



Validation of the cough phenotype TBQ among elderly Finnish subjects

Heikki O. Koskela ^{1,2}, Johanna T. Kaulamo², Tuomas A. Selander³ and Anne M. Lätti ¹

¹Unit for Medicine and Clinical Research, Pulmonary Division, Kuopio University Hospital, Kuopio, Finland. ²School of Medicine, University of Eastern Finland, Kuopio, Finland. ³Science Service Center, Kuopio University Hospital, Kuopio, Finland.

Corresponding author: Heikki Koskela (heikki.koskela@kuh.fi)



Shareable abstract (@ERSpublications)

The previously described cough phenotype TBQ (triggers, background disorders, quality-of-life impairment) and the common cough phenotype can be identified in elderly, retired subjects with current cough <https://bit.ly/3cygez0>

Cite this article as: Koskela HO, Kaulamo JT, Selander TA, *et al.* Validation of the cough phenotype TBQ among elderly Finnish subjects. *ERJ Open Res* 2022; 8: 00284-2022 [DOI: 10.1183/23120541.00284-2022].

Copyright ©The authors 2022

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 9 June 2022
Accepted: 23 Aug 2022

Abstract

Background Phenotypes can be utilised in the clinical management of disorders. 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 analysis included 1109 subjects with current cough (mean±SD age 72.9±5.3 years; 67.7% female). All filled in a comprehensive 86-item questionnaire including the Leicester Cough Questionnaire. Phenotypes were identified utilising k-means partitional clustering.

Results Two clusters were identified. Cluster A included 75.2% of the subjects and cluster B 24.8% of the subjects. The three most important variables to separate the clusters were the number of cough triggers (mean±SD 2.47±2.34 *versus* 7.08±3.16, respectively; p<0.001), Leicester Cough Questionnaire physical domain (5.38±0.68 *versus* 4.21±0.81, respectively; p<0.001) and the number of cough background disorders (0.82±0.78 *versus* 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 common cough phenotype is probably hypersensitivity of the cough reflex arc.

Introduction

Phenotype indicates a single or combination of disease attributes that describe differences between individuals [1]. Recently, approaches to phenotype disorders have evolved from subjective expert opinion to data-driven methodologies like clustering [2]. These methods explore the data through an unsupervised separation of a dataset with little or no ground truth, into a discrete set of hidden data structures [3]. This contrasts with the traditional methods based on human observation and testing of hypotheses using prior knowledge.

Cough is usually classified according to the length of the episode: acute (<3 weeks), subacute (3–8 weeks) or chronic (>8 weeks) [4–12]. However, this classification is one-dimensional and not based on data-driven analyses. We recently performed a cluster analysis in 975 working-age subjects with cough [13]. Two clusters were identified. The smaller one was especially characterised by several triggers of cough, many cough background disorders, and poor cough-related quality of life (TBQ). Those with the phenotype TBQ showed a high tendency for cough prolongation in the follow-up survey 12 months later [13]. The present study was carried out to validate that analysis in a different population, namely in aged, retired subjects, and under different conditions, namely in the middle of the coronavirus disease 2019 (COVID-19) pandemic.



Material and methods

Study design, setting and population

This was an observational, cross-sectional study conducted *via* email among the members of the Finnish Pensioners' Federation. The sample size assessment was based on knowledge about the required sample size for cluster analysis, the response rates in email studies and the prevalence of cough in Finland [14–16]. The 26 205 members (mean age 72.7 years, 63.5% females) who had an email address were sent an invitation to participate along with information about the study. A pre-notification email was sent. The electronic questionnaires were sent as a hyperlink in an email in April 2021 and one reminder was sent 2 weeks later. The responses were recorded in an electronic datasheet.

The study was approved by the ethics committee of Kuopio University Hospital (289/2015; Kuopio, Finland). Permission to conduct the study was obtained from the Finnish Pensioners' Federation. 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 [13]. The 86 questions dealt with social background, lifestyle, general health, doctors' diagnoses and visits, and medications. Appropriate symptom questions for asthma, chronic rhinosinusitis, gastro-oesophageal reflux disease (GORD), obstructive sleep apnoea (OSA) and depressive symptoms were included [17–21]. Respondents with current cough were asked to answer additional cough-related questions, including the Leicester Cough Questionnaire (LCQ). For the present study, new questions were added about symptoms of 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. An English version of the questionnaire can be found in the supplementary material.

Definitions of variables derived from the raw data

Acute, subacute and chronic cough were defined as suggested in international guidelines [4–12]. Current asthma was defined as doctor's diagnosis of asthma at any age and wheezing during the past 12 months [17]. 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 >3 months [18]. GORD was defined as heartburn and/or regurgitation on ≥ 1 day per week during the past 3 months [19]. OSA was defined as presence of two or more of the following features: arterial hypertension, loud snoring, daytime somnolence or observed apnoeas [20]. 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 them up, 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 [21]. Symptom sum was calculated by summing up all reported symptoms except those associated with airway disorders, giving a value from zero to 15. Trigger sum was calculated by summing up 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 >8 weeks) cough in parents, sisters or brothers.

Statistical analysis

All questions of the questionnaire plus the derived variables were included in partitional clustering using the k-means method [3] similarly to our previous cluster analysis [13]. 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.

First, data were pre-processed to transform all variables to the same scale (0–1). Right-skewed (skewness >1) variables were normalised with the $\log(x+1)$ function, since zero values cannot be log-transformed. Ordinal and continuous variables were scaled into 0–1 intervals. Each variable's minimum value or the lowest class was assigned a value of 0 and the maximum value or the highest class was assigned a value of 1. Binary variables remained unchanged. Value 0 indicated the negative or "no" alternative and value 1 the positive or "yes" alternative.

Second, clustering was applied. A distance matrix between observations with scaled variables were calculated using the Manhattan distance function. Diffusion maps dimension reduction algorithm was applied to extract diffusion coordinates from distance matrix, using default settings of the software. The number of clusters was evaluated by the 24 criteria provided by the software. After that, the extracted

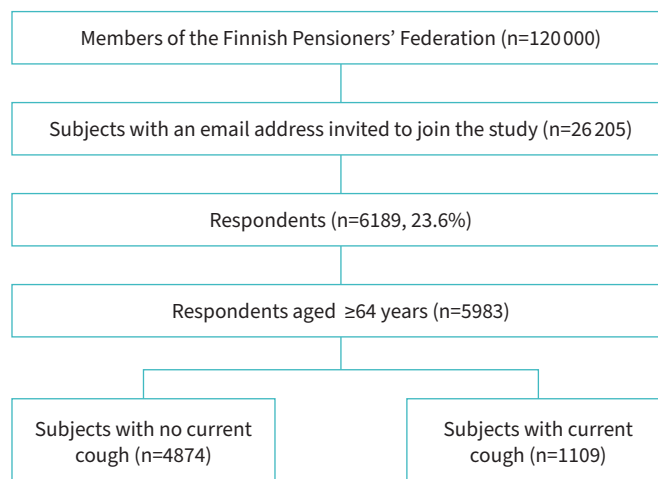


FIGURE 1 Study flow chart.

diffusion map coordinates were clustered into groups using the k-means method. To validate the clustering, the analyses were repeated by excluding those background variables with no plausible biological association with cough (such as hometown, years of education, alcohol consumption, etc.).

Third, cluster membership was added to original data to compare the clusters, utilising the Mann–Whitney U-test or Chi-squared test. The interrelationships of the variables were analysed using the Spearman’s correlation coefficient (r_s) using SPSS software (SPSS Statistics for Windows, version 22.0; IBM, Armonk, NY, USA). Receiver operating characteristic curves (ROC) and the Youden index were utilised to define the cut-off values. The values are expressed by either mean \pm SD, median (interquartile range) or percentages. A p-value <0.05 was accepted as the level of statistical significance.

Results

23.6% of the subjects responded (6189 respondents, mean \pm SD age 72.2 \pm 5.5 years, 66.4% female) (figure 1). 206 respondents were excluded from the analyses because of age <64 years. Of the remaining 5983, 1109 subjects suffered from current cough. They formed the population in which the clustering was applied. Their mean age was 72.9 \pm 5.3 years and 67.7% were female. The proportion of missing values was <2.5%, except for the two OSA-related 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, 11 or 14 clusters, and two suggested 15 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 clusters in acute, subacute, and chronic cough is presented in table 1. Cluster B was represented in all subtypes, although its proportion was largest in chronic cough.

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 interest. Cluster B was especially characterised by

TABLE 1 Distribution of the clusters in acute (duration <3 weeks), subacute (3–8 weeks) and chronic (>8 weeks) cough

	Acute cough	Subacute cough	Chronic cough
Subjects	207	70	809
Cluster A (“common” phenotype)	169 (81.6)	57 (81.4)	587 (72.6)
Cluster B (TBQ phenotype)	38 (18.4)	13 (18.6)	222 (27.4)

Data are presented as n or n (%). 23 subjects could not define the length of the cough episode. TBQ: triggers, background disorders, quality-of-life impairment.

TABLE 2 The clusters and their defining variables among 1109 subjects with current cough

Variable	Cluster A	Cluster B	p-value
Subjects n	834	275	
Order			
1 Trigger sum	2.47±2.34	7.08±3.16	2.39×10 ⁻⁷⁸
2 LCQ physical domain	5.38±0.68	4.21±0.81	1.02×10 ⁻⁷⁵
3 Number of cough background disorders	0.82±0.78	1.99±0.89	9.09×10 ⁻⁶⁵
4 LCQ total score	16.1±2.48	12.5±2.79	7.12×10 ⁻⁶⁴
5 Dyspnoea with wheezing	8.4	55.0	3.38×10 ⁻⁶¹
6 Paints or fumes as a cough trigger	19.2	70.5	5.03×10 ⁻⁵⁶
7 Current asthma	6.4	46.5	2.36×10 ⁻⁵³
8 LCQ social domain	5.63±1.00	4.38±1.12	1.47×10 ⁻⁵²
9 Automobile exhaust fumes as a cough trigger	13.4	59.3	4.37×10 ⁻⁵²
10 LCQ question 3 (tired because of cough) [#]	5.93±1.07	4.57±1.36	2.36×10 ⁻⁴⁸
23 Symptom sum	2.32±1.71	4.13±2.17	9.67×10 ⁻³⁵
30 LCQ question 2 (sputum production when coughing) [#]	4.86±1.84	3.29±1.78	2.20×10 ⁻³⁰
40 Unexplained cough	38.0	4.0	1.93×10 ⁻²⁶
46 Chronic rhinosinusitis	13.9	41.8	1.15×10 ⁻²²
47 Allergy	9.1	33.1	1.37×10 ⁻²¹
51 Number of doctor's visits due to cough in the past 12 months	0.40±1.15	1.24±2.19	4.49×10 ⁻¹⁹
53 Presence of any trigger for cough	71.8	96.7	1.11×10 ⁻¹⁷
56 Speaking as a cough trigger	16.4	40.4	2.98×10 ⁻¹⁶
61 Family history of chronic cough	31.8	58.0	2.03×10 ⁻¹⁴
70 Gastro-oesophageal reflux disease	22.3	43.8	8.87×10 ⁻¹²
77 Sleep apnoea	39.9	63.4	4.20×10 ⁻¹¹
85 Symptoms of flu at the beginning of the current cough	13.9	30.7	6.65×10 ⁻¹⁰
86 Number of cough episodes in the past 12 months	4.92±7.90	7.39±8.98	1.03×10 ⁻⁹
105 Duration of the cough episode	1–5 years (2–12 months to 1–5 years)	1–5 years (1–5 years to 5–10 years)	1.29×10 ⁻⁵
107 Use of cough medicines in the past 12 months	28.2	42.4	1.82×10 ⁻⁵
114 Female	64.7	76.7	0.0003
115 Autoimmune disorder	18.2	28.0	0.0007
121 COPD	2.3	6.2	0.0029
124 Bronchiectasis	1.3	4.4	0.0046
138 Body mass index (kg·m ⁻²)	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

Data are presented as either mean±SD, % or median (interquartile range), unless otherwise stated. The 10 most important variables and 23 variables of special interest expressed, in order of importance, as defined by the p-value obtained by Mann–Whitney U-test or Chi-squared test between the clusters. LCQ: Leicester Cough Questionnaire. #: LCQ questions have seven-step scales, from 1 (all of the time) to 7 (none of the time).

several cough triggers, many cough background disorders, and low LCQ scores (figures 2 and 3). Of the various cough triggers, paints and 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 3 presents the best cut-off values for the 10 most important variables to identify 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 ≥5, LCQ physical domain ≤4.9, at least one cough

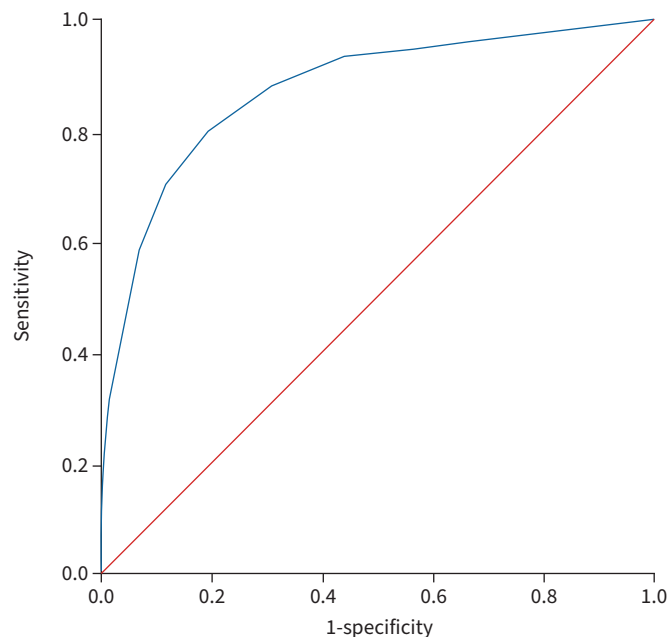


FIGURE 2 A receiver operating characteristic (ROC) curve constructed to demonstrate the separation of the two clusters by the number of cough triggers. Area under the ROC curve is 0.87 (95% CI 0.85–0.90).

background disorder) to separate the clusters (figure 4). The presence of at least two main determinants gave the best Youden index with a sensitivity of 0.96 and specificity of 0.72. The presence of all three main determinants gave a sensitivity of 0.61 and specificity of 0.97.

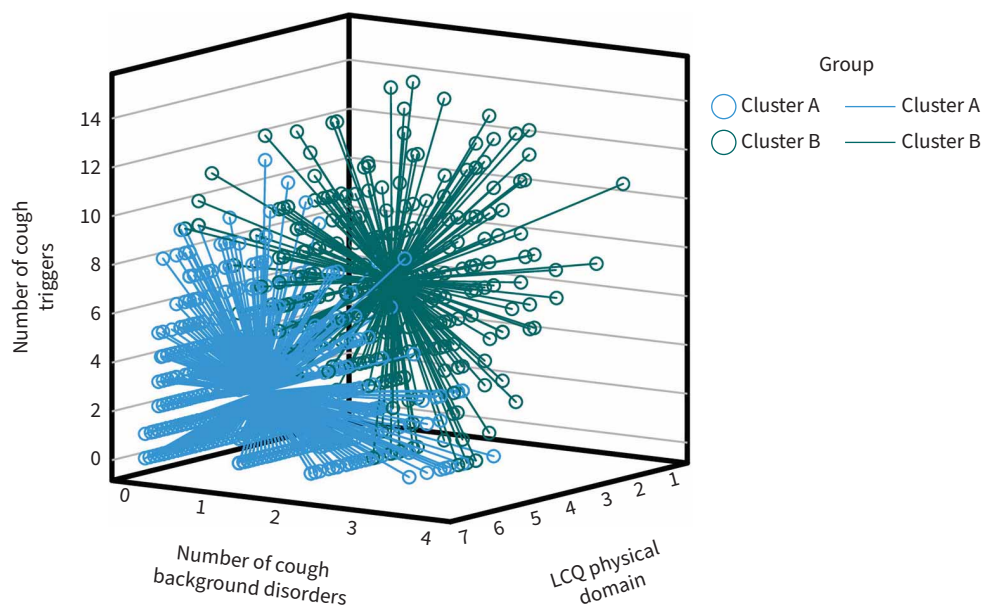


FIGURE 3 Each subject of cluster A ($n=834$) and cluster B ($n=275$) represented in a three-dimensional scatterplot. The number of the cough triggers is represented in the vertical axis, the number of the cough background disorders in the horizontal axis, and the Leicester Cough Questionnaire (LCQ) physical domain in the axis pointing to the observer. The marker of every subject is connected to the cluster mean value by a spike. Most of the subjects aggregate close to the cluster mean value where the markers of each subject are difficult to distinguish. The distances between the individuals are shorter within the cluster than those between the clusters.

TABLE 3 Sensitivity and specificity of the 10 main variables to identify cluster B, utilising the cut-off values giving the best sum of sensitivity and specificity (the Youden index)

	Cut-off value	Sensitivity	Specificity	aROC
Trigger sum	≥ 5	0.80	0.81	0.87
LCQ physical domain	≤ 4.9	0.75	0.81	0.87
Number of cough background disorders	≥ 1	0.72	0.82	0.82
LCQ total score	≤ 14.7	0.76	0.76	0.84
Dyspnoea 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	≤ 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) [#]	≤ 5	0.73	0.70	0.78

aROC: area under the receiver operator characteristic curve; LCQ: Leicester Cough Questionnaire; NA: not applicable. [#]: LCQ questions have seven-step scales from 1 (all of the time) to 7 (none of the time).

Belonging to cluster B increased the likelihood of at least one doctor's visit due to cough in the past 12 months (OR 3.39, 95% CI 2.53–4.55) and the likelihood of having used cough medicines in the past 12 months (OR 1.88, 95% CI 1.41–2.50). The population was also divided according to the length of the cough episode. Presence of chronic cough (>8 weeks' duration) slightly increased the likelihood of doctors' visits (OR 1.91, 95% CI 1.36–2.70), but decreased the likelihood of using cough medicines (OR 0.66, 95% CI 0.50–0.89).

The validation analysis, 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 dyspnoea with wheezing (data not shown).

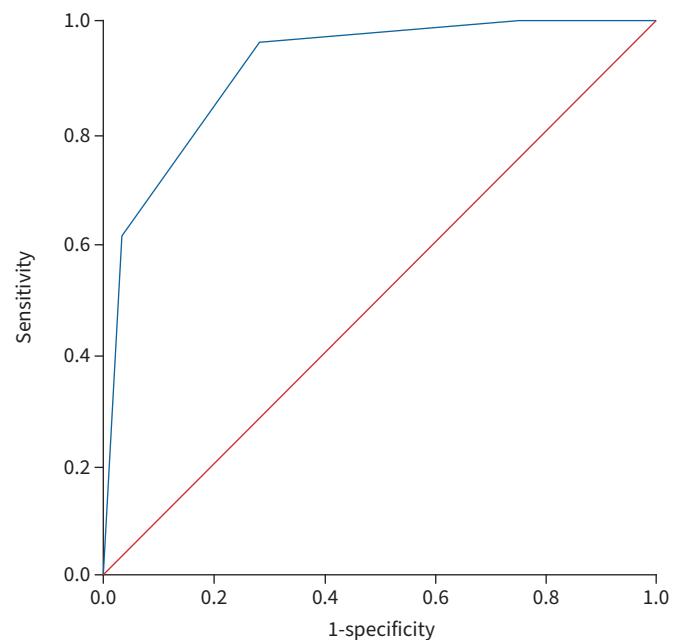


FIGURE 4 A receiver operating characteristic (ROC) curve constructed to demonstrate the separation of the two clusters by the number of the three main determinants (trigger sum ≥ 5 , Leicester Cough Questionnaire physical domain ≤ 4.9 , at least one cough background disorder). Area under the ROC curve is 0.92 (95% CI 0.90–0.93).

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 clustering, which was performed during the COVID-19 pandemic among 1109 elderly, retired subjects with current cough, validates our previous clustering among working-age, employed subjects [13]. Again, two clusters were found. Cluster B, consisting of 24.8% of the subjects, was especially characterised by several cough triggers, many cough background disorders, and poor cough-related quality of life. These features fit the cough phenotype TBQ, which was identified in our previous study. Cluster A, lacking these features, may be called the “common” cough phenotype.

Clustering is a task of grouping subjects in such a way that subjects in the same group (cluster) are more like each other than those in other groups. This is achieved by measuring the distances between the subjects with respect to each variable. These distances are summed and placed on a two-dimensional table representing every possible pair of subjects, called the distance matrix. Different algorithms can then be applied to recognise the clusters of subjects with small distances to each other. Dimension reduction is often necessary to improve the observation to variable ratios, which makes the analysis more reliable.

Both cluster analyses identified the number of cough triggers as the most important variable to separate the phenotypes. Both analyses also identified chemical triggers such as paints, fumes and strong scents as the most important types of triggers [13]. Several studies have shown that subjects with lower airway symptoms induced by chemical irritants are especially sensitive to the cough-provocation test with capsaicin [22–31]. Therefore, we hypothesise that the main underlying pathophysiological feature separating the phenotype TBQ from the common phenotype is hypersensitivity of the cough reflex. Thus, this phenotype might also represent a distinct endotype. Subjects with TBQ phenotype might especially benefit from medication that can decrease the sensitivity of the cough reflex. If that is the case, the phenotype TBQ could also provide a treatable trait. To investigate whether the phenotype TBQ is a distinct cough genotype, studies applying genome analyses should be performed. Of note, family history of chronic cough was far more common in the phenotype TBQ than in the common phenotype.

The common cough phenotype seems to be less associated with the cough reflex hypersensitivity than the phenotype TBQ. Other mechanisms, like excessive mucus production, may be more important in that phenotype [32].

The phenotype TBQ was 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 [13]. Since the phenotype TBQ is related to clinically meaningful outcomes, it fulfils the criteria for clinical phenotype [1]. Its identification may serve as an indication for prompt and comprehensive clinical evaluation regardless of the duration of the cough episode.

The phenotype TBQ resembles the concept “cough hypersensitivity syndrome”, introduced by experienced clinicians [33]. Both emphasise the enhanced response to cough triggers. However, there seems to be two major discrepancies between the entities. First, cough hypersensitivity syndrome has been connected to chronic cough [33], 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 [33–35], whereas in the present 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 [33], 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. Despite the aforementioned differences, it is remarkable that unsupervised, data-driven analyses lead to similar conclusions to those drawn from clinical experience.

A recent study from Australia supports the present analysis. In that study, two clusters could be identified among subjects with various respiratory symptoms. The smaller cluster was characterised by symptoms of laryngeal hypersensitivity and a strong cough response to mannitol [36]. The characteristics of that cluster resemble those of the phenotype TBQ.

For clinical purposes, we calculated the best cut-off values for the most important variables to separate the clusters. They were almost identical to those reported in our previous cluster analysis [13]. Presence of at least two of the three main TBQ determinants 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 material).

There were slight differences in the questionnaires between the present and the previous study [13], which did not affect the main results. It has been shown that there are more cough background disorders in elderly than in younger subjects [37–39]. Therefore, bronchiectasis, pulmonary fibrosis and OSA were added to the variable “number of cough background disorders”. 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. We also asked how many cough episodes the subject had had in the past 12 months. This number was significantly higher in the phenotype TBQ than in the common phenotype. Of note, the number of the cough episodes was more strongly associated with the phenotype TBQ than the length of the current cough episode.

The present study has several limitations. The participation rate was relatively low, which is typical for academic email surveys [15] and of similar magnitude to our previous cluster analysis [13]. However, the age and gender distribution of the responders did not differ significantly from the original population. The high age of the population may have hindered the use of email in some individuals. 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 overrepresentation 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 recommended during the pandemic era [39]. Finally, the analysis was based on the questionnaire data only; spirometry, laboratory and radiography data were missing.

The strengths of the present study included a large population, which was missing in our previous cluster analysis: elderly, retired subjects [13]. The survey was not limited to cough patients, but to a community-based population. Therefore, even those coughing subjects who would never complain about their cough to doctors were also included. The questionnaire was originally planned and further completed to investigate cough and associated conditions. It included a comprehensive list of cough triggers with both external triggers and those representing allotussia [40]. The interindividual variation in how subjects recognise and report symptoms was controlled by the variable “symptom sum”. Furthermore, the symptom questions to define important background disorders were those recommended for epidemiological studies. All raw data plus the derived variables were included in the cluster analysis without hypotheses using prior knowledge. In this way, all relevant features of cough and even undiagnosed but symptomatic background disorders were equally considered in the 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 [13]. The phenotype TBQ was associated with frequent doctor’s visits due to cough, use of cough medicines and a high tendency for cough prolongation [1, 13]. Clinical evaluation of a patient with cough should probably focus on the presence of cough triggers, background disorders and the quality of life rather than on the length of the cough episode.

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

Provenance: Submitted article, peer reviewed.

Support statement: The study was supported by Kuopion Seudun Hengityssäätö and Hengityssairauksien Tutkimussäätiö foundations. 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. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: H.O. Koskela received support for the present manuscript from Kuopion Seudun Hengityssäätiö and Hengityssairauksien tutkimussäätiö; outside the submitted, work the following relationships have been reported: payment for lectures received from Boehringer Ingelheim Ltd and MSD Ltd, and stock owned in Orion Ltd. J.T. Kaulamo received support for the present manuscript from Foundation of the Finnish Anti-Tuberculosis Association/Suomen Tuberkuloosin vastustamisyhdistyksen Säätiö, Respiratory Foundation of Kuopio Region/Kuopion Seudun Hengityssäätiö, Research Foundation of the Pulmonary diseases/Hengityssairauksien tutkimussäätiö, and Väinö and Laina Kivi Foundation/Väinö ja Laina Kiven Säätiö; outside the submitted work, the following relationships have been reported: support for attending a scientific meeting received from Boehringer Ingelheim. T.A. Selander has nothing to disclose. A.M. Lätti received support for the present manuscript from Kuopion Seudun Hengityssäätiö, Hengityssairauksien tutkimussäätiö, KYS:n tutkimussäätiö, Suomen Tuberkuloosin Vastustamisyhdistyksen säätiö sr and Väinö ja Laina Kiven Säätiö; outside the submitted work, the following relationships have been reported: payment for lectures received from Farmasian oppimiskeskus/Pharmaceutical Learning Centre, payment for lectures received from GlaxoSmithKline, payment for group input meeting received from MSD, and support for attending scientific meetings received from Orion Oyj, Boehringer Ingelheim and Roche.

References

- 1 Han MK, Agusti A, Calverley PM, *et al.* Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med* 2010; 182: 598–604.
- 2 Deliu M, Sperrin M, Belgrave D, *et al.* Identification of asthma subtypes using clustering methodologies. *Pulm Ther* 2016; 2: 19–41.
- 3 Xu R, Wunsch DC. Clustering algorithms in biomedical research: a review. *IEEE Rev Biomed Eng* 2010; 3: 120–154.
- 4 Morice AH, Millqvist E, Bieksiene K, *et al.* ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. *Eur Respir J* 2020; 55: 1901136.
- 5 Morice AH, McGarvey L, Pavord I, *et al.* Recommendations for the management of cough in adults. *Thorax* 2006; 61: Suppl. 1, i1–i24.
- 6 Irwin RS, French CL, Chang AB, *et al.* Classification of cough as a symptom in adults and management algorithms: CHEST guideline and expert panel report. *Chest* 2018; 153: 196–209.
- 7 Shankar PS, Korukonda K, Bendre S, *et al.* Diagnoses and management of adult cough: an Indian Environmental Medical Association (EMA) position paper. *Respir Med* 2020; 168: 105949.
- 8 Song DJ, Song WJ, Kwon JW, *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.
- 9 Gibson PG, Chang AB, Glasgow NJ, *et al.* CICADA: cough in children and adults: diagnosis and assessment. Australian cough guidelines summary statement. *Med J Aust* 2010; 192: 265–271.
- 10 Kardos P, Dinh QT, Fuchs KH, *et al.* Leitlinie der Deutschen Gesellschaft für Pneumologie und Beatmungsmedizin zur Diagnostik und Therapie von erwachsenen Patienten mit Husten. [Guidelines of the German respiratory society for diagnosis and treatment of adults suffering from acute, subacute and chronic cough]. *Pneumologie* 2019; 73: 143–180.
- 11 Lai K, Shen H, Zhou X, *et al.* Clinical practice guidelines for diagnosis and management of cough – Chinese Thoracic Society (CTS) Asthma Consortium. *J Thorac Dis* 2018; 10: 6314–6351.
- 12 Mukae H, Kaneko T, Obase Y, *et al.* The Japanese Respiratory Society guidelines for the management of cough and sputum (digest edition). *Respir Investig* 2021; 59: 270–290.
- 13 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.
- 14 Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *BMC Bioinformatics* 2022; 23: 205.
- 15 Sheehan KB. E-mail survey response rates: a review. *JCMC* 2001; 6: JCMC621.
- 16 Lätti AM, Pekkanen J, Koskela HO. Defining the risk factors for acute, subacute and chronic cough: a cross-sectional study in a Finnish adult employee population. *BMJ Open* 2018; 8: e022950.
- 17 Sá-Sousa A, Jacinto T, Azevedo LF, *et al.* Operational definitions of asthma in recent epidemiological studies are inconsistent. *Clin Transl Allergy* 2014; 4: 24.
- 18 Fokkens WJ, Lund VJ, Mullol J, *et al.* European position paper on rhinosinusitis and nasal polyps 2012. *Rhinol Suppl* 2012; 50: 1–298.
- 19 El-Serag HB, Sweet S, Winchester CC, *et al.* Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2014; 63: 871–880.
- 20 Chung F, Yegneswaran B, Liao P, *et al.* STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology* 2008; 108: 812–821.
- 21 Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003; 41: 1284–1292.
- 22 Millqvist E, Bende M, Löwhagen O. Sensory hyperreactivity – a possible mechanism underlying cough and asthma-like symptoms. *Allergy* 1998; 53: 1208–1212.

- 23 Millqvist E, Löwhagen O, Bende M. Quality of life and capsaicin sensitivity in patients with sensory airway hyperreactivity. *Allergy* 2000; 55: 540–545.
- 24 Johansson A, Löwhagen O, Millqvist E, *et al.* Capsaicin inhalation test for identification of sensory hyperreactivity. *Respir Med* 2002; 96: 731–735.
- 25 Ternesten-Hasséus E, Farbrot A, Löwhagen O, *et al.* Sensitivity to methacholine and capsaicin in patients with unclear respiratory symptoms. *Allergy* 2002; 57: 501–507.
- 26 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.
- 27 Millqvist E, Ternesten-Hasséus E, Ståhl A, *et al.* 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.
- 28 Ternesten-Hasséus E, Johansson K, Löwhagen O, *et al.* Inhalation method determines outcome of capsaicin inhalation in patients with chronic cough due to sensory hyperreactivity. *Pulm Pharmacol Ther* 2006; 19: 172–178.
- 29 Elberling J, Dirksen A, Johansen JD, *et al.* The capsaicin cough reflex in eczema patients with respiratory symptoms elicited by perfume. *Contact Dermatitis* 2006; 54: 158–164.
- 30 Holst H, Arendt-Nielsen L, Mosbech H, *et al.* The capsaicin cough reflex in patients with symptoms elicited by odorous chemicals. *Int J Hyg Environ Health* 2010; 213: 66–71.
- 31 Pullerits T, Ternesten-Hasséus E, Johansson EL, *et al.* Capsaicin cough threshold test in diagnostics. *Respir Med* 2014; 108: 1371–1376.
- 32 Koskela HO, Nurmi HM, Birring SS. Utility of cough provocation tests in chronic cough and respiratory diseases: a comprehensive review and introduction of new reference ranges for the capsaicin test. *Allergy Asthma Immunol Res* 2021; 13: 833–849.
- 33 Morice AH, Millqvist E, Belvisi MG, *et al.* Expert opinion on the cough hypersensitivity syndrome in respiratory medicine. *Eur Respir J* 2014; 44: 1132–1148.
- 34 Chung KF. Chronic ‘cough hypersensitivity syndrome’: a more precise label for chronic cough. *Pulm Pharmacol Ther* 2011; 24: 267–271.
- 35 Morice AH. Chronic cough hypersensitivity syndrome. *Cough* 2013; 9: 14.
- 36 Lee JW, Tay TR, Borg BM, *et al.* Laryngeal hypersensitivity and abnormal cough response during mannitol bronchoprovocation challenge. *Respirology* 2022; 27: 48–55.
- 37 Song W, Morice AH, Kim M, *et al.* Cough in the elderly population: relationships with multiple comorbidity. *PLoS One* 2013; 8: e78081.
- 38 Wei W, Yu L, Lu H, *et al.* Comparison of cause distribution between elderly and non-elderly patients with chronic cough. *Respiration* 2009; 77: 259–264.
- 39 Kaulamo JT, Lähti AM, Koskela HO. Cough in the elderly during the COVID-19 pandemic. *Lung* 2022; 200: 161–168.
- 40 Mazzone SB, McGarvey L. Mechanisms and rationale for targeted therapies in refractory and unexplained chronic cough. *Clin Pharmacol Ther* 2021; 109: 619–636.