ERJ Express. Published on December 22, 2010 as doi: 10.1183/09031936.00130910

The impact of using different symptom-based exacerbation algorithms in patients with COPD

Jaap (C.A.) Trappenburg[1*§], Aukje (C.) van Deventer[1*], Thierry Troosters[2*], Theo (J.M.) Verheij[1*], Augustinus (J.P.) Schrijvers[1*], Jan-Willem (J.) Lammers[3*], Evelyn (M.) Monninkhof[1*]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.

[2] Department of Rehabilitation and Respiratory Division, University Hospital Gasthuisberg, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

[3]Department of Respiratory Medicine, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.

*These authors contributed equally to this work

[§]Corresponding author

Email addresses:

JCAT:	J.C.A.trappenburg@umcutrecht.nl
ACD:	a.c.vandeventer@gmail.com
TT:	thierry.troosters@med.kuleuven.be
TV:	T.J.M.Verheij@umcutrecht.nl
AJP:	A.J.P.Schrijvers@umcutrecht.nl
JWJL:	J.W.J.Lammers@umcutrecht.nl
EMM:	E.Monninkhof@umcutrecht.nl

Corresponding author:

Jaap.C.A. Trappenburg, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.

Telephone: +31 887568123

Fax: +31 887555482

Key words: COPD / COPD exacerbations

The impact of using different symptom-based exacerbation algorithms in patients with COPD

Jaap (C.A.) Trappenburg, Aukje (C.) van Deventer, Thierry Troosters, Theo (J.M.) Verheij, Augustinus (J.P.) Schrijvers, Jan-Willem (J.) Lammers, Evelyn (M.) Monninkhof

ABSTRACT

Not all exacerbations are captured by reliance on healthcare contacts. Symptom-based exacerbation definitions have shown to provide more adequate measures of exacerbation rates, severity and duration. Still, no consensus is reached on what is the most useful method and algorithm to identify these events. This article provides an overview of the existing symptom-based definitions and tested the hypothesis that differences in exacerbation characteristics depend on the algorithms used.

We systematically reviewed symptom-based methods and algorithms used in the literature and quantified the impact of the four most referenced algorithms on exacerbation-related outcome using an existing COPD cohort (n=137).

We identified 51 studies meeting our criteria using 14 widely varying symptom algorithms to define onset, severity and recovery. The most (71%) frequently referenced algorithm (modified Anthonisen) identified an incidence rate of 1.7 (1.4 -2.1) episodes per patient-year, while for requiring only one major or two major symptoms this was 1.9 (1.6-2.3) and 1.5 (0.6-1.0) respectively. Studies were generally lacking methods to enhance validity and accuracy of symptom recording.

This review revealed large inconsistencies in definitions, methods and accuracy to define symptom-based COPD exacerbations. We demonstrated that minor changes in symptom criteria substantially affect incidence rates, clustering type and classification of exacerbations.

INTRODUCTION

COPD is a highly prevalent disease and a major cause of mortality and morbidity [1]. The natural course of COPD is interrupted by periods (exacerbations) characterized by a sustained change of patients baseline symptoms which are beyond normal day to day variability and may warrant medical treatment[2]. These exacerbations are important, since they have a serious negative impact on health-related quality of life[3], are associated with accelerated lung function decline[4] and increased mortality[5]. In addition, they represent a significant economic burden due to health-care utilization[6]. The relevance of exacerbations from a clinical, patient and societal perspective, resulted in selecting exacerbation rates as the main outcome parameter in an increasing number of trials[7]. It is surprising that, there is no consensus on the exact operational definition of exacerbations used in studies. Defining a generally accepted standard definition is difficult because there is a large variation in aetiology, type and severity of symptoms between individuals[8].

Several potential approaches can be taken to defining exacerbations. The majority of recently published studies have used event-based exacerbation definitions, i.e. based on increased use of health care services (increased use of reliever medication or treatment with systemic corticosteroids antibiotics or hospitalisation) in the presence of a worsening condition of the patient [2]. Although simple counting of events by healthcare utilization intuitively seems a straightforward and robust approach, it strongly depends on the ability of patients to recognize exacerbations and available healthcare facilities. Therefore, event-based definitions underestimate true exacerbation rates by up to 50%[9-12]. A symptom-based approach takes a patient oriented perspective as it relies on patients experiencing an increase in symptoms for a minimal number of consecutive days, mostly assessed by daily diary registrations. These definitions have shown to capture exacerbations which remained unreported while having substantial and non-negligible negative impact on annual change in health status[12]. This

approach has several challenges. First, due to the heterogeneous nature of COPD and its exacerbations, a standardised symptom-based definition is lacking. This resulted in use of several different symptom-based algorithms and methods to assess onset rates, recovery, severity and recurrence of exacerbations[13]. Apart from the recent promising EXACT (EXAcerbations of Chronic Pulmonary Disease Tool) initiative[14, 15], few of the symptom scores have originated perspective and would meet current criteria for 'Patient Reported Outcomes' (PRO's)[16]. Second, it is unknown whether the clinical and societal severity of an exacerbation is appropriately reflected in this symptom-based approach. More knowledge is needed to provide insight in the magnitude and impact of the disparity in definitions.

The objectives of the current study are twofold. First, it aims at systematically reviewing the different symptom-based definitions and methods used in the literature to assess exacerbation rates. The second objective is to test the hypothesis that exacerbation-related outcomes depend on algorithms applied, using data from an existing COPD cohort.

MATERIAL AND METHODS

Systematic review

In order to review the different symptom-based definitions and methods used in the current literature systematically, we performed a literature search for peer reviewed publications on the following databases: Cochrane Controlled Trials Register, Pubmed, CINAHL and Web of Science (January 1995 – November 2010). Details on the search strategy and data extraction are described in Appendix I.

The following issues were extracted from the included studies: operational definitions of exacerbation onset, duration, recurrence and severity, methods of symptom registration, matching with event-based episodes and method of analysis of the results.

Impact of using different symptom algorithms

To quantify the impact of using different symptom-based definitions on exacerbation-related outcomes, we used data (n = 137, age: 65±10 years, male: 58% male, FEV_1: 59±21% pred) from the ACZiE study; an ongoing multicentre randomized clinical trial. Primary aim of this trial is to evaluate the effectiveness of an individualised 'action plan' as an addition to care as usual. In- and exclusion criteria of patients in the study are described in detail elsewhere[17].

All patients were instructed to record daily diary cards if symptoms were increased over their baseline condition for a period of 6 months. Patients could choose between 'no increase', 'slightly increase' or 'clear increase'. According to the identified algorithms we decided to take into account only symptoms that were reported as 'clearly increased' by the patient. Validity and compliance of symptom registrations was checked and reinforced by telephone contact every four weeks. In case of missing data in the diary, the patient is asked to recall this information. We used forward and backward imputation to replace missing data that could not be recalled[18].

Two researchers independently determined exacerbation rates according to the four most referenced algorithms in the current literature independently. In case of disagreement, consensus was achieved in a meeting, under supervision of a third reviewer To enable comparisons between each algorithm, exacerbation recovery time, severity and type of exacerbation were identified using the same operational criteria.

Exacerbation incidence rate

For each definition the exacerbation incidence rate was calculated, and reported as a weighted exacerbation incidence rate (total number of exacerbations divided by the total follow up time[19]. Exacerbation onset was taken as the first day on which the criteria for the symptom algorithms were met.

Clustering of exacerbations

Three types of episodes were distinguished. Initial exacerbations were patient's first exacerbation assessed after baseline and exacerbations not followed by another exacerbation within 8 weeks. A relapsed exacerbation was defined as an exacerbation that follows within 5 days of onset of a previous exacerbation and is considered to be a part of the same episode. A recurrent exacerbation is an exacerbation that has an onset within 8 weeks of the preceding exacerbation[20].

Classification of exacerbations

Classification of exacerbations was assessed according to Anthonisen et al.[21]. A type-III exacerbation is defined by the presence of one major symptom, type-II by the presence of two major symptoms and type-III as the presence of three majors present on the worst day of an exacerbation. Furthermore, we evaluated exacerbations as combination of height and duration of increased symptoms. The total symptom score, is defined as the sum of the daily symptom

scores (a major symptom accounts for 2 points and 1 point for minor symptoms and ‘slightly increased’ major symptoms).

Recovery time

Exacerbation recovery time was calculated as the time from exacerbation onset for the 3-day moving average of the daily symptom count to return to baseline symptom count (the mean daily symptom count over days 14 tot 8 preceding exacerbation onset). Although counted when calculating exacerbation rates, episodes with a recovery time longer than 35 days are considered as unrecovered and were excluded for specifying exacerbation recovery time.

Concurrence with event-based episodes

Healthcare utilization data were identified by monthly telephone contacts with the patient and evaluation of patients’ medical records (both at the hospital and general practice) after follow-up. An event was considered reported if a patient reported respiratory symptom increase to a healthcare provider in an unscheduled telephone contact, physician or emergency room visit. Treated events were defined as use of oral corticosteroids and/or antibiotics and/or hospitalisation for a worsening in the patient’s respiratory symptoms at the discretion of their usual physician. Concurrence between these events and symptom-based exacerbations algorithms was present if events occurred between 5 days prior and 30 days after the algorithm onset.

RESULTS

Literature review

Study selection

Our initial search retrieved a total of 468 citations, of which 341 abstracts were excluded, see Figure 1. After reviewing the remaining 127 full-text articles, another 84 articles were excluded. Finally, after cross-reference checking, 51 articles met our criteria and were included for analysis in this study. Remarkably, 24 (47%) studies were performed by the same research group, with 20 (39%) studies based on the same longitudinal 'East-London Cohort'. Within these studies, these investigators used consistent operational definitions of evaluating exacerbations and apparently have substantial impact in the findings of this review. A detailed description of their methodology can be found in Appendix II.

It needs to be emphasized that two studies addressing the highly discussed EXACT tool for assessing exacerbation characteristics were not included in the review, since they did not met our criteria for assessing exacerbation frequency[14,15]. These studies described the development and subsequent validation phase but did not yet provide and tested a definition/algorithm for defining exacerbation onset and computing event frequency in a prospective cohort. A brief description: The EXACT tool is a PRO-based 14-item e-diary assessing breathlessness, cough and sputum, chest symptoms, difficulty bringing up sputum, feeling tired or weak, sleep disturbance and feeling scared or worried about their condition, Each item is assessed on a 5 or 6 ordinal scale and summed to yield a total score converted to a 0 – 100 scale.

Characteristics of the included studies

Table 1 summarises characteristics of the included studies of which the majority used a longitudinal cohort design (71%). Twenty-nine percent of the studies evaluated effectiveness of

a pharmacological compound. The median [IQR: Inter Quartile Range] number of patients included was 109 [78 - 259]. Follow-up varied between 3 and 90 months (median 12).

Table 1. Characteristics of the 51 included studies.

Characteristics	**Studies** **n (% from total)**
RCT	15 (29%)
Cohort-study	36 (71%)
Evaluation of a pharmacological compound	15 (29%)
Median number of patients, n [IQR]	109 [78 - 259]
Median follow-up period in months (min - max)	12 (3 - 90)

Symptom-based exacerbation definitions

Fourteen different symptom-based algorithms to determine exacerbation onset were identified. A detailed description of the included studies is available in Appendix II. Within these definitions, 12 different symptoms were used to define exacerbation onset. Coryzal symptoms were scored when studies used a definition including 'cold' (n=5), upper respiratory infection (n=2) or specific symptoms like nasal discharge, nasal congestion or sneezing (n=31).

Nearly all studies referred to the three key [21]: increase of dyspnea (98%) sputum volume (94%) and sputum purulence (94%). Also the 'Anthonisen minor symptoms'[21] were frequently used; cough (86%), wheezing (76%), sore throat (73%) and coryzal symptoms (75%). Only a minority of the studies included fever (16%), chest tightness (8%), fatigue (4%), difficulty with expectoration (4%) and night-time awakenings (2%). Table 2 shows an overview of exacerbation algorithms observed, showing large variations in symptom criteria. The majority (82%) of algorithms, distinguished between major and minor symptoms. Obviously the modified

algorithm by Anthonisen et al. is the most frequently used algorithm (71%), requiring increase in two symptoms including at least one major. Four studies used the same algorithm but also included fever (8%) as a minor symptom. Another four studies (8%) only required one major symptom to change over two consecutive days. Three studies (6%) defined exacerbation onset based on a graded symptom score. One of the studies did not specify the symptoms defining the onset of an exacerbation but only mentioned “increase of symptoms”. Two consecutive days was the most frequently (85%) used minimal time frame in which the symptom criteria should be met, followed by three days (6%) and 1 day (2%). Four (8%) studies did not specify a minimal time frame.

Table 3 indicates that a substantial number of studies did not report on criteria for exacerbation recovery or rules for defining subsequent episodes. Of the 36 studies which stated criteria for recovery by symptom scores returning to a predefined baseline, 28 studies used the recovery rule, introduced by the East-London group as a 3-day moving average to return to the mean symptom count of day -14 to -8. A minority defined recovery as the first day the exacerbation onset criteria did not met. To determine independence of events, 11 studies (22%) reported a minimal stable time period to distinguish exacerbation relapse from recurrence, which varied between 2 and 50 days (median 3 IQR 3-14 days). Concurrence with event-based exacerbations was reported in 39 studies (76%), while three studies (6%) incorporated blinded adjudication by two or more blinded investigators to ensure that events counted as exacerbations were consistent with the study definition of exacerbation.

Eighteen studies (35%) did not attempt to classify exacerbations in terms of severity. Other studies used widely varying approaches; by symptom count, by healthcare utilization (i.e. mild: increase inhalation medication; moderate: course of antibiotics or corticosteroids; severe: hospital admission), by the number of major symptoms (mild=1, moderate=2, severe=3) or by exacerbation length.

Different methods of registration were used to evaluate daily symptom change of which the majority (76%) used a ‘written’ daily diary card to record symptom increase. Seven studies identified predefined episodes of symptom increase by recall, either by telephone or by clinic visits. Three studies (6%) did not report on their methods to record symptom change. The majority (69%) of the studies reported on methods to enhance validity and compliance of diary registration, using run-in periods, standardized telephone or clinic-visit checks. Frequency of these checks varied between weekly and every four months. Of the studies using diary cards, only three (8%) explicitly reported on methods handling missing diary card data (n=2: multiple imputation, n=1 retrospective interviews).

Table 2: Overview of exacerbation onset algorithms in the included studies (n=51)

Algorithm	Studies N =	Criteria for defining onset	Consecutive days	DY	Sputum volume	Sputum purulence	Cough	Wheezing	Sore throat	Coryzal symptoms	Fever	Chest tightness	FA	DE	NA
1. Modified Anthonisen	32	≥ increase in two symptoms including one major	≥ 2 days‡	Ma	Ma	Ma	Mi	Mi	Mi	Mi	-	-	-	-	-
2. Modified Anthonisen+ fever	4	≥ increase in two symptoms including one major	≥ 2 days	Ma	Ma	Ma	Mi	Mi	Mi	Mi	Mi	-	-	-	-
3. Only 'Major' symptoms	4	≥ increase in 1 symptom	≥ 2 days#	+	+	+	-	-	-	-	-	-	-	-	-
4. Vijayasaratha et al.	1	Graded symptom scale; ≥ increase in 1 point in 1 major symptom	≥ 2 days	Ma	Ma	Ma	Mi	-	-	Mi	Mi	Mi	-	-	-
5. Alvarez-Mon et al.	1	≥ increase in 2 symptoms	NS	+	+	+	+	-	-	-	-	-	-	-	-
6. Ekberg et al.	1	≥ 3 points; major symptom=2 points, minor=1 point	≥ 2 days	Mi	Mi	Mi	Ma	-	-	-	-	-	-	Mi	-
7. van Schayk et al.	1	≥ increase in two symptoms including one major	NS	Mi	Mi	Ma	Ma	Mi	Mi	Mi	Mi	Mi	Mi	Mi	-
8. Sund et al.	1	Graded symptom score; ≥ increase in 1 point for 2 symptoms	≥ 2 days	+	+	+	-	-	-	-		+	+	-	-
9. Banerjee et al.	1	'Increase in symptoms'	NS	-	-	-	-	-	-	-	-	-	-	-	-
10. Calverley et al.	1	Graded symptom score, five different criteria described*	≥ 3 days	+	-	-	+	-	-	-	+	-	-	-	+
11. Casaburi et al.	1	≥ increase in 1 symptom	≥ 3 days	+	-	-	+	+	-	-	-	+	-	-	-
12. Dowson et al.	1	≥ increase in 2 symptoms	≥ 1 day	+	+	+	-	-	-	-	-	-	-	-	-
13. Pela et al.	1	Graded symptom score; ≥ increase in 3 points	NS	+	+	+	+	-	-	-	+	-	-	-	-
14. Dahl et al.	1	≥ increase in 1 symptom	≥ 3 days	+	+	+	+	+	-	-	-	-	-	-	-

DY: dyspnea; SP: sputum purulence; SV: sputum volume; WH: wheezing; CO: cough; ST: sore throat; CS: Coryzal Symptoms; FE: fever; CT: chest tightness; FA: fatigue; DE: Difficulty with expectoration; NA: Night-time awakenings; Ma: Major symptom; Mi: Minor symptom; +: included in the algorithm but no distinction between major and minor symptoms; -:not included in the algorithm; NS: Not specified; ‡ Consecutive days, not specified in 3 studies; # Consecutive days, not specified in 2 studies; *See online supplement.

Table 3: Recovery time, severity classification, data registration, attempts to enhance validity of diary registration, blinded adjudication of exacerbations and handling of missing diary data (n=51).

Recovery rules* **n (% from total)**		**Severity classification** **n (% from total)**	
Not reported	15 (29%)	Not reported	18 (35%)
Symptom score back to an individual baseline	28 (55%)	By symptom count at onset‡	24 (47%)
Symptom score back to a fixed value	1 (2%)	By number of major symptoms	4 (8%)
First day not meeting onset criteria	6 (12%)	By duration‡	5 (10%)
All symptoms resolved	1 (2%)	By healthcare utilization	3 (6%)
Data registration **n (% from total)**		**Attempts to enhance validity and compliance of diary registration, n (% from 39 studies using daily diary cards)**	
Not reported	3 (6%)	Not reported	12 (31%)
Daily written diary cards	39 (76%)	Run-in period†	2 (5%)
Daily electronic diary cards	1 (2%)	Telephone check†	4 (10%)
Telephone consultations#	4 (8%)	Clinic visit check	24 (62%)
Clinic visits	4 (8%)	Random home visits	1 (3%)
Blinded adjudication of exacerbations **n (% from total)**		**Handling of missing diary data (checken voor studies die ook daadwerkelijk kaarten gebruiken)**	
Not reported	48 (94%)	Not reported	48 (94%)
Yes	3 (6%)	Multiple imputation	2 (4%)
		Retrospective interviews	1 (2%)
Concurrence with event-based exacerbations **n (% from total)**			
Not reported	12 (24%)		
Yes	39 (76%)		

** Rules to define recovery and/or and to identify new or recurrent events; ‡Three studies used a severity classification by symptom count and duration. # performed monthly in all studies; † two studies combined a run-in period with telephone checks.*

The impact of using different symptom algorithms

In our study, chest tightness, fatigue, difficulty with expectoration and night-time awakenings were not assessed resulting in four different algorithms that could be tested, covering 42 of the 51 studies (83%): 1) modified Anthonisen (n=32); 2) modified Anthonisen including fever (n=4); 3) at least one 'major' (dyspnea, sputum amount and purulence) symptom (n=4); 4) at least two 'major' symptoms (n=2).

Table 4 illustrates the effects on exacerbation related outcomes when applying different symptom-based definitions. Algorithm 1 and 2 generated equal number of 119 (1.7 CI 1.4-2.1 per person-year) exacerbations and subsequent characteristics, indicating that adding fever as a minor symptom is not decisive in capturing additional events. Algorithm 3, requiring only 1 major symptom identified the highest number of 132 (1.9 CI 1.6-2.3 per person-year) exacerbations, counting 1.11 and 2.44 times algorithm 1 and 4, respectively. In addition, this algorithm also provides a higher crude number of relapsed and recurrent exacerbations but similar within group distributions in clustering type. In terms of classification, lowering the threshold (compared to modified Anthonisen) results in a shift towards increased identification of Anthonisen type-III exacerbations and subsequently a lower median symptom count of 57 [IQR 27-94] and 51 [IQR 23-91] respectively. Although requiring 2 major symptoms (algorithm 4) resulted in a 120% (n=54) lower incidence (1.5 CI 0.6-1.0 per person-year), compared to algorithm 1. At the same time, a lower ratio of patients with >1 exacerbation was seen, both within the total group (8%; algorithm 1: 23%; algorithm 3: 25%) as the group of patients having at least one exacerbation (37%; algorithm 1: 55%; algorithm 3: 53%). Algorithm 4 excludes type-III exacerbation, since it requires the presence of two major symptoms. The number of type-II exacerbations is also lower in algorithm 4 due to the fact that the increase in two majors had to be present for 2 consecutive days. This results in a lower ratio of type-I to type-II exacerbations compared to the other algorithms. Subsequently this approach produced the highest mean symptom count of 70 [IQR 42-125] including a higher proportion of exacerbations to be reported and subsequently treated. Nevertheless, 11 treated events and 4 hospital admission/emergency room visits identified by algorithm 1 to 3 would have been missed by algorithm 4.

Table 4. Exacerbation rate, type, severity and concurrence with reported / treated events of events using four different algorithms.

	Symptom algorithm	No. *	Incidence rate (per person year)	No. per patient (% of total) (% of patients with an exacerbation)			Clustering type (% of total)			Recovery time median [IQR]	Anthonisen classificaton by no. of majors (% of total)			Total symptom count median [IQR]	Concurrence with Reported and treated events		
				0	1	>1	Initial	Relapsed	Recurrent		I	II	III		Reported events	Treated events	Hospital admission / emergency room visits missed
1	Modified Anthonisen	119	1.7 (1.4 -2.1)	81 (59) -	25 (18) (45)	31 (23) (55)	69 (58)	13 (11)	50 (42)	14 [7-22]	27 (23)	38 (32)	54 (45)	57 [27-94]	62 (52%)	31 (26%)	1
2	Modified Anthonisen + fever	119	1.7 (1.4 -2.1)	81 (59) -	25 (18) (45)	31 (23) (55)	69 (58)	13 (11)	50 (42)	14 [7-22]	27 (23)	38 (32)	54 (45)	57 [27-94]	62 (52%)	31 (26%)	1
3	At least one ‘major’ symptom	132	1.9 (1.6-2.3)	73 (53) -	30 (22) (47)	34 (25) (53)	75 (57)	14 (11)	57 (43)	13 [7-22]	27 (20)	39 (30)	66 (50)	51 [23-91]	62 (47%)	31 (24%)	1
4	At least two ‘major’ symptoms	54	1.5 (0.6-1.0)	105 (77) -	20 (15) (63)	12 (8) (37)	35 (65)	2 (4)	19 (35)	14 [8-21]	27 (50)	27 (50)	-	70 [42-125]	41 (76%)	22 (41%)	5

* Relapsed exacerbations excluded (original exacerbation and relapse are considered as the same event)

DISCUSSION

Our systematic review revealed significant inconsistencies in definitions, methods and accuracy to define symptom-based COPD exacerbations. Differences in the most referenced definitions were tested in an existing COPD cohort to quantify impact on exacerbation related outcome. We demonstrated that minor changes in symptom criteria substantially affect incidence rates, clustering type and classification of exacerbations.

This review demonstrates that symptom-based exacerbations have been frequently used in trials but mainly in longitudinal studies. The 51 studies meeting the inclusion criteria, showed large variations in defining onset of exacerbations. Fourteen different symptoms algorithms were found, aimed at the same objective; defining exacerbation onset or rates. The most prominent applied algorithm is based on a modification by Anthonisen et al. requiring increase in at least dyspnea, sputum volume and sputum purulence[21]. Besides these three generally agreed cardinal symptoms, nine other symptoms were used, reflecting the inconsistent attempts to cover the heterogeneity in aetiology of exacerbations[8]. The 2000 Aspen consensus statement has partly eliminated this complexity by not specifying symptoms but requiring a straightforward 'worsening of the patient condition'[2]. Although this aggregation secures covering all exacerbations, it is lacking the operational discriminative properties needed in prospective studies.

The algorithms differentiated from the literature used widely varying definitions of COPD exacerbations including type and number of symptoms and days these criteria should be met. Using data from our COPD cohort, we demonstrated that minor changes in the four most referenced algorithms have substantial impact on incidence rates of exacerbations. Adding fever as a minor symptom to the most frequently used Anthonisen modification did not show any added value in capturing additional exacerbations. Lowering the threshold by not including minor symptoms however produced 11% increase in the number of episodes. Otherwise, increasing the threshold lowered exacerbation incidence substantially and increased the change

of missing important treated-events including hospital admissions. These apparently, subtle adaptations in the threshold, also affected the distribution and group characteristics in terms of exacerbation type and classification. It needs to be emphasized that the straightforward modified Anthonisen definition in our test cohort seemed to steer a middle course between other thresholds, and therefore could be considered the best available trade-off for optimal classification of events. Other studies showed that it is quite unusual to have increase in only one major symptom without at least one minor symptom[10, 22]. Therefore, including a minor symptom might decrease the risk of overestimation. In addition, in line with the lower incidence found for algorithm 4 in our test cohort, the risk of underestimation might increase if a threshold of at least 2 major symptoms is applied. Thanks to the cumulative experience of the East-London group, the (modified) Anthonisen definition has proven to be operational and validated against important outcomes including, airway inflammatory markers[23] quality of life[3] and lung function decline[4]. Nevertheless, we need to be reserved considering it as the gold standard, because like others, it also fails to cover all of the heterogeneity of COPD and its exacerbations.

Our review revealed other methodological issues, which can substantially alter exacerbation outcomes. A critical aspect in counting exacerbations episodes is how to handle with multiple events within individual patients and how to distinguish exacerbation relapse from recurrence. Almost 30% of the studies did not report criteria when an index episode recovered and subsequently a new event was counted implicating that reported rates can be biased. Exacerbations have shown not to be random events, but seem to cluster in time[20], emphasizing the importance of defining recovery rules and criteria for subsequent events. The studies that did define recovery rules, used different methods. The most referenced rules were meeting the exacerbation onset criteria and recovery of a symptom count to a predefined individual baseline. Defining individual normal day-to-day variations requires assessment over a sufficient number of days not including prodromal increase of symptoms. A frequently used

method to assess baseline stability that includes these aspects, is taking an average symptom count (each symptom reflects 1 point) of day 14 to 8 preceding an exacerbation[24] . Only 29 % of the studies included a minimal number of stable days for a new event could be identified or used a moving mean symptom count (mostly 3 days) to meet the pre-exacerbation baseline. Both methods deal with the essential requirement of assuring that exacerbation relapse is not counted as a new subsequent episode.

Assessing and analyzing exacerbations based on symptom criteria is highly complex and needs to be performed as accurate as possible. Surprisingly, 8 studies (16%) used monthly evaluations based on recall (telephone review or clinic visits) to assess episodes retrospectively. It's highly debatable whether testing symptom-based algorithms (including symptom criteria to be present for at least x number of days) by monthly recall results in a valid outcome. This loss of accuracy not only applies for defining exact onset of exacerbation but also leads to impossible identification of the aforementioned aspects for recovery and recurrence.

The majority of the studies used daily diary recording to examine symptom increase. To achieve optimal validity and avoid missing data, appropriate efforts should be taken in enhancing patients understanding and compliance. Although the majority of the studies reviewed patients regularly by telephone or clinic visit checks, they did not report on how validity and compliance was improved. Symptom-recording demands a certain degree of cognitive skills and therefore missing data can rarely be avoided in which subjects are requested to complete questionnaires with many items[25]. Although only few studies reported on completion-rates, diaries in the East-London cohort were completed reasonably well (~ 85 % of the time in the study)[4, 10]. Surprisingly, only three studies reported on how missing or invalid data were managed. Analysis of available data only, will lead to loss of efficiency and since missing data in daily symptom recording is expected to be not a random phenomenon, to biased results. Two studies used multiple imputation of missing data using auxiliary data. Although statistical methods like joint modelling or multiple imputation have shown to be potential when missing at random occurs in

longitudinal studies[26], this has never been simulated in COPD patients. Next to dealing with missing data it's highly desirable that future studies put more efforts in pro-actively avoiding missing and invalid data. Ideally, follow-up is preceded by a run-in period with appropriate length in which patients receive feedback on the importance of complete diary recording. It also enhances patients to be well instructed and if necessary corrected on how to record symptom increase. Another method, assisting patients in correctly identifying these episodes to be beyond normal day to day variability is to provide an individualized and dynamic "what is normal" card. Such cards can easily be adapted if a patient does not return to it's stable condition.

Although there is no general agreement definition of an exacerbation, exacerbation rate is a clinically important outcome in COPD research. Since event-driven approaches almost certainly fail to capture all events [9, 12, 25] and adequate biomarkers are not available until now, we suggest that studies continue to use symptom-based definitions but should pay more attention into their methods of data recording since this will contribute to the accuracy of data retrieval and assessment of exacerbations. This review did not aim at identifying 'the best' definition but rather at quantifying the variety and consequences of using different symptom-based approaches to count and analyze exacerbations. The widely varying symptom criteria for analyzing exacerbation episodes complicate comparisons between certain studies or populations. This study illustrated that in- or exclusion of a single symptom substantially affects exacerbation related outcomes and therefore choosing a certain definition truly matters. These findings emphasize that inconsistent use of definitions leads to widely varying exacerbation related outcome. Besides these differences in crude rates, this also significantly affects effect sizes of interventions evaluated in randomized trials, as illustrated in another Dutch COPD cohort [13]. Ideally, an instrument assessing occurrence and severity of exacerbations with appropriate measurement properties, sensitivity and an origin in the patients' perception of exacerbations would be a true asset. Possibly, a promising new initiative, EXACT-PRO is able

to accomplish these requirements to a substantial extend[14, 15]. In a recent study comparing stable patients and patients visiting the clinic for an exacerbation, EXACT was found to be a valid tool to in discriminating between stable periods and exacerbations. Furthermore it was internal consistent and sensitive to change with recovery of an exacerbation[15]. Although this PRO-based tool could be considered the best available method to measure the magnitude and duration of exacerbations, no data is available on its discriminative performance in prospectively identifying exacerbation onset. The pivotal objective of diary recording is to assess symptom-increase beyond normal day-to-day variability. The degree of this change represents exacerbation severity. The majority of studies required the patient to note a worsening of the symptom for a certain number of consecutive days, but did not specify by how much symptoms should deteriorate. For this reason, it is questionable if simply counting symptoms, as performed by 28 studies provides a valid assessment of severity. The Anthonisen classification and total symptom count address certain exacerbation characteristics, but have never been validated in terms of severity. In the same way as EXACT, four studies used a graded diary card suggesting specific thresholds to be exceeded. Although this provides a much better way to quantify the degree of change (and severity), the limits of variations of individual daily symptom scoring should be well established to separate changes due to spontaneous variation (“bad days”) from true exacerbations. Future longitudinal studies are needed to evaluate the discriminative performance of the proposed EXACT scoring algorithms and its surplus value against existing definitions like the modified Anthonisen.

This review revealed large inconsistencies in definitions, methods and accuracy to define symptom-based COPD exacerbations. In an existing COPD cohort, we demonstrated that minor changes in symptom criteria substantially affect incidence rates, clustering type and classification of exacerbations. These results stress the importance of leading organizations and investigators to increase their efforts in reaching consensus on operational symptom-based definitions and quality requirements of counting and analyzing exacerbations.

ACKNOWLEDGMENTS

We gratefully acknowledge all the respiratory nurses, general and family physicians, and research assistants involved in this study but especially Suzanne van Vooren, Joyce Heijneman, Diederik Bolkestein, Justine Sinsai, Bregje Peeters, Rosalie Foppen, Mariska Manten, Anne-Marie Bulten en Carien Roos.

COMPETING INTERESTS

None

FUNDING

This study was not funded by any grant.

REFERENCES

1 Rabe KF, Hurd S, Anzueto A, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176: 532-555.

2 Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. *Chest* 2000; 117: 398S-401S.

3 Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998; 157: 1418-1422.

4 Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 2002; 57: 847-852.

5 Soler-Cataluna JJ, Martinez-Garcia MA, Roman SP, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax* 2005; 60: 925-931.

6 Miravitlles M, Murio C, Guerrero T, Gisbert R. Pharmacoeconomic evaluation of acute exacerbations of chronic bronchitis and COPD. *Chest* 2002; 121: 1449-1455.

7 Aaron SD, Fergusson D, Marks GB, *et al*. Counting, analysing and reporting exacerbations of COPD in randomised controlled trials. *Thorax* 2008; 63: 122-128.

8 Sapey E, Stockley RA. COPD exacerbations . 2: aetiology. *Thorax* 2006; 61: 250-258.

9 Langsetmo L, Platt RW, Ernst P, Bourbeau J. Underreporting exacerbation of chronic obstructive pulmonary disease in a longitudinal cohort. *Am J Respir Crit Care Med* 2008; 177: 396-401.

10 Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2000; 161: 1608-1613.

11 Vijayasaratha K, Stockley RA. Reported and unreported exacerbations of COPD: analysis by diary cards. *Chest* 2008; 133: 34-41.

12 Xu W, Collet JP, Shapiro S, *et al*. Negative impacts of unreported COPD exacerbations on health-related quality of life at 1 year. *Eur Respir J* 2010; 35: 1022-1030.

13 Effing TW, Kerstjens HA, Monninkhof EM, *et al*. Definitions of exacerbations: does it really matter in clinical trials on COPD? *Chest* 2009; 136: 918-923.

14 Leidy NK, Wilcox TK, Jones PW, *et al*. Development of the EXAcerbations of Chronic Obstructive Pulmonary Disease Tool (EXACT): A Patient-Reported Outcome (PRO) Measure. *Value Health* 2010.

15 Leidy NK, Wilcox TK, Jones PW, Roberts L, Powers JH, Sethi S. Standardizing Measurement of COPD Exacerbations: Reliability and Validity of a Patient-Reported Diary. *Am J Respir Crit Care Med* 2010.

16 U.S.Department of Health and Human Services FDA Center for Drug evaluation and Research, U.S.Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S.Department of Health and Human Services FDA Center for Devices and Radiological Health. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims. 2009.
Ref Type: Report

17 Trappenburg JC, Koevoets L, de Weert-van Oene GH, *et al*. Action Plan to enhance self-management and early detection of exacerbations in COPD patients; a multicenter RCT. *BMC Pulm Med* 2009; 9: 52.

18 Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol* 2003; 56: 968-976.

19 Suissa S. Statistical treatment of exacerbations in therapeutic trials of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; 173: 842-846.

20 Hurst JR, Donaldson GC, Quint JK, Goldring JJ, Baghai-Ravary R, Wedzicha JA. Temporal clustering of exacerbations in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2009; 179: 369-374.

21 Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med* 1987; 106: 196-204.

22 Calverley P, Pauwels DR, Lofdahl CG, *et al*. Relationship between respiratory symptoms and medical treatment in exacerbations of COPD. *Eur Respir J* 2005; 26: 406-413.

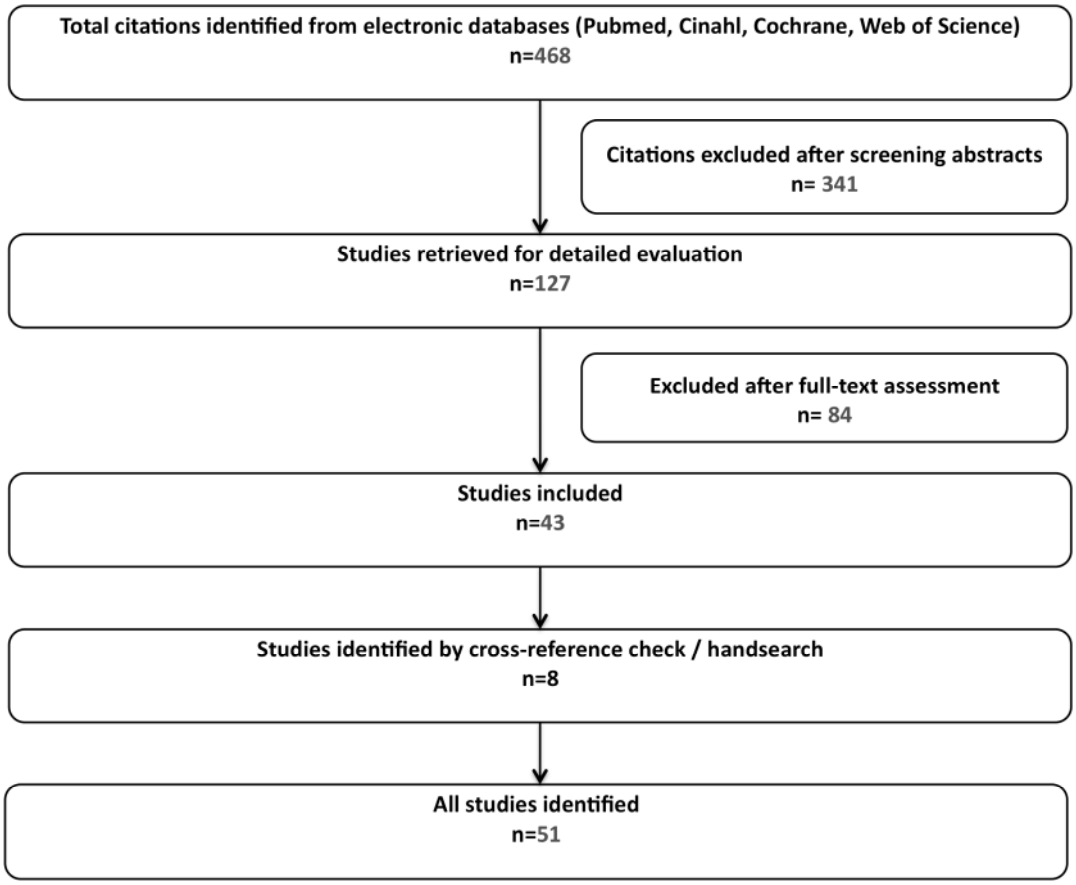
23 Bhowmik A, Seemungal TA, Sapsford RJ, Wedzicha JA. Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. *Thorax* 2000; 55: 114-120.

24 Wilkinson TM, Donaldson GC, Hurst JR, Seemungal TA, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004; 169: 1298-1303.

25 Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. *Eur Respir J Suppl* 2003; 41: 46s-53s.

26 Wang C, Hall CB. Correction of bias from non-random missing longitudinal data using auxiliary information. *Stat Med* 2010; 29: 671-679.

figure 1


Total citations identified from electronic databases (Pubmed, Cinahl, Cochrane, Web of Science)
n=468
Citations excluded after screening abstracts
n= 341
Studies retrieved for detailed evaluation
n=127
Excluded after full-text assessment
n= 84
Studies included
n=43
Studies identified by cross-reference check / handsearch
n=8
All studies identified
n=51