TITLE: Clinically important differences in the intensity of chronic refractory breathlessness.

AUTHORS:

Miriam J Johnson, MD. Professor of Palliative Medicine, Hull York Medical School, University of Hull, UK

J Martin Bland, PhD. Professor for Medical Statistics, Department of Health Sciences, University of York, Heslington, York, UK

Stephen G Oxberry, PhD. Consultant in Palliative Medicine, Kirkwood Hospice, Huddersfield, UK

Amy P Abernethy, MD. Associate Professor, Division of Medical Oncology, Department of Medicine, Duke University Medical Centre, Durham, USA.

Discipline, Palliative and Supportive Services, Flinders University Australia.

David C Currow, FRACP. Professor, Discipline, Palliative and Supportive Services, Flinders University, Adelaide, Australia.

AUTHOR FOR CORRESPONDENCE: M. J. Johnson, Hertford Building, The University of Hull, Cottingham Rd, Hull, UK. HU6 7RX Tel: +44(01482 463482 Fax: +44(0)1482 464705; Email: miriam.johnson@hyms.ac.uk

Tables: 2

Figures: 1

References: 30

WORD COUNT: 2296

RUNNING TITLE: **MCID; chronic refractory breathlessness**

ABSTRACT

Context: Clinically important differences in chronic refractory breathlessness are ill-defined, but important in clinical practice and trial design.

Objectives: To estimate the clinical relevance of differences in breathlessness intensity using distribution and patient anchor methods.

Methods: A retrospective data analysis from 213 datasets from four clinical trials for refractory breathlessness. Linear regression was used to explore the relationship between study effect size and change in breathlessness score (0 – 100mm visual analogue scale) and to estimate the change in score equivalent to small, moderate and large effect sizes.

Pooled individual blinded patient preference data from three randomized controlled trials were analysed. The difference between the mean change in day 4 minus baseline scores between preferred and non-preferred arms was calculated.

Results: There was a strong relationship between change in score and effect size, ($p=0.001$; $R^2=0.98$). Values for small, moderate and large effects were −5.5mm, −11.3mm and −18.2mm. The participant preference change in score was −9mm (95% CI −15.8 to −2.1) ($p = 0.008$).

Conclusions: This larger dataset supports a clinically important difference of 10mm. Studies should be powered to detect this difference.

WORD COUNT: 181

KEY WORDS: Minimal clinical important difference; breathlessness; dyspnoea; opioid

INTRODUCTION

Chronic refractory breathlessness is a challenging symptom for those who experience it, their caregivers and clinicians. Assessment sufficient to monitor clinical benefit and for clinical trial outcome measurement is difficult because of the multi-factorial, subjective nature of the symptom (1-5), demonstrated by the plethora of measurement tools . There is consensus that the 0 -100mm visual analogue scale [VAS], 0 – 10 numerical rating scale [NRS] and modified Borg scale are appropriate uni-dimensional measures of chronic breathlessness intensity. (3;5;6) Despite this, there are few data-derived estimates of the minimal clinically important difference (MCID) for chronic refractory breathlessness which is needed to design studies with adequate power to determine a statistically *and* clinically significant change. The MCID, different from a statistically significant difference, is defined as the change in a breathlessness intensity score discernable to a patient as indicative of improvement or deterioration.(7)

There are two main methods of identifying an MCID for an outcome measure; distribution and anchor-based. A full description is beyond the scope of this paper, but useful summary discussions are available.(8;9)

Distribution based calculation of MCID in chronic breathlessness

The effect size (change after intervention divided by standard deviation of baseline scores) can be used as a statistical analysis method to determine the lowest meaningful detectable difference based on the underlying variability of the reported data. The smaller the group data variability, the greater the "clinical precision" in the change as detected by the individuals studied. Furthermore, if there is a true relationship between change in measure and effect size, then different study populations should have a similar level of precision over the

detection of the change in measure under study. However, it can be difficult to judge whether such detectable differences are considered relevant by the patients concerned.

Anchor-based methods of MCID in breathlessness

A patient rated global impression of change can provide an anchor to help define a small, moderate or large change.(10) However, variation amongst and between individuals,(10)or the measurement error of the measurement tool is not taken into account.(7;8) Also, by definition as a single estimate using recall estimation, the global rating of change has an un-measurable baseline error. Recent approaches have tried to account for this by giving ranges for the MCID.(7)

In recognition of the limitations of each method, we present anchor- and distribution-based estimates of the MCID for chronic refractory breathlessness in order to frame the categorisation of effect size in the context of patient report. The aim of this paper is to quantify the clinically noticeable difference in scores, and not the *mechanism* whereby the intervention effected that change. However, the studies contributing to this work were all clinical trials of opioids with the relief of chronic refractory breathlessness as a primary outcome.(11-14)

METHOD

This was a retrospective analysis of data from a total of 178 clinical trial participants with chronic refractory breathlessness from four clinical trials of morphine or oxycodone for the management of breathlessness. Of these participants, 35 provided data from separate periods of exposure to two different opioids to give a total of 213 sets of data(12) and were treated as

two studies for the distribution method calculation. Baseline characteristics of participants, and the NRS/VAS used, are seen in Table 1. Confirmation was obtained that NHS ethical permission was not required for secondary analysis of anonymised data.

The methods of the four contributing studies are described in detail elsewhere.(11-14) In summary, three were placebo controlled, randomised, crossover trials, and compared 4 days of morphine with 4 days of placebo.(11-13) Oxberry et al added a third arm with 4 days of oxycodone.(12) One was a cohort study examining breathlessness response to morphine dose increments, with respondents graduating to a pharmacovigilance study in participants with breathlessness that improved with morphine.(14) Two studies included only those with CHF; in the other studies, participants' breathlessness was predominantly due to chronic obstructive pulmonary disease.(13;14) All participants had chronic refractory breathlessness.

Three studies used the VAS to measure breathlessness intensity and the others used NRS. The NRS intensity scores were represented as equivalent 0 – 100mm in the combined dataset; NRS and VAS appear to be reported by patients in a linear fashion in both breathlessness and pain. (15;16)

Statistical analysis

The MCID was calculated using two methods.

The distribution-based approach calculated the effect size (change after opioid divided by standard deviation of baseline scores) in each study. Only data from the opioid arm of the placebo controlled studies were used.

Cohen's guidelines for interpreting small, moderate and large effect sizes were used (sizes chosen as small = 0.25 medium = 0.50, large = 0.80).(17) Assessment of the linear relationship between change in breathlessness score and the effect size was explored using a linear regression model to determine the change in VAS score equivalent to the three effect sizes.

The anchor-based method is from analysis of a pooled dataset of individual participant data from randomised controlled trials. The placebo controlled RCTs had data for blinded patient preference for the arm of the study which they felt provided better benefit for breathlessness for all participants. Participants in the Oxberry study had data for each opioid intervention (morphine and oxycodone arms), but only one set of data on placebo.
Participant perception of change in breathlessness intensity (day 4 minus baseline VAS scores) was examined in relation to preferred study arm (opioid or placebo). The difference between the mean change in breathlessness scores in preferred study arm and the mean change in the non-preferred arm was calculated.

RESULTS

Distribution calculation

The effect sizes for each study, ranging from 0.1 to 0.6, are shown in Table 2, and the regression line shown in Figure 1. The regression equations for the differences in VAS breathlessness score and effect sizes 0.25, 0.5, 0.8 are shown below. The slope of the line was

statistically significant (p = 0.001) and the percentage variation of the data explained by the fitted line was extremely high (R-squared = 0.98).

Small change: $0.2380574 + 23.0974 \times (-0.25) = -5.5$

Moderate change: $0.2380574 + 23.0974 \times (-0.5) = -11.3$

Large change: $0.2380574 + 23.0974 \times (-0.8) = -18.2$

Blinded patient preference calculation

In total, there were 113 evaluable preference responses from the 93 study participants. A stated preference for an intervention arm was given in 93 (62 drug; 31 placebo). Eighteen participants had no preference for any intervention arm. The Day 4 minus baseline VAS change in the preferred arm (−13mm; SD 24.7mm) was greater than in the non-preferred arm (−5mm; SD 19.6mm). Two did not provide data for the opioid arm, and two did not provide data for the placebo arm although data were given for their preferred arm. The VAS difference calculated from 93 datasets with data for both preferred and non-preferred arms was −9mm (95% CI −15.8 to −2.1) (p = 0.008).

DISCUSSION

Using a distribution method, we have estimated the changes in VAS breathlessness intensity scores that relate to a small, moderate or large improvement in symptoms using individual patient data generated in four studies that evaluated the effect of opioids on chronic refractory breathlessness. In keeping with recommendations to estimate MCID by using both a distribution method and a patient data anchor method, we have used the blinded patient preference data from the placebo-controlled, crossover RCTs. This recognises that there is a distinction between a *clinically detectable* change and a *clinically important* difference. Our data estimates that a reduction in VAS scores of 9mm represents a change that patients consider significant enough to warrant choice of one intervention over another. Although there are wide confidence intervals and relatively modest sample size for such a calculation, the present study includes data from more participants than the only other published anchor-based estimate in chronic refractory breathlessness. (7) The use of blinded patient preference data from RCTs is a novel patient-defined, anchor-based approach and reflects a clinically important difference, without the difficulties in analysis inherent relying only on the global impression of change.

Rationale for using both anchor based and distributed based methods

The distinction between a detectable difference and a clinically important difference is difficult to deduce from distribution methods alone. (17) Cohen's grading system was devised in the context of psychological research that shares parallel challenges of multi-factorial and subjective study endpoints, so its application to chronic breathlessness is reasonable.

Although the size of the clinical effect can be estimated using the grading suggested by Cohen, this can be criticised as being fairly arbitrary. The terms "small", "moderate" and "large" are relative to each other and will be different for individuals and perhaps also within

sub-populations. This should be taken into consideration when interpreting results. This is illustrated by data from a total of 95 participants from five studies in people with COPD where a VAS was used as the outcome measurement (18); three studies of pulmonary rehabilitation (19-21); and one each in oxygen supplementation(22); and (23) lung reduction surgery. Ries et al found effect sizes ranging between 0.48 (11.9 mm) and 1.76 (30.3mm) equating to a moderate to large improvements.(18) However, the participants varied considerably with regard to gender split (one study was solely of men), baseline intensity of breathlessness (25mm to 79.6mm) and intervention.(18) For example, a patient's perception of what constitutes a large change may depend on how long they have been living with daily breathlessness, limitations to their activities of daily living due to breathlessness and the burden of the intervention. Net clinically important benefit may be different for those with severe baseline breathlessness who underwent lung reduction surgery as a last desperate attempt to reduce symptoms, when compared to those with similar baseline breathlessness intensity who attended pulmonary rehabilitation classes even though both approaches have been demonstrated to have large effects.

Thus anchor based methods, with their direct use of the patient's subjective experience with regard to whether the symptom is worse or better, adds further information to the effect size, helping to confirm that the detected difference does indeed have clinical significance to the patient. In our data, we can see the mean change that was considered by study participants to be significant enough to warrant a choice between two interventions, and the figure of 9mm lies very near the 'moderate effect size' as calculated by the distribution method.

Comparison with other published estimates for the MCID for breathlessness

The MCID for *acute* breathlessness due to asthma(24) or acute decompensated heart failure(25) has been estimated using anchor based methods (global impression of change),

where a difference of 21.1 mm and 22mm respectively on a 100mm VAS was found to be equivalent to subjectively "feeling a little less breathless". Our data indicate that the MCID is less than this for people with *chronic* breathlessness. These people had more intense breathlessness when they were assessed initially and might expect, at least in asthma, to return to no breathlessness when successfully treated. In acute breathlessness, both absolute and relative improvements were almost double the levels needed in chronic breathlessness (where there may be little expectation of breathlessness returning to zero). In addition, acute breathlessness is usually accompanied by severe distress and is likely to represent a very different patient experience than that of a person who experiences breathlessness daily at rest or on minimal exertion over many months or years.

People with chronic breathlessness tend to adapt to limitations caused by breathlessness over time in a resourceful and stoical manner, which can render it less visible as a symptom.(26) Medical care may only be sought when there is a crisis, with little expectation that interventions could provide clinically relevant benefit in its chronic state when a change in intensity equating to a smaller change in the absolute VAS score could have a clinically meaningful benefit for individuals coping with daily breathlessness.(27) In addition, clinicians may perceive chronic breathlessness which persists despite optimally tolerated management of the underlying disease as difficult to treat and inevitable.(28;29) Such therapeutic nihilism may mean that optimal assessment and treatment does not occur, and clinical focus is directed to other symptoms considered more responsive to interventions. Recognition that even a relatively small improvement in the intensity of breathlessness is perceived by the patient as a useful change may drive a more systematic approach to breathlessness management.

Comparison with other published estimates for the MCID in chronic breathlessness

Ries et al estimated that a change of 10 – 20 mm VAS represented a moderate change using their distribution methods.(18) Our data supports this, as 10mm change in VAS is just below the value of a moderate change (distribution method) and just greater than the clinically important difference measured by patient anchor. Oxberry et al reported a moderate effect size (0.5) using a distribution method.(7) Assuming 10mm VAS is equivalent to 1 point NRS, values from both distribution (5.5mm small; 11.3mm moderate; 18.2mm large) and anchor methods (9mm) in this paper fall within the range (0.5 – 2.0 for average/24 hours NRS score) calculated from the anchor based estimation in the Oxberry study.

Our combined data used VAS/NRS measures framed as "breathlessness now" and "average breathlessness over the past 24 hours". Given these are both measures in the context of chronic daily breathlessness, rather than acute exacerbations (we did not use the "worst breathlessness over the past 24 hours" measure that we also had in the Oxberry dataset) we have made the assumption that an absolute change in either will reflect a clinically important difference in chronic refractory breathlessness. However, this should be taken into account when interpreting our findings.

The consensus statement recommends that a clinically important difference for chronic refractory breathlessness scores is 1 point on an NRS scale or 10mm on a VAS scale.(30) Our data supports this value confirming that this level of improvement should be the therapeutic target. Studies of chronic refractory breathlessness should be powered to detect this difference.

DISCLOSURES AND ACKNOWLEDGMENTS.

The Authors have no conflict of interest in relation to this manuscript

ACKNOWLEDGEMENTS With thanks to Ms Tracey Hawkes for able assistance in combining the datasets.

CRITERIA FOR AUTHORSHIP: MJ and DC conceived the question and design; MJ, DC, SO and AA provided anonymised data; MJ and MB conducted the analysis; MJ, MB, AA, SO and DC drafted the manuscript. All authors agreed the final manuscript.

FUNDING: none

REFERENCES

(1) Johnson MJ, Oxberry SG, Cleland JG, Clark AL. Measurement of breathlessness in clinical trials in patients with chronic heart failure: the need for a standardized approach: a systematic review. Eur J Heart Fail 2010 Feb;12(2):137-47.

(2) Pang PS, Cleland JG, Teerlink JR, Collins SP, Lindsell CJ, Sopko G, et al. A proposal to standardize dyspnoea measurement in clinical trials of acute heart failure syndromes: the need for a uniform approach. Eur Heart J 2008 Mar;29(6):816-24.

(3) Bausewein C, Farquhar M, Booth S, Gysels M, Higginson IJ. Measurement of breathlessness in advanced disease: a systematic review. Respir Med 2007 Mar;101(3):399-410.

(4) Bausewein C, Booth S, Higginson IJ. Measurement of dyspnoea in the clinical rather than the research setting. Curr Opin Support Palliat Care 2008 Jun;2(2):95-9.

(5) Dorman S, Byrne A, Edwards A. Which measurement scales should we use to measure breathlessness in palliative care? A systematic review. Palliat Med 2007 Apr;21(3):177-91.

(6) Dorman S, Jolley C, Abernethy A, Currow D, Johnson M, Farquhar M, et al. Researching breathlessness in palliative care: consensus statement of the National Cancer Research Institute Palliative Care Breathlessness Subgroup. Palliat Med 2009 Apr;23(3):213-27.

(7) Oxberry SG, Bland JM, Clark AL, Cleland JG, Johnson MJ. Minimally Clinically Important Difference (MCID) in chronic breathlessness: Every little helps. American Heart Journal 2012;164(2):229-35.

(8) de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 2006;4:54.

(9) Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002 Apr;77(4):371-83.

(10) Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989 Dec;10(4):407-15.

(11) Johnson MJ, McDonagh TA, Harkness A, McKay SE, Dargie HJ. Morphine for the relief of breathlessness in patients with chronic heart failure-a pilot study. Eur J Heart Fail 2002 Dec;4(6):753-6.

(12) Oxberry SG, Torgerson DJ, Bland JM, Clark AL, Cleland JG, Johnson MJ. Short-term opioids for breathlessness in stable chronic heart failure: a randomized controlled trial. Eur J Heart Fail 2011 Sep;13(9):1006-12.

(13) Abernethy AP, Currow DC, Frith P, Fazekas BS, McHugh A, Bui C. Randomised, double blind, placebo controlled crossover trial of sustained release morphine for the management of refractory dyspnoea. BMJ 2003 Sep 6;327(7414):523-8.

(14) Currow DC, McDonald C, Oaten S, Kenny B, Allcroft P, Frith P, et al. Once-daily opioids for chronic dyspnea: a dose increment and pharmacovigilance study. J Pain Symptom Manage 2011 Sep;42(3):388-99.

(15) Powers J, Bennett SJ. Measurement of dyspnea in patients treated with mechanical ventilation. Am J Crit Care 1999 Jul;8(4):254-61.

(16) Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011 Jun;41(6):1073-93.

(17) Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum; 1988.

(18) Ries AL. Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. COPD 2005 Mar;2(1):105-10.

(19) de Torres JP, Pinto-Plata V, Ingenito E, Bagley P, Gray A, Berger R, et al. Power of outcome measurements to detect clinically significant changes in pulmonary rehabilitation of patients with COPD. Chest 2002 Apr;121(4):1092-8.

(20) Foglio K, Bianchi L, Bruletti G, Battista L, Pagani M, Ambrosino N. Long-term effectiveness of pulmonary rehabilitation in patients with chronic airway obstruction. Eur Respir J 1999 Jan;13(1):125-32.

(21) Reardon J, Awad E, Normandin E, Vale F, Clark B, ZuWallack RL. The effect of comprehensive outpatient pulmonary rehabilitation on dyspnea. Chest 1994 Apr;105(4):1046-52.

(22) Alvisi V, Mirkovic T, Nesme P, Guerin C, Milic-Emili J. Acute effects of hyperoxia on dyspnea in hypoxemia patients with chronic airway obstruction at rest. Chest 2003 Apr;123(4):1038-46.

(23) Martinez FJ, de Oca MM, Whyte RI, Stetz J, Gay SE, Celli BR. Lung-volume reduction improves dyspnea, dynamic hyperinflation, and respiratory muscle function. Am J Respir Crit care Med 1997 Jun;155(6):1984-90.

(24) Karras DJ, Sammon ME, Terregino CA, Lopez BL, Griswold SK, Arnold GK. Clinically meaningful changes in quantitative measures of asthma severity. Acad Emerg Med 2000 Apr;7(4):327-34.

(25) Ander DS, Aisiku IP, Ratcliff JJ, Todd KH, Gotsch K. Measuring the dyspnea of decompensated heart failure with a visual analog scale: how much improvement is meaningful? Congest Heart Fail 2004 Jul;10(4):188-91.

(26) Gysels M, Higginson IJ. Access to services for patients with chronic obstructive pulmonary disease: the invisibility of breathlessness. J Pain Symptom Manage 2008 Nov;36(5):451-60.

(27) Gysels M, Higginson IJ. The experience of breathlessness: the social course of chronic obstructive pulmonary disease
296. J Pain Symptom Manage 2010 Mar;39(3):555-63.

(28) Currow DC, Smith J, Davidson PM, Newton PJ, Agar MR, Abernethy AP. Do the trajectories of dyspnea differ in prevalence and intensity by diagnosis at the end of life? A consecutive cohort study. J Pain Symptom Manage 2010 Apr;39(4):680-90.

(29) Seow H, Barbera L, Sutradhar R, Howell D, Dudgeon D, Atzema C, et al. Trajectory of performance status and symptom scores for patients with cancer during the last six months of life
10. J Clin Oncol 2011 Mar 20;29(9):1151-8.

(30) Improving research methodology in breathlessness: a meeting convened by the MRC clinical trials unit and the Cicely Saunders Foundation. Palliat Med 2006 Apr;20(3):219-20.

Figure 1. Regression line of effect size (change in VAS score divided by standard deviation of baseline scores) and patient rated change in breathlessness score (VAS) from the contributing studies (Table 2)

Table 1. Baseline data from contributing studies in order to calculate a distribution-based estimate of minimal clinically important difference in dyspnoea intensity scores for people with chronic refractory dyspnoea.

Variable		Abernethy (13)		Currow (14)		Johnson (11)		Oxberry (12) Morphine		Oxberry (12) Oxycodone		Combined	
Number		48	%	85	%	10	%	35	%	35	%	213	%
Age	Mean	73.1		74.8		66.2		70.2		70.2		73.1	
	SD	9.5		8.9		11.6		11.1		11.1		9.5	
	Range	41-89		51-88		45-83		41-89		41-89		41- 89	
Gender	M	35	75	53	62	10	100	30	86	30	86	158	74
	F	13	25	32	38	0		5	14	5	14	55	26
Disease	HF	0	0	3	3.5	10	100	35	100	35	100	83	39
	COPD	42	89	42	49	0	0	0	0	0	0	84	39
	MND	1	2	0	0	0	0	0	0	0	0	1	<1
	Cancer	3	6	24	28	0	0	0	0	0	0	27	13
	RLD	2	4	16	19	0	0	0	0	0	0	18	8
Intensity of breathlessness	Mean	42		48		35		41		48		45	
	SD	23		22		25		20		22		22	
	Range	23- 86		1 - 85		1- 77		10- 90		10-100		1 - 100	
Karnofsky Performance Scale	mean	63.6		63.3		67.6		69.1		69.4		65.5	
	SD	16.4		9.5		3.9		6.1		6.1		10.7	
	range	20-85		40-80		62-70		60-80		60-80		20-85	
VAS/NRS scale		100 mm VAS; 0 mm = ‘‘no breathlessness’’; 100 mm = ‘‘worst imaginable breathlessness.’’						0-10 NRS; average breathlessness over last 24 hours. Verbal anchors as for Currow and Abernethy					

SD standard deviation; HF heart failure; COPD Chronic obstructive pulmonary disease; MND motor neurone disease; RLD restrictive lung disease

Table 2. Contributing studies

study	d4 minus d1	Baseline breathlessness intensity	SD difference	SD baseline	effect
Abernethy	−2.5	42	19	23.2	−0.11
Currow	−7.0	48`	18.0	21.8	−0.32
Johnson	−12.0	35	27.0	24.6	−0.48
Oxberry (morphine)	−4.0	41	25.1	19.7	−0.20
Oxberry (oxycodone)	−12.9	48	21.9	21.7	−0.60