PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The rate of normal lung function decline in ageing adults: a systematic review of prospective cohort studies
AUTHORS	Thomas, Elizabeth; Guppy, Michelle; Straus, Sharon; Bell, Katy; Glasziou, Paul

VERSION 1 - REVIEW

REVIEWER	Shyamali Dharmage The University of Melbourne
REVIEW RETURNED	12-Dec-2018

GENERAL COMMENTS	The paper aimed to summarize data on lung function decline rates in healthy population. Major comments: 1. Search strategy In search strategy in page 60, there were some typos in search syntaxes e.g. “Forced Vital Capacity, Timed”; “FEVt”; “Forced Vital Capacity, Timed”; “Timed Vital Capacity”; “Capacity, Timed Vital”, etc in No8. These phrases give no hits in Pubmed. I wonder why the authors used phrases such as: oximetry, blood gas analysis, bronchial provocation test, etc in No8 as they are not really relevant (the inclusion criterion focused on (i.e. FEV1, FVC, peak expiratory flow rate [PEFR]) I just try with a simpler but more comprehensive syntax for No8. ((((((((((“forced expiratory volume”[Title/Abstract]) OR FEV[Title/Abstract]) OR "forced vital capacity"[Title/Abstract]) OR FVC[Title/Abstract]) OR spirometry[MeSH Terms]) OR spirometry[Title/Abstract]) OR "lung function"[Title/Abstract]) OR "pulmonary function"[Title/Abstract])) OR "Expiratory Flow"[Title/Abstract]) There are 91792 hits. No6: (Humans[Mesh] OR Humans[tiab] OR Human[tiab] OR Population[tiab]). Sometimes these key words are not present in title/abstract or indexed. I suggest to use them in the filter function and compare results. No4 and No5 are both about study design but why are they separate? I just wonder why the authors used phrases:

	Compared[tiab] OR Investigated[tiab] OR Evaluating[tiab] OR Analysis[tiab] OR Analyzed[tiab] OR Statistics[tiab] OR Data[tiab] OR Baseline[tiab]) in No5 Because the authors combined No4 AND No5, papers were likely to be missed. Relevant papers were also missed because the authors combined No1 AND No3 with others. Although the authors aimed to search studies reporting lung function decline in older populations, I am concerned that key words in No1 and No3 are not frequently present in title/abstract or indexed as Mesh terms. 2. Three key inclusion criteria included:1-Participants did not have a known risk factor for respiratory disease (such as smoking, occupational inhalation), though studies could have included a comparator arm with participants with risk factors;2- Participants without respiratory symptoms such as wheeze, dyspnoea, chronic cough;3- Participants without known respiratory disease (chronic airways disease, asthma). These criteria were very hard to assessed in all studies. Although some studies reported stratified lung function decline rates for smoking status and the authors could extract decline rate among never smokers, information about symptoms, respiratory diseases or occupation exposure were simply not reported in all studies. Thus, this information could not be assessed; The authors may need to revise these criteria as well as the legend of table 2. 3. One key inclusion criterion was: Longitudinal studies that followed adults past the age of 65 years. Was this mean/median age, minimum age or maximum age among the study sample? For example, in the study by Bartholomew (1998) in table 1, mean age at baseline was 41.6±16.1 and study duration was 6 years. What were participants' age at end of follow-up? Minor comments Table 1 and 2: please label the column "mean age" clearer. The authors stated that "We also accounted for the proportion of the cohort that subsequently developed symptoms or disease during the course of the follow-up". This is hardly feasible because studies rarely report diseases and symptoms at end of follow-up. I cannot find some data in table 2 from original papers. Did the authors contact papers' authors to get data, for example the decline rate -22.4 (36.4) from Triebner 2017 and -25.8 (14.0) from Liao 2015. SDs are different from original papers. I cannot find decline rates for FEV1 in table 2 from the original paper. Did the authors calculate them? If so, a footnote should be included in table 2. Column "confounding variables" in table 2: if smoking was adjusted as a confounder, the estimated decline rate was not in never smokers.

	Contents of section “comparison with previous research” in the discussion does not match with the subheading. The phrase “age-specific decline rates by decade of age” is a bit not clear in text. It can imply baseline age or age at midpoint. Can you authors make it clearer how the age specific decline rates were calculated in table 3? The conclusion that “The decline in absolute and relative lung function parameters also accelerates with age” (lines 16-17 page 25) is hard to drawn as only 3 studies had these data and one of them did not show an accelerated decline. Lines 28-30: Dose the phrase “baseline organ function, organ function measurements” mean “…lung…”? The authors should improve the legends of figures 3A, 3B in pages 45 and 46 for the ease of understanding, e.g. the size of circles. Error bars may be easier to understand.

REVIEWER	Mark L Levy Self employed sessional GP, United Kingdom
REVIEW RETURNED	15-Dec-2018

GENERAL COMMENTS	Thank you for this review which may help change the GOLD (& other) guideline committee to abandon the ill advised use of 70% as an absolute cut off for diagnosing post-bronchodilator airflow limitation & COPD. Some comments for consideration: i) The authors rightly draw attention to the decline in lung function with age and conclude that further longitudinal work is required to inform calculation of normal values (LLN in particular), however they don’t discuss whether they have tested the rates of decline in their study, against the normal values currently used by the major spirometer producers. In other words, does this work contribute new practical, useable data or supplement current knowledge related to decline in lung function by gender, ethnicity relative to age? ii) Page 9 line 23, and page 11: Inclusion criteria included studies that measured spirometry and PEF: An additional possible source of bias relates to the quality of spirometry – was it quality assured? Regarding PEF – it is known that different meters vary between others in consistency and accuracy – did the authors ascertain whether the same PEF meter was used in patients included in the studies? Furthermore as many patients over 80 are diagnosed by GPs with COPD, it is worth adding that these included studies only report data on patients below 80 years of age (according to the tables in the paper) These factors should be included in the section on inclusion criteria and as possible sources of bias in the discussion & conclusions. iii) Page 27, line 49: The ‘Horse racing effect’ may explain the heterogeneity; however so might the possibility of poor technique or poorly maintained or calibrated equipment. Another possible factor would include undiagnosed COPD (or other lung disease at entry, or patients who developed late onset respiratory, cardiac,

	renal or other diseases during the course of the studies (those diagnosed would probably have been excluded, but who knows??) (PS I don't really grasp the analogous reference to horse racing- is this just me, or is it a bad analogy??) iv) Guidelines recommend (inappropriately in my view - see reference 10) that an absolute cut off of 70% ratio of FEV1 to FVC is used to diagnose airflow obstruction for the purpose of diagnosing COPD. As few of the papers reported the FEV1/FVC ratio and that the raw data does not appear to be available: do FEV1 and FVC independently decline proportionately with age?

VERSION 1 – AUTHOR RESPONSE

1. Reviewer 1. Major comments: Search strategy In search strategy in page 60, there were some typos in search syntaxes e.g. “Forced Vital Capacity, Timed”; “FEVt”; “Forced Vital Capacity, Timed”; “Timed Vital Capacity”; “Capacity, Timed Vital”, etc in No8. These phrases give no hits in Pubmed. I wonder why the authors used phrases such as: oximetry, blood gas analysis, bronchial provocation test, etc in No8 as they are not really relevant (the inclusion criterion focused on (i.e. FEV1, FVC, peak expiratory flow rate [PEFR]) I just try with a simpler but more comprehensive syntax for No8. ((((((((((“forced expiratory volume”[Title/Abstract]) OR FEV[Title/Abstract]) OR “forced vital capacity”[Title/Abstract]) OR FVC[Title/Abstract]) OR spirometry[MeSH Terms]) OR spirometry[Title/Abstract]) OR “lung function”[Title/Abstract]) OR “pulmonary function”[Title/Abstract])) OR “Expiratory Flow”[Title/Abstract]) There are 91792 hits. No6: (Humans[Mesh] OR Humans[tiab] OR Human[tiab] OR Population[tiab]).	# Thank you for this suggested search line for lung studies. We have incorporated this in our updated search. Along with your suggested “lung function” string, we made a slight modification to the search strategy to include one more MeSH term for follow up studies. With this adjustment we found all of the possible studies with indexed titles, abstracts and MeSH terms. # We tested the search strategy using the filter as suggested, but found this to be less sensitive and missing studies. # Line 4 is a study design filter, Line 5 is a line	We screened an additional 1881 studies and only identified one additional study. The new study by Luoto et al was published in December 2018.

Sometimes these key words are not present in title/abstract or indexed. I suggest to use them in the filter function and compare results. No4 and No5 are both about study design but why are they separate? I just wonder why the authors used phrases: Compared[tiab] OR Investigated[tiab] OR Evaluating[tiab] OR Analysis[tiab] OR Analyzed[tiab] OR Statistics[tiab] OR Data[tiab] OR Baseline[tiab]) in No5 Because the authors combined No4 AND No5, papers were likely to be missed. Relevant papers were also missed because the authors combined No1 AND No3 with others. Although the authors aimed to search studies reporting lung function decline in older populations, I am concerned that key words in No1 and No3 are not frequently present in title/abstract or indexed as Mesh terms.	which detects if data analysis took place. Although these look similar, they capture more studies with these different search concepts. # We tested this hypothesis and after adding the MeSH term for "follow-up studies" it found all of eligible studies. # Once again we tested this hypothesis and found it to be incorrect, as the 2 lines run separately also found all the relevant studies.	
2. Three key inclusion criteria included: 1-Participants did not have a known risk factor for respiratory disease (such as smoking, occupational inhalation), though studies could have included a comparator arm with participants with risk factors; 2- Participants without respiratory symptoms such as wheeze, dyspnoea, chronic cough; 3- Participants without known respiratory disease (chronic airways disease, asthma). These criteria were very hard to assessed in all studies. Although some studies reported stratified lung function decline rates for smoking status and the authors could extract decline rate among never smokers, information about symptoms, respiratory diseases or occupation exposure were simply not reported in all studies. Thus, this	We agree that not all of the studies reported this information. We have clarified the statement of our selection criteria, so that this now specifies that where studies specifically included participants with symptoms, respiratory disease or known risk factors, we excluded these studies.	"We excluded studies if the participants did not meet the pre-specified age criteria; if the population of interest were reported to include smokers or those with risk factors such as occupational inhalation; if participants were reported to have respiratory symptoms such as wheeze, dyspnea or chronic cough; if the study included participants with known respiratory disease such as asthma or COPD." Methods, Search strategy and inclusion criteria, Page 6

information could not be assessed; The authors may need to revise these criteria as well as the legend of table 2.		Figure 2 legend now reads “Table 2. Reported annual rates of absolute and relative lung function decline (FEV1, FVC, PEFR, FEV0.75) in 16 prospective cohort studies.” Page 31.
3. One key inclusion criterion was: Longitudinal studies that followed adults past the age of 65 years. Was this mean/median age, minimum age or maximum age among the study sample? For example, in the study by Bartholomew (1998) in table 1, mean age at baseline was 41.6±16.1 and study duration was 6 years. What were participants’ age at end of follow-up?	We included studies where any or all of the participants were followed past the age of 65. Therefore we used maximum age of the sample. In the study by Bartholomew et al, the original paper reports in Table 1 that the age of the female sample was 46.3 (SD 15.6). Since they were followed up for six years, some of this sample would have reached and/or passed the age of 65 during the course of the study.	First inclusion criteria now reads “Longitudinal studies that followed some or all of the adult participants past the age of 65 years” Methods, Search strategy and inclusion criteria, Page 6
4. Minor comments: Table 1 and 2: please label the column “mean age” clearer.	We have edited the column headings in both tables to make it clearer.	This now reads “Mean age of sample (years, SD)” Table 1, Page 10 Table 2, Page 13
5. The authors stated that “We also accounted for the proportion of the cohort that subsequently developed symptoms or disease during the course of the follow-up”. This is hardly feasible because studies rarely report diseases and symptoms at end of follow-up.	Thank you for pointing this out. We aimed to report if any of the participants subsequently developed symptoms or disease or died during the course of follow up. Only two studies (Proctor et al, Lange et al) reported these outcomes. We have changed the wording of the sentence to reflect this.	“We also aimed to report the proportion of the cohort that subsequently developed symptoms or disease during follow-up”. Methods, Study selection and data extraction, Page 7
6. I cannot find some data in table 2 from original papers. Did the authors contact papers’ authors to get data, for example the decline rate -22.4 (36.4) from Triebner 2017 and -25.8 (14.0)	We did contact one author to obtain original data (Triebner). For clarification, we have attached a supplementary file with the details of how	See supplementary file 3

from Liao 2015. SDs are different from original papers.	each result was extracted or calculated.	
7. I cannot find decline rates for FEV1 in table 2 from the original paper. Did the authors calculate them? If so, a footnote should be included in table 2.	Please see above comment.	See supplementary file 3
8. Column “confounding variables” in table 2: if smoking was adjusted as a confounder, the estimated decline rate was not in never smokers.	Thank you for this comment, the term confounding variables here appears to be a misnomer. This column was supposed to refer to studies that compared different populations (i.e. those that smoked, different ethnicities, BMI) to the reference population (asymptomatic, non-smokers). The title of the column has been modified for clarification.	Column heading now reads “Variables reported to alter the rate of change”. Table 2, Page 13
9. Contents of section “comparison with previous research” in the discussion does not match with the subheading.	We have added a sentence at the beginning of this paragraph to better link the contents of this paragraph to the subheading.	The paragraph now begins with “To date, there have been no systematic reviews or meta-analyses examining the rate of lung function decline with age, to assess the potential impact of the fixed threshold on COPD misdiagnosis.” Discussion, Comparison with previous research, Page 23
10. The phrase “age-specific decline rates by decade of age” is a bit not clear in text. It can imply baseline age or age at midpoint. Can you authors make it clearer how the age specific decline rates were calculated in table 3?	Burchfiel et al reported the average annual rate of FEV1 decline based on the age at baseline (reported in Table 3 in the original paper). Pearson et al, and Luoto et al also reported the change rates for each decade of age, according to baseline age.	The column heading of the table now reads “Baseline age (years)” A footnote has been added. “The estimates from Burrows were derived from formulae modelling change in FEV1 with age. See Supplementary File 3 for calculations.” Table 3, Page 16.

	The decline rates in Burrows et al, were derived from two formulae modelling the predicted change in FEV_1 for males and females. As shown in the newly added supplementary file of calculations, the rate of decline was calculated at each 5-year time point and these were combined to fit the decade-specific analysis.	
11. The conclusion that "The decline in absolute and relative lung function parameters also accelerates with age" (lines 16-17 page 25) is hard to drawn as only 3 studies had these data and one of them did not show an accelerated decline.	Overall, six studies reported the rates of decline with different age groups, however only three of these studies conformed to a decade-specific analysis. Two of these three studies demonstrated an accelerated decline (the remaining study had a very small sample size). Another study demonstrated increased relative decline with age (though this was not observed for absolute decline). However, two other studies (Bartholomew et al, Lange et al) that used different age groups to our analysis also demonstrated that the rates of decline increase with age. We have presented this descriptively in the section "Age-specific lung function decline by decade of age", after Table 3.	"An age-specific analysis suggests that the rate of FEV1 function decline may accelerate with each decade of age." Abstract, Conclusions, Page 2 Discussion, Statement of principal findings, Page 20
12. Lines 28-30: Dose the phrase "baseline organ function, organ function measurements" mean "…lung…"?	Thank you, this has now been clarified.	Now reads "baseline lung function, lung function measurements" Methods, Study selection and data extraction, Page 7

13. The authors should improve the legends of figures 3A, 3B in pages 45 and 46 for the ease of understanding, e.g. the size of circles. Error bars may be easier to understand.	Thank you for this suggestion. For clarity, we have re-phrased the legends of Figure 3A and 3B.	"Figure 3A. The rate of FEV1 decline in twelve study populations by years of follow-up. The size of the circle corresponds to individual study sample size. Figure 3B. Sensitivity analysis, excluding studies with less than ten years of follow-up. The size of the circle corresponds to individual study sample size." Page 31
Reviewer 2: Thank you for this review which may help change the GOLD (& other) guideline committee to abandon the ill advised use of 70% as an absolute cut off for diagnosing post-bronchodilator airflow limitation & COPD.	Thank you for your comments.	
1. The authors rightly draw attention to the decline in lung function with age and conclude that further longitudinal work is required to inform calculation of normal values (LLN in particular), however they don't discuss whether they have tested the rates of decline in their study, against the normal values currently used by the major spirometer producers. In other words, does this work contribute new practical, useable data or supplement current knowledge related to decline in lung function by gender, ethnicity relative to age?	Thank you for this suggestion, which we agree would contribute an interesting layer of analysis to our study. We analysed the rate of decline in the reference values provided by the NHANES III study (which has provided reference values for major spirometer producers) and included this in our discussion.	We have now added: "Spirometers used in practice commonly derive their reference values from the National Health and Nutrition Examination Survey (NHANES), a cross-sectional study which was conducted in the USA between 1988 – 1994. Though the predicted values do reflect a decline in FEV1 and FEV1/FVC with age, these decline rates may not be as reliable as the estimates from longitudinal studies included in our review. According to the NHANES III, the median rate of FEV1 decline for a Caucasian male of 1.75m aged between 30-80 is 32ml/year and a female with an average height of 1.6m has an FEV1 that declines a median of 25ml/year. Both of these estimates are lower than

		the median FEV1 decline of the studies in our review, which was 43.5ml/year and 30.5ml/year for men and women respectively. Therefore the predicted age-specific lung function used in spirometers may often mislabel people as having abnormal lung function when they are actually within normal limits" Discussion, Meaning of the study: possible explanations and implications for clinicians and policymakers, Page 24.
2. Page 9 line 23, and page 11: Inclusion criteria included studies that measured spirometry and PEF: An additional possible source of bias relates to the quality of spirometry – was it quality assured? Regarding PEF – it is known that different meters vary between others in consistency and accuracy – did the authors ascertain whether the same PEF meter was used in patients included in the studies? Furthermore as many patients over 80 are diagnosed by GPs with COPD, it is worth adding that these included studies only report data on patients below 80 years of age (according to the tables in the paper) These factors should be included in the section on inclusion criteria and as possible sources of bias in the discussion & conclusions.	Thank you for raising this important comment. Only eight of the included studies reported on the reproducibility and acceptability of their spirometry measurements. We cannot be sure that the two studies that measured PEFR used the same peak flow meter as one of the studies (Proctor et al) did not comment on this. We also agree with the suggestion that there is a requirement for more studies in the elderly as many patients over 80 are diagnosed with COPD by GPs. We have added these suggestions to our discussion in the limitations section.	Now added "Quality of spirometry, as well as properly maintained and calibrated equipment is another source of bias, which may have contributed to variation in the results ….did not specify which peak flow meter they used." Discussion, Strengths and weaknesses of the study, Page 22 Also added "Our study aimed to examine the rate of lung function change in the elderly, however the majority of included studies did not focus on this age group. COPD misdiagnosis particularly affects those older than 80 years of age, therefore more studies are required in the elderly." Discussion, Strengths and weaknesses of the study, Page 21
3. Page 27, line 49: The 'Horse racing effect' may explain the heterogeneity; however so might	Thank you for this helpful comment and we have included these other	"Variation within the results may be also explained by the "horse-

	the possibility of poor technique or poorly maintained or calibrated equipment. Another possible factor would include undiagnosed COPD (or other lung disease at entry, or patients who developed late onset respiratory, cardiac, renal or other diseases during the course of the studies (those diagnosed would probably have been excluded, but who knows??) (PS I don't really grasp the analogous reference to horse racing- is this just me, or is it a bad analogy??)	reasons as possible causes of the heterogeneity. We have also tried to make the 'horse racing' analogy more explicit – which we agree might otherwise be a bit obtuse.	racing effect", where an initially low FEV1 measurement may reflect a greater loss of function in the preceding years and hence predicts faster decline in subsequent years (just as the position of the horse in halfway through the race is related to its speed in the early part of the race and hence speed for the final part of the race." Discussion, Strengths and weaknesses of the study, Page 23
4.	Guidelines recommend (inappropriately in my view - see reference 10) that an absolute cut off of 70% ratio of FEV1 to FVC is used to diagnose airflow obstruction for the purpose of diagnosing COPD. As few of the papers reported the FEV1/FVC ratio and that the raw data does not appear to be available: do FEV1 and FVC independently decline proportionately with age?	While this would have been an interesting addition to our analysis, it is difficult to say whether the declines in FEV1 and FVC are proportional since only a paucity of studies measured both outcomes. In studies that report both FEV1 and FVC decline, one study (Triebner et al), reports the FEV1 declines at a higher rate than FVC, but in other studies (Ahmadi, Bartholomew, Griffith) the FVC declines at a faster rate (See Table 2). We really require longitudinal studies that specifically measure the FEV1/FVC for a more reliable measure. We have made a comment in this in our discussion.	"Five studies separately measured changes in both FEV1 and FVC, however is difficult to conclude whether the rate of decline in FEV1 and FVC is proportional. Out of the four studies that reported both FEV1 and FVC decline, only one study demonstrated that FEV1 declines faster than FVC, but in the three remaining studies, the FVC declines at a faster rate (See Table 2). Longitudinal studies that specifically measure the FEV1/FVC would provide the most reliable measure of this decline." Discussion, Strengths and weaknesses of the study, Page 20-21.

VERSION 2 – REVIEW

REVIEWER	Mark L Levy Locum General Practitioner, London , United Kingdom
REVIEW RETURNED	05-Mar-2019

GENERAL COMMENTS	The revision has addressed the concerns/points I raised in my first review off this paper. The results do provide a clear directive for the GOLD Committee to include LLN in the definition of clinically significant airflow obstruction. I suggest this is included in an editorial statement.