REVIEW

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

J.C.A. Trappenburg*,+, A.C. van Deventer*,+, T. Troosters#,+, T.J.M. Verheij*,+, A.J.P. Schrijvers*,+, J-W.J. Lammers¶,+ and E.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. However, no consensus has been 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 tests 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 chronic obstructive pulmonary disease (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 episodes·patient-yr^{-1} (95% CI 1.4–2.1), while for requiring only one major or two major symptoms this was 1.9 episodes·patient-yr^{-1} (95% CI 1.6–2.3) and 1.5 episodes·patient-yr^{-1} (95% CI 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.

KEYWORDS: Chronic obstructive pulmonary disease, chronic obstructive pulmonary disease exacerbations

Chronic obstructive pulmonary disease (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) characterised 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], and are associated with accelerated lung function decline [4] and increased mortality [5]. In addition, they represent a significant economic burden due to healthcare utilisation [6]. The relevance of exacerbations from a clinical, patient and societal perspective has resulted in the selection of 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 healthcare services (increased use of reliever

medication, or treatment with systemic corticosteroids anti-biotics or hospitalisation) in the presence of a worsening condition of the patient [2]. Although simple counting of events by healthcare utilisation intuitively seems a straightforward and robust approach, it strongly depends on the ability of patients to recognise exacerbations and available healthcare facilities. Therefore, event-based definitions underestimate true exacerbation rates by ⩽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'' (PROs) [16]. Secondly, 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 two-fold. First, it aims to systematically review 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 in the online supplementary material.

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
In order to quantify the impact of using different symptom-based definitions on exacerbation-related outcomes, we used data (n=137; mean±SD age 65±10 yrs, 58% males; forced expiratory volume in 1 s 59±21% predicted) from the ACZiE (Action Plan in Patients with COPD to Enhance Self-management and Early Detection of Exacerbations Study), an ongoing multicentre randomised clinical trial. The primary aim of this trial is to evaluate the effectiveness of an individualised ''action plan'' in addition to care as usual. Inclusion and exclusion criteria of patients in the study are described in detail elsewhere [17].

All patients were instructed to record on daily diary cards if symptoms were increased over their baseline condition for a period of 6 months. Patients could choose between ''no increase'', ''slight 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 4 weeks. In case of missing diary data, 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. In the case of disagreement, consensus was achieved in a meeting, under the supervision of a third reviewer. In order to enable comparisons between each algorithm, exacerbation recovery time, and 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 the 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 I 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 major symptoms 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 to 8 preceding exacerbation onset). Although counted when calculating exacerbation rates, episodes with a recovery time >35 days are considered unrecovered and were excluded from specification of exacerbation recovery time.

Concurrence with event-based episodes

Healthcare utilisation data were identified by monthly telephone contacts with the patient and evaluation of the patient's 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 health-care provider in an unscheduled telephone contact, or physician or emergency room visit. Treated events were defined as use of oral corticosteroids, 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 exacerbation algorithms was present if events occurred between 5 days before 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 (fig. 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
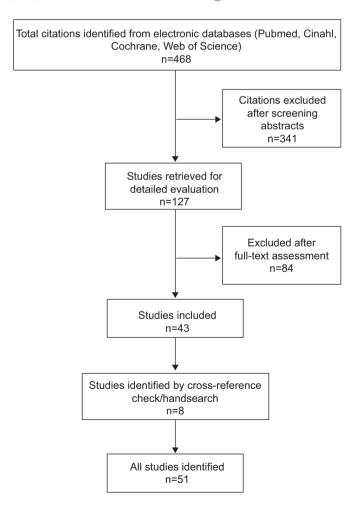
Cohort". Within these studies, these investigators used consistent operational definitions for evaluating exacerbations and apparently have substantial impact in the findings of this review. A detailed description of their methodology can be found in Appendix II in the online supplementary material.

It needs to be emphasised that two studies addressing the highly discussed EXACT tool for assessing exacerbation characteristics were not included in the review, as they did not meet our criteria for assessing exacerbation frequency [14, 15]. These studies described the development and subsequent validation phase, but did not yet provide and test 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 electronic 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 five or six 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%). 29% of the studies evaluated effectiveness of a pharmacological compound. The median number of patients included was 109 (interquartile range (IQR) 78–259). Follow-up varied between 3 and 90 months (median 12 months).

Symptom-based exacerbation definitions

14 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 such as nasal discharge, nasal congestion or sneezing (n=31).

Nearly all studies referred to the three key symptoms [21]: increase of dyspnoea (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

FIGURE 1. Flow chart of included and excluded studies.

TABLE 1	Characteristics of the 51 included studies

Characteristics	Studies
RCT	15 (29)
Cohort study	36 (71)
Evaluation of a pharmacological compound	15 (29)
Patients median (IQR)	109 (78–259)
Follow-up period months median (range)	12 (3–90)

Data are presented as n (%), unless otherwise stated. RCT: randomised controlled trial; IQR: interquartile range.

TABLE 2	Overview of exacerbation onset algorithms in the included studies[#]														
Algorithm	Studies n	Criteria for defining onset	Consecutive days	DY	SV	SP	CO	WH	ST	CS	FE	CT	FA	DE	NA
Modified Anthonisen	32	Increase in ≥2 symptoms including 1 major	≥2[+]	Ma	Ma	Ma	Mi	Mi	Mi	Mi	-	-	-	-	-
Modified Anthonisen + fever	4	Increase in ≥2 symptoms including 1 major	≥2	Ma	Ma	Ma	Mi	Mi	Mi	Mi	Mi	-	-	-	-
Only major symptoms	4	Increase in ≥1 symptom	≥2[§]	+	+	+	-	-	-	-	-	-	-	-	-
Vijayasaratha [11]	1	Graded symptom scale; increase in ≥1 point in 1 major symptom	≥2	Ma	Ma	Ma	Mi	-	-	Mi	Mi	Mi	-	-	-
Alvarez-Mon [22]	1	Increase in ≥2 symptoms	NS	+	+	+	+	-	-	-	-	-	-	-	-
Ekberg-Jansson [23]	1	≥3 points; major symptom=2 points, minor=1 point	≥2	Mi	Mi	Mi	Ma	-	-	-	-	-	-	Mi	-
Van Schayk [24]	1	Increase in ≥2 symptoms including 1 major	NS	Mi	Mi	Ma	Ma	Mi	Mi	Mi	Mi	Mi	Mi	Mi	-
Sund [25]	1	Graded symptom score; increase of ≥1 point for 2 symptoms	≥2	+	+	+	-	-	-	-	-	+	+	-	-
Banerjee [26]	1	''Increase in symptoms''	NS	-	-	-	-	-	-	-	-	-	-	-	-
Calverley [27]	1	Graded symptom score, 5 different criteria described[¶]	≥3	+	-	-	+	-	-	-	+	-	-	-	+
Casaburi [28]	1	Increase in ≥1 symptom	≥3	+	-	-	+	+	-	-	-	+	-	-	-
Dowson [29]	1	Increase in ≥2 symptoms	≥1	+	+	+	-	-	-	-	-	-	-	-	-
Pela [30]	1	Graded symptom score; increase in ≥3 points	NS	+	+	+	+	-	-	-	+	-	-	-	-
Dahl [31]	1	Increase in ≥1 symptom	≥3	+	+	+	+	+	-	-	-	-	-	-	-

DY: dyspnoea; SV: sputum volume; SP: sputum purulence; CO: cough; WH: wheezing; ST: sore throat; CS: coryzal symptoms; FE: fever; CT: chest tightness; FA: fatigue; DE: difficulty with expectoration; NA: night-time awakenings; NS: not specified; Ma: major symptom; +: included in the algorithm but no distinction between major and minor symptoms; -: not included in the algorithm; Mi: minor symptom. [#]: n=51. [¶]: see online supplementary material. [+]: not specified in three studies. [§]: not specified in two studies.

algorithms observed, showing large variations in symptom criteria. The majority (82%) of algorithms distinguished between major and minor symptoms. Obviously, the modified algorithm of ANTHONISEN *et al.* [21] is the most frequently used algorithm (71%), requiring increase in two symptoms, including at least one major symptom. Four (8%) studies used the same algorithm but also included fever as a minor symptom. Another four (8%) studies only required one major symptom to change over 2 consecutive days. Three (6%) studies 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''. 2 consecutive days was the most frequently (85%) used minimal time frame in which the symptom criteria should be met, followed by 3 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 that 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– -8. A minority defined recovery as the first day the exacerbation onset criteria were not met. To determine independence of events, 11 (22%) studies reported a minimal stable time period

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[#]

Criteria	Studies n (%)
Recovery rules[¶]	
Not reported	15 (29)
Symptom score back to an individual baseline	28 (55)
Symptom score back to a fixed value	1 (2)
First day not meeting onset criteria	6 (12)
All symptoms resolved	1 (2)
Data registration	
Not reported	3 (6)
Daily written diary cards	39 (76)
Daily electronic diary cards	1 (2)
Telephone consultations[+]	4 (8)
Clinic visits	4 (8)
Blinded adjudication of exacerbations	
Not reported	48 (94)
Yes	3 (6)
Concurrence with event-based exacerbations	
Not reported	12 (24)
Yes	39 (76)
Severity classification	
Not reported	18 (35)
By symptom count at onset[§]	24 (47)
By number of major symptoms	4 (8)
By duration[§]	5 (10)
By healthcare utilisation	3 (6)
Attempts to enhance validity and compliance of diary registration[ƒ]	
Not reported	12 (31)
Run-in period[##]	2 (5)
Telephone check[##]	4 (10)
Clinic visit check	24 (62)
Random home visits	1 (3)
Handling of missing diary data	
Not reported	48 (94)
Multiple imputation	2 (4)
Retrospective interviews	1 (2)

[#]: n=51; [¶]: rules to define recovery and/or identify new or recurrent events; [+]: performed monthly in all studies; [§]: three studies used a severity classification by symptom count and duration; [ƒ]: from 39 studies using daily diary cards; [##]: two studies combined a run-in period with telephone checks.

to distinguish exacerbation relapse from recurrence, which varied between 2 and 50 days (median 3 days, IQR 3–14 days). Concurrence with event-based exacerbations was reported in 39 (76%) studies, while three (6%) studies incorporated blinded adjudication by two or more blinded investigators to ensure that events counted as exacerbations were consistent with the study definition of exacerbation.

18 (35%) studies did not attempt to classify exacerbations in terms of severity. Other studies used widely varying approaches: by symptom count, by healthcare utilisation (*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 and 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 (6%) studies 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, or standardised telephone or clinic-visit checks. Frequency of these checks varied between weekly and every 4 months. Of the studies using diary cards, only three (8%) explicitly reported on methods handling missing diary card data (two studies used multiple imputation and one study used retrospective interviews).

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 (83%) of the 51 studies: 1) modified Anthonisen (n=32); 2) modified Anthonisen including fever (n=4); 3) at least one major (dyspnoea, and sputum amount and purulence) symptom (n=4); and 4) at least two major symptoms (n=2).

Table 4 illustrates the effects on exacerbation-related outcomes when applying different symptom-based definitions. Algorithms 1 and 2 generated an equal number of 119 exacerbations and subsequent characteristics (incidence 1.7 exacerbations·person-yr^{-1}, 95% CI 1.4–2.1 exacerbations·person-yr^{-1}), indicating that adding fever as a minor symptom is not decisive in capturing additional events. Algorithm 3, requiring only one major symptom, identified the highest number of 132 exacerbations (1.9 exacerbations·person-yr^{-1}, 95% CI 1.6–2.3 exacerbations·person-yr^{-1}), which was 1.11 and 2.44 times the number of exacerbations according to algorithms 1 and 4, respectively. In addition, this algorithm also provides a higher crude number of relapsed and recurrent exacerbations, but similar within-group distributions of 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 two major symptoms (algorithm 4) resulted in a 120% (n=54) lower incidence (1.5 exacerbations·person-yr^{-1}, 95% CI 0.6–1.0 exacerbations·person-yr^{-1}), compared with algorithm 1, at the same time, a

Symptom algorithm	Exacerbations# n	Incidence exacerbations·person-yr⁻¹ (95% CI)	Exacerbations n (% total) (% patients with an exacerbation)			Recovery time days	Clustering type n (% total)			Anthonisen classificaton n (% total)			Total symptom count n	Concurrence with reported and treated events		
			0	1	>1		Initial	Relapsed	Recurrent	I	II	III		Reported events n (%)	Treated events n (%)	Missed admissions¶ n
1.	119	1.7 (1.4–2.1)	81 (59)	25 (18) (45)	31 (23) (55)	14 (7–22)	69 (58)	13 (11)	50 (42)	27 (23)	38 (32)	54 (45)	57 (27–94)	62 (52)	31 (26)	1
2.	119	1.7 (1.4–2.1)	81 (59)	25 (18) (45)	31 (23) (55)	14 (7–22)	69 (58)	13 (11)	50 (42)	27 (23)	38 (32)	54 (45)	57 (27–94)	62 (52)	31 (26)	1
3. ≥1	132	1.9 (1.6–2.3)	73 (53)	30 (22) (47)	34 (25) (53)	13 (7–22)	75 (57)	14 (11)	57 (43)	27 (20)	39 (30)	66 (50)	51 (23–91)	62 (47)	31 (24)	1
4. ≥2	54	1.5 (0.6–1.0)	105 (77)	20 (15) (63)	12 (8) (37)	14 (8–21)	35 (65)	2 (4)	19 (35)	27 (50)	27 (50)		70 (42–125)	41 (76)	22 (41)	5

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

Data are presented as median (interquartile range), unless otherwise stated. The symptom alogrithms are: 1. Modified Anthonisen; 2. Modified Anthonisen+fever; 3. ≥1 major symptom; 4. ≥2 major symptoms. #: relapsed exacerbations excluded (original exacerbation and relapse are considered as the same event); ¶: hospital and emergency room visits missed.

lower ratio of patients with more than one exacerbation was seen, both within the total group (8% *versus* 23% for algorithm 1 and 25% for algorithm 3) as the group of patients having at least one exacerbation (37% *versus* 55% for algorithm 1 and 53% for algorithm 3). 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 major symptoms had to be present for 2 consecutive days. This results in a lower ratio of type I to type II exacerbations compared with 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 four hospital admission/emergency room visits identified by algorithms 1–3 would have been missed by algorithm 4.

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 their 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. 14 different symptom algorithms were found, all with the same objective, *i.e.* defining exacerbation onset or rates. The most prominent applied algorithm is based on a modification of that of ANTHONISEN *et al.* [21], requiring increase in at least one symtpom out of dyspnoea, sputum volume and sputum purulence. 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 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 for which 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 an 11% increase in the number of episodes. Otherwise, increasing the threshold lowered exacerbation incidence substantially and increased the change in 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 emphasised 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, 27]. Therefore, including a minor symptom might decrease the risk of overestimation. In addition, consistent with the lower incidence found for algorithm 4 in our test cohort, the risk of underestimation might increase if a threshold of at least two 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 [32], quality of life [3] and lung function decline [4]. Nevertheless, we need to be reserved when 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 that can substantially alter exacerbation outcomes. A critical aspect in counting exacerbation episodes is how to handle multiple events within individual patients and 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], emphasising 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 [33]. Only 29% of the studies including a minimal number of stable days for a new event or using a moving mean symptom count (mostly 3 days) to meet the pre-exacerbation baseline could be identified. Both methods deal with the essential requirement of assuring that exacerbation relapse is not counted as a new subsequent episode.

Assessing and analysing exacerbations based on symptom criteria is highly complex and needs to be performed as accurately as possible. Surprisingly, eight (16%) studies used monthly evaluations based on recall (telephone review or clinic visits) to assess episodes retrospectively. It is highly debatable whether testing symptom-based algorithms (including criteria requiring symptoms 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 skill and, therefore, missing data can rarely be avoided when subjects are requested to complete questionnaires with many items [34]. Although only a 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 a loss of efficiency and, as missing data in daily symptom recording is expected not to be a random phenomenon, to biased results. Two studies used multiple imputation of missing data using auxiliary data. Although statistical methods, such as joint modelling or multiple imputation, have been shown to be of potential use when missing data occurs at random in longitudinal studies [35], this has never been simulated in COPD patients. As well as dealing with missing data, it is highly desirable that future studies put more effort into proactively avoiding missing and invalid data. Ideally, follow-up should be preceded by a run-in period of appropriate length in which patients receive feedback on the importance of complete diary recording. It also enhances compliance if patients are well instructed and, if necessary, corrected on how to record symptom increase. Another method for assisting patients in correctly identifying these episodes as beyond normal day-to-day variability is to provide an individualised and dynamic ''What is normal?'' card. Such cards can easily be adapted if a patient does not return to their stable condition.

Although there is no general agreement on the 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, 34] and adequate biomarkers are not available at present, we suggest that studies continue to use symptom-based definitions but should pay more attention into their methods of data recording, as this will contribute to the accuracy of data retrieval and assessment of exacerbations. This review did not aim to identify ''the best'' definition but, rather, to quantify the variety and consequences of using different symptom-based approaches to count and analyse exacerbations. The widely varying symptom criteria for analysing exacerbation episodes complicate comparisons between certain studies or populations. This study illustrated that inclusion or exclusion of a single symptom substantially affects exacerbation-related outcomes and, therefore, choosing a certain definition truly matters. These findings emphasise that inconsistent use of definitions leads to widely varying exacerbation-related outcomes; 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, 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 in discriminating between stable periods and exacerbations. Furthermore, it was internally 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 are 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 in 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.

In conclusion, 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 organisations and investigators increasing their efforts to reach a consensus on operational symptom-based definitions and quality requirements for counting and analysing exacerbations.

STATEMENT OF INTEREST
None declared.

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, *et al.* 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, *et al.* 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, *et al.* Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax* 2005; 60: 925–931.

6 Miravitlles M, Murio C, Guerrero T, *et al.* 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, *et al.* 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, *et al.* 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; 3: 965–975.

15 Leidy NK, Wilcox TK, Jones PW, *et al.* Standardizing measurement of COPD exacerbations: reliability and validity of a patient-reported diary. *Am J Respir Crit Care Med* 2010; [Epub ahead of print DOI: 10.1164/rccm.201005-0762OC].

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

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, *et al.* 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, *et al.* Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med* 1987; 106: 196–204.

22 Alvarez-Mon M, Miravitlles M, Morera J, *et al.* Treatment with the immunomodulator AM3 improves the health-related quality of life of patients with COPD. *Chest* 2005; 127: 1212–1218.

23 Ekberg-Jansson A, Larson M, MacNee W, *et al.* N-isobutyrylcysteine, a donor of systemic thiols, does not reduce the exacerbation rate in chronic bronchitis. *Eur Respir J* 1999; 13: 829–834.

24 Van Schayck CP, Dompeling E, van Herwaarden CL, *et al.* Continuous and on demand use of bronchodilators in patients with non-steroid dependent asthma and chronic bronchitis: four-year follow-up randomized controlled study. *Br J Gen Pract* 1995; 45: 239–244.

25 Sund ZM, Powell T, Greenwood R, *et al.* Remote daily real-time monitoring in patients with COPD – a feasibility study using a novel device. *Respir Med* 2009; 103: 1320–1328.

26 Banerjee D, Khair OA, Honeybourne D. The effect of oral clarithromycin on health status and sputum bacteriology in stable COPD. *Respir Med* 2005; 99: 208–215.

27 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.

28 Casaburi R, Mahler DA, Jones PW, *et al.* A long-term evaluation of once-daily inhaled tiotropium in chronic obstructive pulmonary disease. *Eur Respir J* 2002; 19: 217–224.

29 Dowson LJ, Guest PJ, Stockley RA. The relationship of chronic sputum expectoration to physiologic, radiologic, and health status characteristics in α_1-antitrypsin deficiency (PiZ). *Chest* 2002; 122: 1247–1255.

30 Pela R, Calcagni AM, Subiaco S, *et al. N*-acetylcysteine reduces the exacerbation rate in patients with moderate to severe COPD. *Respiration* 1999; 66: 495–500.

31 Dahl R, Chung KF, Buhl R, *et al.* Efficacy of a new once-daily long-acting inhaled β2-agonist indacaterol *versus* twice-daily formoterol in COPD. *Thorax* 2010; 65: 473–479.

32 Bhowmik A, Seemungal TA, Sapsford RJ, *et al.* Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. *Thorax* 2000; 55: 114–120.

33 Wilkinson TM, Donaldson GC, Hurst JR, *et al.* Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004; 169: 1298–1303.

34 Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. *Eur Respir J* 2003; 21: Suppl. 41, 46s–53s.

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