), OSA, and depressive symptoms were included [16–20]. Respondents with current cough were asked to answer additional cough-related questions, including the Leicester cough questionnaire (LCQ). An English version of the questionnaire can be found as a supplementary file.

## **Definitions of variables derived from the raw data in the questionnaire**

Current acute, subacute, and chronic cough were defined as suggested in international guidelines [3–11]. Current asthma was defined as doctor's diagnosis of asthma at any age and wheezing during the last 12 months [16]. Chronic rhinosinusitis was defined as either nasal blockage or nasal discharge (anterior or posterior nasal drip) and either facial pain/pressure or reduction/loss of smell for more than three months [17]. Gastroesophageal reflux disease was defined as heartburn and/or regurgitation on at least one day per week during the last three months [18]. OSA was defined as presence of two or more of the following features: Arterial hypertension, loud snoring, daytime somnolence or observed apneas [19]. These disorders, in addition to doctor's diagnoses of bronchiectasis and pulmonary fibrosis, were defined as cough background disorders. The number of cough background disorders was calculated by summing up them, giving a value from zero to six. Unexplained cough was defined as absence of any of these disorders. Autoimmune disorder was defined as presence of a doctor's diagnosis of hypothyroidism, rheumatoid arthritis, or other autoimmune disorders. Presence of depressive symptoms was defined as a Patient Health Questionnaire-2 score of three or more [20]. Symptom sum was calculated by summing all reported symptoms except those associated with airway disorders, giving a value from zero to 15. Trigger sum was calculated by summing all reported cough triggers, giving a value from zero to 15. Allergy was defined as a self-reported allergy to pollens, animals, or food. A family history of chronic cough was defined as the presence (now or in the past) of chronic (duration more than eight weeks) cough in parents, sisters, or brothers.

## **Statistical analysis**

All questions of the questionnaire plus the derived variables were included in partitional clustering with K-means method [14] similarly to our previous cluster analysis [15]. Dimension reduction and cluster analysis steps were performed using R statistical software version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) with `diffusionMap`, `NbClust` and `cluster` packages.

At first phase, data were preprocessed. Right skewed (skewness > 1) variables were normalized with  $\log(x + 1)$  function. Next, ordinal and continuous variables were scaled into 0–1 interval. Variable's minimum value or the lowest class got value 0 and maximum value or the highest class 1. Binary variables remained unchanged. Value 0 indicated negative or 'no' alternative and value 1 positive or 'yes' alternative. After that, distance matrix between observations with scaled variables were calculated using Manhattan distance function. Diffusion maps dimension reduction method was applied to extract diffusion coordinates from distance matrix with function *diffuse* using default settings.

The number of clusters was evaluated by the 24 criteria provided by the software. After that, the extracted diffusion map coordinates were clustered into groups with k-means method. Cluster membership was added to original data for further analysis. To validate the clustering, the analyses were repeated by excluding those background variables with no plausible biological association with cough (like hometown, years of education, alcohol consumption *etc.*).

Statistical analyses between the clusters were performed by Mann-Whitney U test, chi-square test, and logistic regression analysis, and the interrelationships of the variables were analysed by the Spearman's correlation coefficient ( $r_s$ ) using SPSS software version 22.0 (IBM SPSS Statistics for Windows. Armonk, NY, USA). Receiver operating characteristic curves (ROC) and the Youden index were utilized to define the cut-off values. The values are expressed by either means and standard deviations, medians and ranges, or percentages. A p value less than 0.05 was accepted as the level of statistical significance.

## Results

Of the 26 205 subjects who were sent the questionnaire, 23.6% responded (6189 respondents, mean (SD) age 72.2 (5.5), 66.4% female, (Fig. 1)). 206 respondents were excluded from the analyses because of age less than 64 years. Of the remaining 5983 respondents, 1109 subjects suffered from current cough. They formed the population in which the clustering was applied. Their mean age was 72.9 years (range 64–90 years, SD 5.3) and 67.7% were females. The proportion of missing values was less than 2.5%, except for the 3 OSA-related symptom questions (3.1–3.7%).

Nine of the criteria provided by the R statistical software suggested two as the best number of clusters, five suggested three clusters, three suggested five clusters, one criterion suggested seven, eight, nine, eleven, or fourteen clusters, and two suggested fifteen clusters. Therefore, the extracted diffusion map coordinates were clustered into two groups, called cluster A (834 subjects (75.2%)) and cluster B (275 subjects (24.8%)).

The distribution of the subjects to the two clusters according to the length of the cough episode is presented in Table 1. Cluster B was represented both in acute, subacute, and chronic cough, though its proportion was largest in chronic cough.

Table 1

Distribution of the clusters in acute (duration < 3 weeks), subacute (3–8 weeks), and chronic (> 8 weeks) cough. 23 subjects could not define the length of the cough episode.

	<b>Acute cough (N = 207)</b>	<b>Subacute cough (N = 70)</b>	<b>Chronic cough (N = 809)</b>
Cluster A (the 'common' phenotype)	169 (81.6%)	57 (81.4%)	587 (72.6%)
Cluster B (the phenotype TBQ)	38 (18.4%)	13 (18.6%)	222 (27.4%)
TBQ: Triggers, Background disorders, Quality of life impairment.			

Table 2 presents those 10 variables that most strongly separated the clusters, according to the p value between the clusters, as well as 23 other variables of particular interest. The cluster B was especially characterized by several reported cough triggers, many cough background disorders, and low LCQ scores (Fig. 2). Of the various cough triggers, paints, fumes, and automobile exhaust fumes most strongly separated the clusters. Of the cough background disorders, asthma most strongly separated the clusters. Of the three LCQ domains, the physical domain most strongly separated the clusters.

Table 2

The clusters and their defining variables among 1109 subjects with current cough. The ten most important variables and 23 variables of special interest expressed, in order of importance. The order was defined by the p value obtained by Mann-Whitney U test or chi-square test between the clusters. The values are expressed by either means (standard deviations) or percentages, unless stated otherwise.

Order	Variable	Cluster A (N = 834)	Cluster B (N = 275)	P value
1	Trigger sum	2.47 (2.34)	7.08 (3.16)	2.39E-78
2	LCQ physical domain	5.38 (0.68)	4.21 (0.81)	1.02E-75
3	Number of cough background disorders	0.82 (0.78)	1.99 (0.89)	9.09E-65
4	LCQ total score	16.1 (2.48)	12.5 (2.79)	7.12E-64
5	Dyspnea with wheezing	8.4%	55.0%	3.38E-61
6	Paints or fumes as a cough trigger	19.2%	70.5%	5.03E-56
7	Current asthma	6.4%	46.5%	2.36E-53
8	LCQ social domain	5.63 (1.00)	4.38 (1.12)	1.47E-52
9	Automobile exhaust fumes as a cough trigger	13.4%	59.3%	4.37E-52
10	LCQ question 3 (tired because of cough) <sup>1</sup>	5.93 (1.07)	4.57 (1.36)	2.36E-48
23	Symptom sum	2.32 (1.71)	4.13 (2.17)	9.67E-35
30	LCQ question 2 (sputum production when coughing) <sup>a</sup>	4.86 (1.84)	3.29 (1.78)	2.20E-30
40	Unexplained cough	38.0%	4.0%	1.93E-26
46	Chronic rhinosinusitis	13.9%	41.8%	1.15E-22
47	Allergy	9.1%	33.1%	1.37E-21

<sup>a</sup> Leicester Cough Questionnaire (LCQ) questions have 7-step scales from 1 = all of the time to 7 = none of the time.

Order	Variable	Cluster A (N = 834)	Cluster B (N = 275)	P value
51	Number of doctor's visits due to cough in the last 12 months	0.40 (1.15)	1.24 (2.19)	4.49E-19
53	Presence of any trigger for cough	71.8%	96.7%	1.11E-17
61	Family history of chronic cough	31.8%	58.0%	2.03E-14
70	Gastroesophageal reflux disease	22.3%	43.8%	8.87E-12
77	Sleep apnea	39.9%	63.4%	4.20E-11
85	Symptoms of flu at the beginning of the current cough	13.9%	30.7%	6.65E-10
86	Number of cough episodes in the last 12 months	4.92 (7.90)	7.39 (8.98)	1.03E-09
105	Duration of the cough episode (median, (interquartile range))	1–5 years (2 to 12 months – 1 to 5 years)	1–5 years (1 to 5 years – 5 to 10 years)	1.29E-05
107	Use of cough medicines in the last 12 months	28.2%	42.4%	1.82E-05
114	Female gender	64.7%	76.7%	0.0003
115	Autoimmune disorder	18.2%	28.0%	0.0007
121	Chronic obstructive pulmonary disease	2.3%	6.2%	0.0029
124	Bronchiectasis	1.3%	4.4%	0.0046
138	Body mass index (kg/m <sup>2</sup> )	27.7 (4.5)	28.2 (5.4)	0.19
146	Pulmonary fibrosis	0.8%	1.8%	0.30
149	Age (years)	72.8 (5.2)	73.1 ((5.4)	0.36
150	Current smoking	1.8%	2.9%	0.38
193	History of Covid-19 infection	1.0%	0.7%	0.72
<sup>a</sup> Leicester Cough Questionnaire (LCQ) questions have 7-step scales from 1 = all of the time to 7 = none of the time.				

Table 3 presents the best cut-off values for the ten most important variables to identify the cluster B and their sensitivity, specificity, and area under the ROC values. After that, a ROC curve was constructed to

evaluate the best number of the main determinants (trigger sum  $\geq 5$ , LCQ physical domain  $\leq 4.9$ , at least one cough background disorder) to separate the clusters (Fig. 3). The presence of at least 2 determinants gave the best Youden index with the sensitivity of 0.96 and specificity of 0.72. The presence of all three determinants gave the sensitivity of 0.61 and the specificity of 0.97.

Table 3

Sensitivity and specificity of the ten main variables to identify cluster B, utilizing the cut-off values giving the best sum of sensitivity and specificity (The Youden index).

Variable	Cut-off value	Sensitivity	Specificity	aROC
Trigger sum	$\geq 5$	0.80	0.81	0.87
LCQ physical domain	$\leq 4.9$	0.75	0.81	0.87
Number of cough background disorders	$\geq 1$	0.72	0.82	0.82
LCQ total score	$\leq 14.7$	0.76	0.76	0.84
Dyspnea with wheezing	Yes/No	0.55	0.92	NA
Paints or fumes as a cough trigger	Yes/No	0.70	0.81	NA
Current asthma	Yes/No	0.46	0.94	NA
LCQ social domain	$\leq 5.6$	0.57	0.90	0.81
Automobile exhaust fumes as a cough trigger	Yes/no	0.59	0.87	NA
LCQ question 3 (tired because of cough) <sup>a</sup>	$\leq 5$	0.73	0.70	0.78

<sup>a</sup> Leicester Cough Questionnaire (LCQ) questions have 7-step scales from 1 = all of the time to 7 = none of the time. aROC = Area under the receiver operator characteristic curve. NA = not applicable

Belonging to cluster B increased the likelihood of at least one doctor's visit due to cough in the last 12 months (OR 3.39 (95% CI 2.53–4.55)). Presence of chronic cough, when compared to those with shorter cough episodes, also increased this likelihood, but less (OR 1.91 (1.36–2.70)). Belonging to cluster B increased the likelihood of having used cough medicines in the last 12 months (OR 1.88 (1.41–2.50)). On the contrary, presence of chronic cough decreased this likelihood (OR 0.66 (0.50–0.89)).

The validation analysis by excluding those background variables with no plausible biological association with cough gave almost identical results. The five most important variables in that analysis were trigger sum, LCQ physical domain, number of cough background disorders, LCQ question 9 (paints or fumes as a cough trigger), and dyspnea with wheezing (data not shown).

There were significant interrelationships between the most important variables: The number of cough triggers was associated with the number of cough background disorders and the LCQ physical domain ( $r_s = 0.28$ ,  $p < 0.001$ , and  $r_s = -0.34$ ,  $p < 0.001$ , respectively), and the number of cough background disorders was associated with the LCQ physical domain ( $r_s = -0.40$ ,  $p < 0.001$ ).

## Discussion

This cluster analysis, which was performed in the middle of Covid-19 pandemic among 1109 elderly, retired subjects with current cough, validates our previous cluster analysis among working-age, employed subjects [15]. Again, two clusters were found. The cluster B, consisting of 24.8% of the subjects, was especially characterized by several triggers of cough, many cough background disorders, and poor cough-related quality of life. These features fit to the cough phenotype TBQ, which was identified in our previous study. The cluster A, lacking these features, may be called as the 'common' cough phenotype.

Clustering is a hypothesis-generating method, with the assumption that the inherent patterns within the data may be a reflection of different underlying pathophysiological mechanisms or genetic basis, i.e., that the identified phenotypes may also represent distinct cough endotypes or genotypes [13]. Both cluster analyses identified the number of reported cough triggers as the most important variable to separate the two phenotypes. Both analyses also identified the chemical triggers like paints, fumes, strong scents and automobile exhaust fumes, as the most important types of cough triggers [15]. Several studies have shown that subjects with lower airway symptoms induced by chemical irritants are especially sensitive to the cough-provocation test with capsaicin [21–30]. Therefore, we hypothesize that the main underlying pathophysiological feature separating the phenotype TBQ from the common phenotype is hypersensitivity of the cough reflex arc. Thus, this phenotype might also represent a distinct endotype. To test this hypothesis, studies applying cough provocation tests to subjects with TBQ and the common phenotype should be carried out. To investigate whether the phenotype TBQ is a distinct cough genotype, studies applying genome analyses on well-characterized subjects with cough should be carried out. Of note, family history of chronic cough was far more common in phenotype TBQ than in the common phenotype.

Though the phenotype TBQ was more common in chronic cough than in acute and subacute cough, the length of the current cough episode was far from the most important variables defining the clusters. This was also the case in our previous cluster analysis [15]. This finding speaks against the classification of cough according to the length of the episode, which is widely utilized in cough guidelines [3–11]. The phenotype TBQ was much more strongly associated with doctor's visits due to cough and the use of cough medicines than the presence of chronic cough. Furthermore, our previous study showed a strong tendency for cough prolongation in the phenotype TBQ [15]. Therefore, cough phenotype TBQ has clinical significance, and identification of it may serve as an indication for prompt and comprehensive clinical evaluation regardless of the duration of the cough episode.

Cough phenotype TBQ resembles the previously described concept 'cough hypersensitivity syndrome' [31]. Both emphasize the enhanced cough response to various external triggers. However, there seems to be two major discrepancies between the entities. First, 'cough hypersensitivity syndrome' has been connected to chronic cough [31] but the present study shows that features of cough hypersensitivity can be present in acute and subacute cough as well. Second, it has been postulated that the 'cough hypersensitivity syndrome' is present in the majority of subjects with chronic cough [31–33] whereas in

the present cluster analysis just 27.4% of subjects with chronic cough showed the features of the phenotype TBQ. These discrepancies may be best explained by the fact that ‘cough hypersensitivity syndrome’ has been described among subjects attending special cough clinics [31] whereas our cluster analyses are based on community-based populations. Given the documented high tendency of the subjects with the phenotype TBQ to seek medical attention, they are probably overrepresented in the population attending special cough clinics. ‘Cough hypersensitivity syndrome’ is mainly based on expert opinion [31]. Despite the above-mentioned differences, it is highly interesting that unsupervised, data-driven analyses lead to similar conclusions to those made by experienced clinicians.

A recent study from Australia supports the present phenotype analysis. In that study, two clusters could be identified among adult subjects with various respiratory symptoms. The smaller cluster, including 32% of the subjects, was characterized by symptoms of laryngeal hypersensitivity and a strong cough response to mannitol [34]. The characteristics of that cluster resemble those of the phenotype TBQ. To our knowledge, there are no cluster analyses in non-Caucasian cough populations.

For clinical purposes, we calculated the best cut-off values to separate the clusters. They were almost identical to those reported in our previous cluster analysis among working-age subjects [15]. Presence of at least two of the three main TBQ determinants was found to be a very sensitive and modestly specific indicator of the cough phenotype TBQ. It gave the largest sum of sensitivity + specificity and is thus the most suitable clinical criterion for the cough phenotype TBQ. Reliable clinical demonstration of the phenotype TBQ requires that a comprehensive list of triggers is presented to the patient in a written form since patients often forget some triggers when asked openly. The list of the 15 triggers asked in the present study can be found in the questionnaire (Supplementary file).

There were slight differences in the questionnaires between the present and the previous study [15], which did not affect the main results. Others and we have shown that there are more cough background disorders in elderly than in younger subjects [35–38]. Therefore, bronchiectasis, pulmonary fibrosis, and obstructive sleep apnea were added to the variable ‘number of cough background disorders’. All six background disorders, which were utilized in this analysis, have been shown to be statistically significantly associated with cough in the present population [38]. Questions about Covid-19 infection and vaccination, and symptoms of flu at the beginning of the current cough episode were added due to the current pandemic. The recurrence of cough was now investigated by asking how many cough episodes the subject had had in the last 12 months. The number of cough episodes was significantly higher in the phenotype TBQ than in the common phenotype. Of note, the recurrence of cough episode was more strongly associated with the TBQ phenotype than the length of the current cough episode.

The present study has several limitations. The participation rate was relatively low, which is typical for e-mail-based surveys. However, the age and gender distribution of the responders did not differ significantly from the original population, though females were slightly more willing to respond than males. High age of the population may have hindered the use of e-mail in some individuals. However, the participation rate in our previous survey among working-age subjects was of similar magnitude [15]. It is possible that

patients with severe cough have been more willing to participate than patients with mild cough. This may have led to an over-representation of the TBQ phenotype. The proportion of current smokers was small, which may have reduced the impact of smoking on the analysis. The prevalence of short, infection-associated cough subtypes was low in the present population, probably due to personal protective and social measures that were highly recommended during the pandemic era[38]. Finally, the analysis was based on the questionnaire data only; spirometry, laboratory, and x-ray data were missing.

The strength of the present study included a large population, which was missing in our previous cluster analysis: Elderly, retired subjects. In addition, the geographic distribution of the present population was wider than that in our previous study [15]. The survey was not directed to subjects complaining of cough but to a community-based population. Therefore, even those coughing subjects who would never complain their cough to doctors were also included. The questionnaire was originally planned and further completed to investigate cough and associated conditions, including also the highly validated LCQ and a comprehensive cough trigger list. The interindividual variation in how subjects recognise and report symptoms was controlled by the variable 'symptom sum'. Furthermore, there were symptom questions that are recommended for epidemiologic studies, to define important background disorders of cough. All raw data plus the derived variables were included in the cluster analysis without hypotheses using prior knowledge. By this way, all relevant features of cough and even undiagnosed but symptomatic background disorders were equally considered in the cluster analysis.

## Conclusions

The phenotypes TBQ and the common phenotype could be identified among elderly, retired subjects with cough, thus validating the previous phenotyping among working-age subjects [15]. The cough phenotype TBQ is associated with frequent doctor's visits due to cough, use of cough medicines, and a high tendency for cough prolongation [15]. Therefore, an identification of the phenotype TBQ may serve as an indication for prompt and comprehensive clinical evaluation regardless of the duration of the cough episode. Our two cluster analyses, performed under very different epidemiologic circumstances, now involve over 2000 non-selected Caucasian subjects with current cough altogether, with a wide variation in age and socio-economic background. In future, data-driven methodologies to define cough phenotypes should be applied also to non-Caucasian populations. Further studies are also needed to explore whether the phenotype TBQ represents a distinct endotype or genotype.

## Abbreviations

GORD: Gastro-oesophageal reflux disease

LCQ: Leicester cough questionnaire

OSA: Obstructive sleep apnea

TBQ: Triggers, Background disorders, Quality of life impairment

# Declarations

## *Ethics approval and consent to participate*

The study was approved by the Ethics Committee of Kuopio University Hospital (289/2015). Permission to conduct the study was obtained from the Finnish Pensioners` Federation. The invitation mail requesting participation in the study included detailed information about the study. The decision of the subject to reply was considered as an informed consent.

## *Consent for publication*

Not applicable

## *Availability of data and materials*

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## *Competing interests*

Heikki Koskela has received funding from Kuopion Seudun Hengityssäitiö and Hengityssairauksien Tutkimussäitiö, has got payment for lectures from Boehringer Ingelheim Ltd and MSD Ltd, and owns shares of Orion Ltd.

Johanna Kaulamo has received funding from Suomen Tuberkuloosin Vastustamisyhdistyksen säitiö, Kuopion Seudun Hengityssäitiö, Hengityssairauksien Tutkimussäitiö, and Väinö ja Laina Kiven säitiö. She has got support for attending scientific meetings Boehringer Ingelheim Ltd.

Tuomas Selander has no competing interests

Anne Lätti has received funding from Kuopion Seudun Hengityssäitiö, Hengityssairauksien tutkimussäitiö, KYS tutkimussäitiö, Suomen Tuberkuloosin Vastustamisyhdistyksen säitiö, and Väinö ja Laina Kiven säitiö. She has got payments for lectures from Farmasian Oppimiskeskus and GlaxoSmithKline Ltd. She has got support for attending scientific meetings from Orion Ltd, Boehringer Ingelheim Ltd, and Roche Ltd

## *Funding*

The study was supported by Kuopion Seudun Hengityssäitiö (grant number not applicable) and Hengityssairauksien Tutkimussäitiö foundations (grant number not applicable). The funders had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

## *Authors' Contributions:*

HOK was mainly responsible for the conception and design of the work. He participated the collection of data and participated the analysis and interpretation of data. He mainly wrote the manuscript. He has approved the final version to be submitted.

JTK made substantial contributions to the conception and design of the work, participated the collection of data, and participated the analysis and interpretation of data. She has revised the manuscript critically for important intellectual content. She has approved the final version to be submitted.

TAS made substantial contributions to the conception and design of the work, mainly performed the statistical analysis, and participated the interpretation of data. He has revised the manuscript critically for important intellectual content. He has approved the final version to be submitted.

AML made substantial contributions to the conception and design of the work, participated the collection of data, and participated the analysis and interpretation of data. She has revised the manuscript critically for important intellectual content. She has approved the final version to be submitted.

### *Acknowledgments*

We thank Seppo Hartikainen (Istekki Oy, Kuopio, Finland) for his assistance in creating the electronic questionnaire.

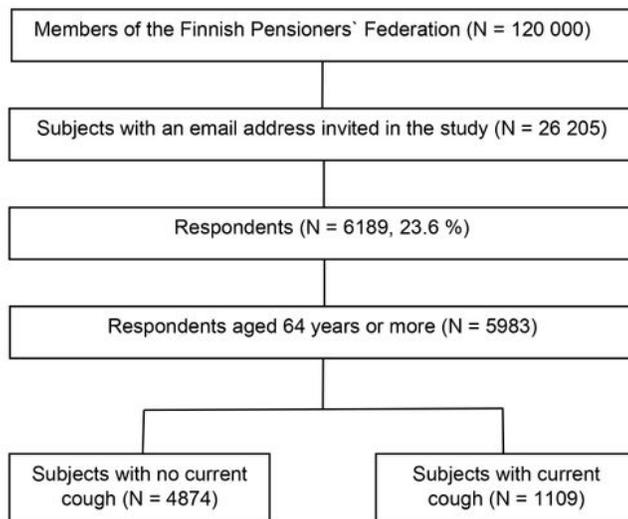
## **References**

1. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med.* 2012;18:716–725.
2. Sidhaye VK, Nishida K, Martinez FJ. Precision medicine in COPD: where are we and where do we need to go? *Eur Respir Rev.* 2018;27.
3. Morice AH, Millqvist E, Bieksiene K, Biring SS, Dicpinigaitis P, Domingo Ribas C, et al. ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. *Eur Respir J.* 2020;55:1901136.
4. Morice AH, McGarvey L, Pavord I, British Thoracic Society Cough Guideline Group. Recommendations for the management of cough in adults. *Thorax.* 2006;61 Suppl 1:i1-i24.
5. Irwin RS, French CL, Chang AB, Altman KW. Classification of Cough as a Symptom in Adults and Management Algorithms: CHEST Guideline and Expert Panel Report. *Chest.* 2018;153:196–209.
6. Shankar PS, Korukonda K, Bendre S, Behera D, Mirchandani L, Awad NT, et al. Diagnoses and management of adult cough: An Indian Environmental Medical Association (EMA) position paper. *Respir Med.* 2020;168:105949.
7. Song DJ, Song WJ, Kwon JW, Kim GW, Kim MA, Kim MY, et al. KAAACI Evidence-Based Clinical Practice Guidelines for Chronic Cough in Adults and Children in Korea. *Allergy Asthma Immunol Res.* 2018;10:591–613.

8. Gibson PG, Chang AB, Glasgow NJ, Holmes PW, Katelaris P, Kemp AS, et al. CICADA: Cough in Children and Adults: Diagnosis and Assessment. Australian cough guidelines summary statement. *Med J Aust*. 2010;192:265–271.
9. Kardos P, Dinh QT, Fuchs K, Gillissen A, Klimek L, Koehler M, et al. [Guidelines of the German Respiratory Society for Diagnosis and Treatment of Adults Suffering from Acute, Subacute and Chronic Cough]. *Pneumologie*. 2019;73:143–180.
10. Lai K, Shen H, Zhou X, Qiu Z, Cai S, Huang K, et al. Clinical Practice Guidelines for Diagnosis and Management of Cough-Chinese Thoracic Society (CTS) Asthma Consortium. *J Thorac Dis*. 2018;10:6314–6351.
11. Mukae H, Kaneko T, Obase Y, Shinkai M, Katsunuma T, Takeyama K, et al. The Japanese respiratory society guidelines for the management of cough and sputum (digest edition). *Respir Investig*. 2021;59:270–290.
12. Song W, Chang Y, Faruqi S, Kim J, Kang M, Kim S, et al. The global epidemiology of chronic cough in adults: a systematic review and meta-analysis. *Eur Respir J*. 2015;45:1479–1481.
13. Deliu M, Sperrin M, Belgrave D, Custovic A. Identification of Asthma Subtypes Using Clustering Methodologies. *Pulm Ther*. 2016;2:19–41.
14. Xu R, Wunsch DC. Clustering algorithms in biomedical research: a review. *IEEE Rev Biomed Eng*. 2010;3:120–154.
15. Koskela HO, Selander TA, Lätti AM. Cluster analysis in 975 patients with current cough identifies a phenotype with several cough triggers, many background disorders, and low quality of life. *Respir Res*. 2020;21:219.
16. Sa-Sousa A, Jacinto T, Azevedo LF, Morais-Almeida M, Robalo-Cordeiro C, Bugalho-Almeida A, et al. Operational definitions of asthma in recent epidemiological studies are inconsistent. *Clin Transl Allergy*. 2014;4:24–24. eCollection 2014.
17. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. European Position Paper on Rhinosinusitis and Nasal Polyps 2012. *Rhinol Suppl*. 2012;50:1–298.
18. El-Serag HB, Sweet S, Winchester CC, Dent J. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut*. 2014;63:871–880.
19. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology*. 2008;108:812–821.
20. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41:1284–1292.
21. Millqvist E, Bende M, Löwhagen O. Sensory hyperreactivity—a possible mechanism underlying cough and asthma-like symptoms. *Allergy*. 1998;53:1208–1212.
22. Millqvist E, Löwhagen O, Bende M. Quality of life and capsaicin sensitivity in patients with sensory airway hyperreactivity. *Allergy*. 2000;55:540–545.

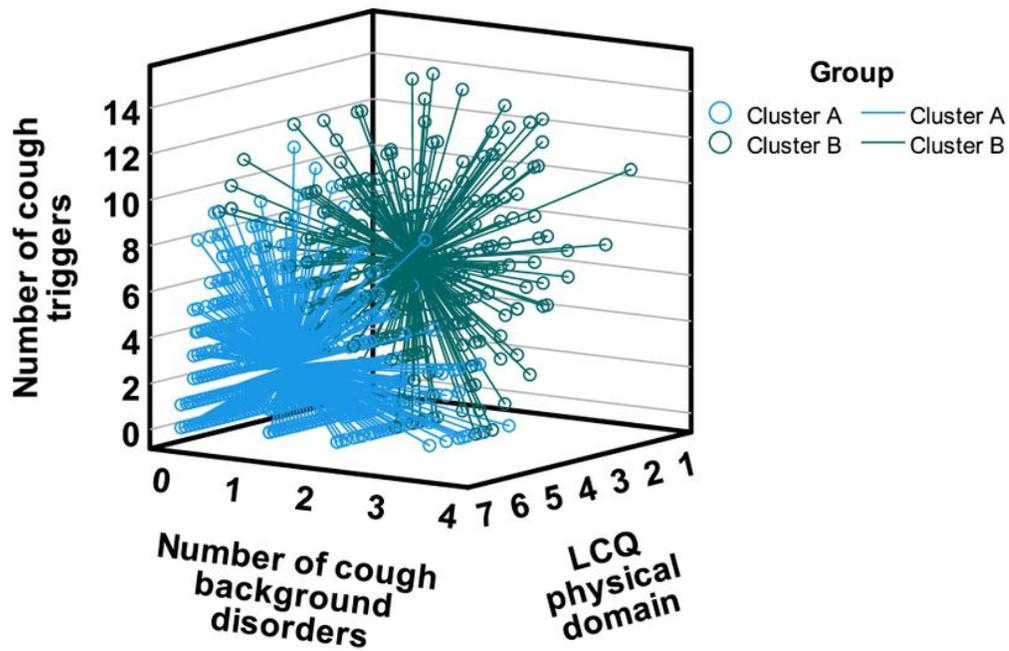
23. Johansson A, Löwhagen O, Millqvist E, Bende M. Capsaicin inhalation test for identification of sensory hyperreactivity. *Respir Med.* 2002;96:731–735.
24. Ternesten-Hasséus E, Farbrot A, Löwhagen O, Millqvist E. Sensitivity to methacholine and capsaicin in patients with unclear respiratory symptoms. *Allergy.* 2002;57:501–507.
25. Millqvist E, Johansson A, Bende M. Relationship of airway symptoms from chemicals to capsaicin cough sensitivity in atopic subjects. *Clin Exp Allergy.* 2004;34:619–623.
26. Millqvist E, Ternesten-Hasséus E, Ståhl A, Bende M. Changes in levels of nerve growth factor in nasal secretions after capsaicin inhalation in patients with airway symptoms from scents and chemicals. *Environ Health Perspect.* 2005;113:849–852.
27. Ternesten-Hasseus E, Johansson K, Lowhagen O, Millqvist E. Inhalation method determines outcome of capsaicin inhalation in patients with chronic cough due to sensory hyperreactivity. *Pulm Pharmacol Ther.* 2006;19:172–178.
28. Elberling J, Dirksen A, Johansen JD, Mosbech H. The capsaicin cough reflex in eczema patients with respiratory symptoms elicited by perfume. *Contact Derm.* 2006;54:158–164.
29. Holst H, Arendt-Nielsen L, Mosbech H, Vesterhauge S, Elberling J. The capsaicin cough reflex in patients with symptoms elicited by odorous chemicals. *Int J Hyg Environ Health.* 2010;213:66–71.
30. Pullerits T, Ternesten-Hasseus E, Johansson EL, Millqvist E. Capsaicin cough threshold test in diagnostics. *Respir Med.* 2014;108:1371–1376.
31. Morice AH, Millqvist E, Belvisi MG, Bieksiene K, Birring SS, Chung KF, et al. Expert opinion on the cough hypersensitivity syndrome in respiratory medicine. *Eur Respir J.* 2014;44:1132–1148.
32. Chung KF. Chronic 'cough hypersensitivity syndrome': a more precise label for chronic cough. *Pulm Pharmacol Ther.* 2011;24:267–271.
33. Morice AH. Chronic cough hypersensitivity syndrome. *Cough.* 2013;9:14.
34. Lee JW, Tay TR, Borg BM, Sheriff N, Vertigan A, Abramson MJ, et al. Laryngeal hypersensitivity and abnormal cough response during mannitol bronchoprovocation challenge. *Respirology.* 2022;27:48–55.
35. Song W, Morice AH, Kim M, Lee S, Jo E, Lee S, et al. Cough in the elderly population: relationships with multiple comorbidity. *PLoS One.* 2013;8:e78081.
36. Wei W, Yu L, Lu H, Wang L, Shi C, Ma W, et al. Comparison of cause distribution between elderly and non-elderly patients with chronic cough. *Respiration.* 2009;77:259–264.
37. Won H, Yoon S, Song W. The double-sidedness of cough in the elderly. *Respir Physiol Neurobiol.* 2018;257:65–69.
38. Kaulamo JT, Lätti AM, Koskela HO. Cough in the Elderly During the COVID-19 Pandemic. *Lung.* 2022;200:161–168.

## Figures



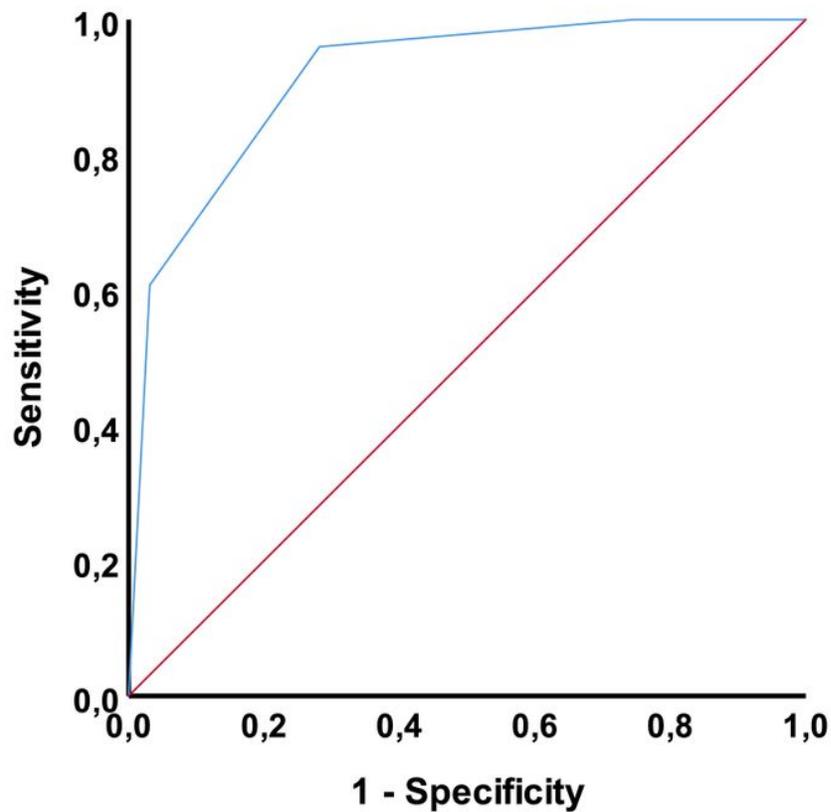
**Figure 1**

The flow chart.



**Figure 2**

Each patient of the cluster A (N = 834, blue color) and cluster B (N = 275, green color) represented in a 3-dimensional figure according to the number of the cough triggers, the number of the cough background disorders, and the Leicester Cough Questionnaire (LCQ) physical domain. The marker of every patient is connected to the cluster mean value by a spike.



**Figure 3**

A receiver operating characteristic curve (ROC) constructed to demonstrate the separation of the two clusters by the number of the three main determinants (trigger sum  $\geq 5$ , LCQ physical domain  $\leq 4.9$ , at least one cough background disorder). Area under the ROC curve is 0.92 (95 % CI 0.90 – 0.93).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ElderlyQuestionnaireEnglishversion03012022.docx](#)