1. Title: How outcomes are defined in clinical trials of mechanically ventilated adults and children

2. Bronagh Blackwood¹, Mike Clarke², Danny F McAuley¹, ³, Peter J McGuigan³, John C Marshall⁴, Louise Rose⁴, ⁵

3. ¹ Centre for Infection and Immunity, School of Medicine, Dentistry and Biomedical Sciences, Queen’s University Belfast, Northern Ireland
² Centre for Public Health and Medical Research Council All Ireland Hub for Trials Methodology Research, School of Medicine, Dentistry and Biomedical Sciences, Queen’s University Belfast, Northern Ireland
³ Regional Intensive Care Unit, Royal Hospitals, Belfast Trust, Northern Ireland
⁴ Li Ka Shing Knowledge Institute, St. Michael’s Hospital, University of Toronto, Canada
⁵ Lawrence S. Bloomberg Faculty of Nursing, University of Toronto; Provincial Centre of Weaning Excellence, Toronto East General Hospital; Mount Sinai Hospital, Toronto, Canada

4. Corresponding author: Bronagh Blackwood
Address: Centre for Infection and Immunity, Health Sciences Building, 97 Lisburn Road, BT9 7AE, Northern Ireland
Email: b.blackwood@qub.ac.uk; Tel: +44 28 9097 2616; Fax: +44 28 9097 2671

5. Author contributions: BB, LR, DMcA, and MC have been involved in the conception and design of the study; BB, PMcG, and LR have been involved in data acquisition and data extraction; BB, LR, DMcA, JM and MC have been involved in analysis and interpretation of the data; all authors have contributed to writing the article prior to submission.

6. Running title: How outcomes are defined in ventilation trials

7. No sources of support

8. Descriptor number 4.4 (Clinical Trials in Critical Care Medicine)

9. Body of the manuscript, 2258 words; abstract, 180 words

10. Scientific Knowledge on the Subject: International standards for trial registration call for sufficient detail on outcomes measured in clinical trials. Analyses of clinical trial outcomes documented in intervention reviews and trial databases indicate a lack of sufficient detail to aid comparison and synthesis of findings. What this Study adds to the Field: Ventilation outcomes are not well defined and outcome time point measures are not consistent among trials. The outcomes reported generally fell into a number of domains that suggests the possibility of generating a core outcome set.
**Abstract**

Systematic reviews have considerable potential to provide evidence-based data to aid clinical decision-making. However there is growing recognition that trials involving mechanical ventilation lack consistency in the definition and measurement of ventilation outcomes, creating difficulties in combining data for meta-analyses. To address the inconsistency in outcome definitions, international standards for trial registration and clinical trial protocols published recommendations, effectively setting the ‘gold standard’ for reporting trial outcomes. In this Critical Care Perspective, we review the problems resulting from inconsistent outcome definitions and inconsistent reporting of outcomes (outcome sets). We present data highlighting the variability of the most commonly reported ventilation outcome definitions. Ventilation outcomes reported in trials over the last six years typically fall into four domains: measures of ventilator dependence; adverse outcomes; mortality; and resource utilization. We highlight the need first, for agreement on outcome definitions and second, for a minimum core outcome set for trials involving mechanical ventilation. A minimum core outcome set would not restrict trialists from measuring additional outcomes, but would overcome problems of variability in outcome selection, measurement and reporting thereby enhancing comparisons across trials.

**Keywords:** core outcome measures; critical care trials; systematic reviews
**Background**

Many interventions applied in the critically ill influence the duration of mechanical ventilation. Measures of ventilator dependence, such as duration of ventilation, are frequently used as primary and secondary outcomes. However, variations in outcome definitions lead to differences in estimates of treatment effects and in a systematic review this ‘artefactual difference’ may dilute the real effect (1). As such, it is highly desirable that standardized outcome definitions are used in trials examining interventions likely to affect duration of mechanical ventilation. This will enable treatment effects to be compared in an unbiased, reliable and robust manner.

Consistency in the measurement and reporting of trial outcomes is lacking. Williamson and colleagues noted that the most accessed and cited Cochrane reviews in 2009 all described inconsistencies in trial outcomes (2). Two recent systematic reviews of the effectiveness of protocolized weaning highlighted inconsistencies in measurement time points for ventilation outcomes (3, 4). Current international standards for trial registration (CONSORT, WHO registry) and the recent SPIRIT 2013 recommendations for writing clinical trial protocols stress that investigators should report the specific measurement variable and time frame for each outcome measure when registering trials (5-7). However, analysis of a cohort of records from the clinicaltrials.gov database indicated that 36% of trials registered a domain only and lacked definition of the specific measure, metric used or method of aggregation of results. (8). Furthermore, outcomes included in trial protocols are not always reported in trial publications (9, 10).
In addition to the need for consistency in outcome definitions, Williamson and colleagues called for the development and use of core outcome sets (2), defined as an “agreed, standardized collection of outcomes measured and reported in all trials for a specific clinical area” (11). Agreement on a core outcome set should avoid problems associated with selective reporting of outcomes. The COMET (Core Outcome Measures in Effectiveness Trials) database has registered more than 260 references of planned work in developing core outcome sets and among them there are three studies exploring ventilation outcomes; long-term outcomes in acute respiratory failure; and rehabilitation following critical illness (12).

Little is known about how ventilation outcomes are defined and measured. To investigate this we sought to: first, identify trials involving mechanically ventilated adults and children that reported ventilation outcomes and describe how these outcomes were reported; and second, explore how trials specifically evaluating ventilation interventions reported ventilation outcomes and whether these might reveal a core outcome set.

**Review process**

We included trials published in the top ranked journals between January 2007 and December 2012 in general medicine (New England Journal of Medicine; Lancet; Journal of American Medical Association; Pediatrics) and critical care journals (American Journal of Respiratory and Critical Care Medicine; Critical Care Medicine; Intensive Care Medicine; Pediatric Critical Care Medicine). We specifically chose top ranked journals because of the
likelihood of locating high quality trial reports. Rankings were based on the impact factors in the ISI Web of Knowledge, Journal Citation Reports 2011, Thomson Reuters. Our review only included randomized controlled trials involving adults or children receiving invasive mechanical ventilation and measuring outcomes pertaining to the duration of ventilation and its discontinuation. Three authors (BB, LR, PMcG) divided the journals and independently searched, screened and extracted data. Titles and abstracts of papers in the journals were reviewed: full texts including supplementary material, published protocols or protocol registrations of all potentially relevant trials were retrieved. Details of the trials’ aims, primary and secondary outcome measures and their definitions were extracted onto pre-piloted data extraction forms (see online data supplement, appendix 1). These forms were reviewed by the three authors to confirm inclusion: a fourth author (DMcA) acted as arbitrator.

**Results**

We included 66 reports of randomized trials (13-78) and associated documentation (59 trial registrations; 34 electronic supplementary materials; 13 published protocols) (Figure 1). Interventions addressed management of ventilation; sedation; physiotherapy; nutrition; renal/fluid management; medications and infection prevention. The 66 trials reported 30 different primary outcomes. Primary outcomes reported by more than 1 trial and secondary outcomes reported by more than 10 trials are shown in Figure 2.

Nine ventilation outcomes were reported across included trials reflecting measures of ventilator dependence or occurrence of events (typically adverse outcomes) (Table 1).
Duration of ventilation was the most commonly reported outcome yet only 12 trials (25%) provided a definition (20, 24, 30, 32, 35, 42-44, 61, 64, 71, 73) with substantial variation in time point measures. Twenty-five trials reported ventilator-free days as an outcome, 9 (36%), (18, 19, 23, 51, 53, 62, 63, 72, 73) provided no definition and variable start and end points were reported in the 16 (64%) providing a definition (24, 26, 31, 45, 50, 52, 54, 57, 58, 65, 66, 68-70, 72, 76) there were variable start and end points (Table 2). Three trials reporting weaning duration as a secondary outcome (24, 62, 76) and only one provided a definition (24). When reintubation was reported as a trial outcome (4 trials), follow-up was measured at 24-hours (73), 48-hours (49, 61) or 7-days (32). When reintubation was recorded as an adverse event (4 trials) (31, 40, 55, 71), the follow-up period within which this was measured was not provided.

Outcomes reported in trials specifically evaluating a ventilation intervention

Twelve trials tested a ventilation intervention including sedation and ventilation weaning methods (20, 30, 31, 43, 44, 49, 68); ventilator modes (27, 78); automated systems to facilitate ventilator weaning (61, 64); and early non-invasive ventilation following invasive ventilation (32). Primary and secondary outcomes reported in these trials are presented in Figure 3. Outcome measures reflected four domains; measures of ventilator dependence (e.g. duration of ventilation, duration of weaning, separation potential); adverse outcomes (e.g. VAP, reintubation, self extubation); mortality and survival outcomes (e.g. survival, ICU mortality, hospital mortality); and resource utilization (e.g. ICU LOS, hospital LOS, clinical workload).
There was considerable variation in measuring primary outcomes. Ventilation trials reported either duration of mechanical ventilation or ventilator-free days, but not both. Start and endpoints are shown in Table 3. In secondary outcomes, mortality was reported for different follow up periods that included: ICU (20, 49, 68, 78); hospital (20, 68, 78); 28-day (31, 64); 30-day (30); 90-day (31); and prior to ventilator separation potential and extubation (61). One trial did not define the follow up period (27) and one trial did not measure mortality (43). Two trials measured survival, at 1-year (31) and at an undefined time point (32).

Discussion

We found substantial variation in the outcome sets used. Outcome definitions differed between trials, often measuring different time points and different follow up periods. Furthermore, a large number of trials lacked detail in their outcome definitions.

It is important to highlight the effect related to the competing risk of death in using duration of ventilation as an outcome measure. Various statistical methods have tried to address this issue. Egleston and colleagues (79) point out that by using a basic approach of examining outcomes in survivors only it is possible that a harmful intervention will increase mortality in a vulnerable population. These remaining healthier survivors may have a reduced duration of ventilation giving the impression that the intervention has a beneficial effect. (79) Therefore it is important to consider the duration of ventilation as an outcome in the context of mortality. Measuring ventilator-free days (i.e. a composite outcome of mortality and duration of ventilation) is one method to address competing risks. (80) Our work has demonstrated that ventilator-free days are often poorly defined or not reported.
Cause specific hazard models that fail to take into account the competing event (death) results in the patient being censored from the analysis and may falsely make the intervention appear beneficial. (81) When a sub-distribution hazard model is employed patients are not censored despite the occurrence of the competing event. When events are mutually exclusive such as death or unassisted breathing and discharge home a parametric mixture survival model may be the most appropriate method to determine the true effect of an intervention. (82)

The issue of competing risk is complex and does not just apply to mortality. An intervention intended to reduce the duration of mechanical ventilation may potentially lead to complications making it unclear whether the treatment causes the complication or results in more patients being alive to develop the complication. Statistical approaches such as cause specific hazard, cumulative incidence function and event-free survival are used to detect the true effect of an intervention. However, their ability to do so varies depending on whether the competing risk is affected in the same or opposite manner to the primary event being studied. (83) Regardless of these issues our finding of variability in mechanical ventilation outcome measures is important.

The outcome ‘duration of mechanical ventilation’ was reported by 73% of all trials and was measured from either intubation or initiation of ventilation (which may or may not occur simultaneously). It is generally interpreted as intubation and initiation of mechanical ventilation occurring on ICU admission. However, initiation may occur in the Operating Room or Emergency Department and the start time point may be difficult to obtain or
inconsistently recorded. Furthermore, some trials reported *randomization* as the start point for ‘duration of mechanical ventilation’. Firstly, the criteria for randomization vary between trials. Secondly, enrollment in some trials, for example, investigating weaning methods, are dependent upon physician assessment and are prone to bias. Finally, randomization may be delayed by the process of gaining consent, with consent windows up to 24 hours.

We found the end time point for measuring duration of mechanical ventilation was reported as either ‘free from mechanical ventilation’ or extubation. Free from mechanical ventilation was defined as either freedom from invasive ventilation or both invasive and non-invasive ventilation. Timing of extubation may be influenced by organizational issues such as workload and staff availability (84, 85). These organizational issues may vary widely between institutions and across trials. This wide variability in outcome definitions regarding start and end points of mechanical ventilation will not be problematic within a trial provided both arms are equally affected. However, it may lead to systemic variations when comparing trials, highlighting the need for agreed outcome definitions.

In the 12 trials specifically evaluating a ventilation intervention, there was also time point variation in secondary outcomes. When reporting mortality, it is reasonable to choose a longer duration of follow up when delayed effects are expected, such as in ARDS or severe sepsis (86). However, consensus on duration of follow up is required if trials are to be compared. Length of stay (ICU and hospital) was frequently measured as a secondary outcome. Length of stay is an important healthcare resource utilization outcome. However, it has limitations, particular in comparing trials across countries with different healthcare
funding models and resource availability. Furthermore, contextual differences in end of life care practices may also affect duration of ventilation and length of stay.

Although this paper is the first to provide data showing substantial variation in outcome definitions in ventilation trials, it is not the first to call for improvements in standardizing outcomes in critical care trials. A workshop convened by the National Heart, Lung and Blood Institute (87) recommended standardization in describing and collecting endpoints to facilitate meta-analyses of acute lung injury trials; and a Society of Critical Care Medicine stakeholder conference (88) highlighted the necessity of gaining consensus on a standard set of outcome measures for long term outcomes following ICU discharge.

Our analysis indicates that outcomes reported in trials of ventilation interventions typically measure outcomes in a number of domains: measures of ventilator dependence; adverse outcomes; mortality and survival outcomes; and resource utilization. For ICU patients, being free from ventilation and survival are clearly important outcomes, but length of stay may not be. Longer term outcomes such as cognitive and physical function and quality of life are often under reported. (87, 88) However, the effects of critical illness impacts on patients long after hospital discharge and these longer term outcomes are increasingly being recognized as being important by investigators. (89-91) The optimal duration of long term follow-up remains to be determined.

The common domains that are addressed give rise to the possibility of obtaining agreement on outcome definitions and a core outcome set. A core outcome set would not restrict trialists from measuring additional outcomes, but would overcome problems of variability in
outcome selection, measurement and reporting thereby enhancing valid comparisons across trials. To address this, we are undertaking a study that will use the Delphi technique to achieve international consensus on core outcome definitions and a core outcome set for use in trials involving mechanical ventilation (http://www.comet-initiative.org/studies/details/292?result=true). We will liaise closely with the International Forum for Acute Care Trialists (InFACT) and the Delphi panel will draw upon relevant stakeholders including patient groups, professional societies, clinical trial groups, and industry.

Limitations

Our search for trials was restricted to a short, but recent, time period and a small number of journals. It does not provide a full comprehensive overview of ventilation outcomes in trials across a longer time period and a wider range of journals; however, we are confident that we have presented sufficient data to demonstrate significant variability in outcome reporting in recent trials accepted for publication in high impact journals.

Conclusion

We show substantial variation in the choice of outcome measures and their definition in randomized trials evaluating interventions likely to influence the duration of ventilation. We anticipate the recent SPIRIT 2013 statement (7) outlining guidance for reporting clinical trial protocols will help investigators provide clear definitions enabling more appropriate comparisons. Expert consensus on, and implementation of, standardized outcome
definitions and core outcome sets is fundamental to reducing bias when comparing effects across trials.
Table 1  Summary of ventilation outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No of trials</th>
<th>Reported as primary</th>
<th>Reported as adverse event</th>
<th>Definition provided N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures of ventilation dependence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of ventilation</td>
<td>48</td>
<td>9</td>
<td>-</td>
<td>12 (25%)</td>
</tr>
<tr>
<td>Ventilator-free days</td>
<td>25</td>
<td>4</td>
<td>-</td>
<td>16 (64%)</td>
</tr>
<tr>
<td>Weaning time</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>1 (33%)</td>
</tr>
<tr>
<td>Time to separation potential</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Occurrence of events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reintubation</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>4 (50%)</td>
</tr>
<tr>
<td>Extubation failure</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2 (50%)</td>
</tr>
<tr>
<td>Use of post extubation</td>
<td>4</td>
<td>-</td>
<td>3</td>
<td>1 (25%)</td>
</tr>
<tr>
<td>NIV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weaning failure</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>Successful extubation</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>

Key: NIV, non-invasive ventilation
<table>
<thead>
<tr>
<th>Start point</th>
<th>Endpoint</th>
</tr>
</thead>
</table>

**Duration of mechanical ventilation**

Commencement of IMV (20, 30, 71)  
Intubation (43)  
Randomization (24, 35, 42, 44, 61, 64, 73)  
---

1\textsuperscript{st} extubation (24, 30, 44, 61, 64)  
Successful extubation (35, 61, 64)  
Successful extubation (UAB for 24(73), 48(43, 44) or 72-hours(24))  
1\textsuperscript{st} SBT (64)  
Successful SBT (61, 64)  
Successful extubation (24-hours) or successful SBT (UAB for 48-hours)(42)  
Ventilator free (71)  
Free from IMV or NIV (for 48-hours)(35)  
Extubation from IMV and NIV stop or reintubation for NIV group (32)  
Ventilation time within 28-days (64)  
Successful weaning (20, 61)  
---

**VFD**

Day 1 (50, 54)  
Intubation (50)  
Randomization (24, 30, 40, 46, 49, 51, 52, 81)  
Not specifically stated (26, 65, 69, 70, 76)  
---

Day 28 (24, 31, 45, 50, 52, 58, 65, 66, 68-70, 76)  
Day 60 (24)  
Day 90 (57)  
Day 28 and 90 (54)  
Not specifically stated (72)  
---

Key: IMV, invasive mechanical ventilation; NIV, noninvasive mechanical ventilation; SBT, spontaneous breathing trial; UAB, unassisted breathing. Some trials measured more than one endpoint.
### Table 3  Start and end point variability for ventilation outcomes in 12 ventilation trials

<table>
<thead>
<tr>
<th>Start point</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration of mechanical ventilation</strong></td>
<td></td>
</tr>
<tr>
<td>Commencement of IMV (20, 30)</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; extubation (30, 61, 64)</td>
</tr>
<tr>
<td>Intubation (43, 61)</td>
<td>Successful extubation (61)</td>
</tr>
<tr>
<td>Randomization (39, 76, 77)</td>
<td>Extubation or a trache mask for 48-hours (43, 44)</td>
</tr>
<tr>
<td></td>
<td>Extubation from IMV and NIV stop or reintubation for NIV group (32)</td>
</tr>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; SBT (32)</td>
</tr>
<tr>
<td></td>
<td>Successful SBT (32)</td>
</tr>
<tr>
<td></td>
<td>Successful weaning (20)</td>
</tr>
<tr>
<td><strong>VFD</strong></td>
<td></td>
</tr>
<tr>
<td>Intubation (68)</td>
<td>Day 28 (31, 68)</td>
</tr>
<tr>
<td>Randomization (31, 68)</td>
<td></td>
</tr>
</tbody>
</table>

Key: IMV, invasive mechanical ventilation; NIV, noninvasive mechanical ventilation; SBT, spontaneous breathing trial. Some trials measured more than one start and endpoint.
References


