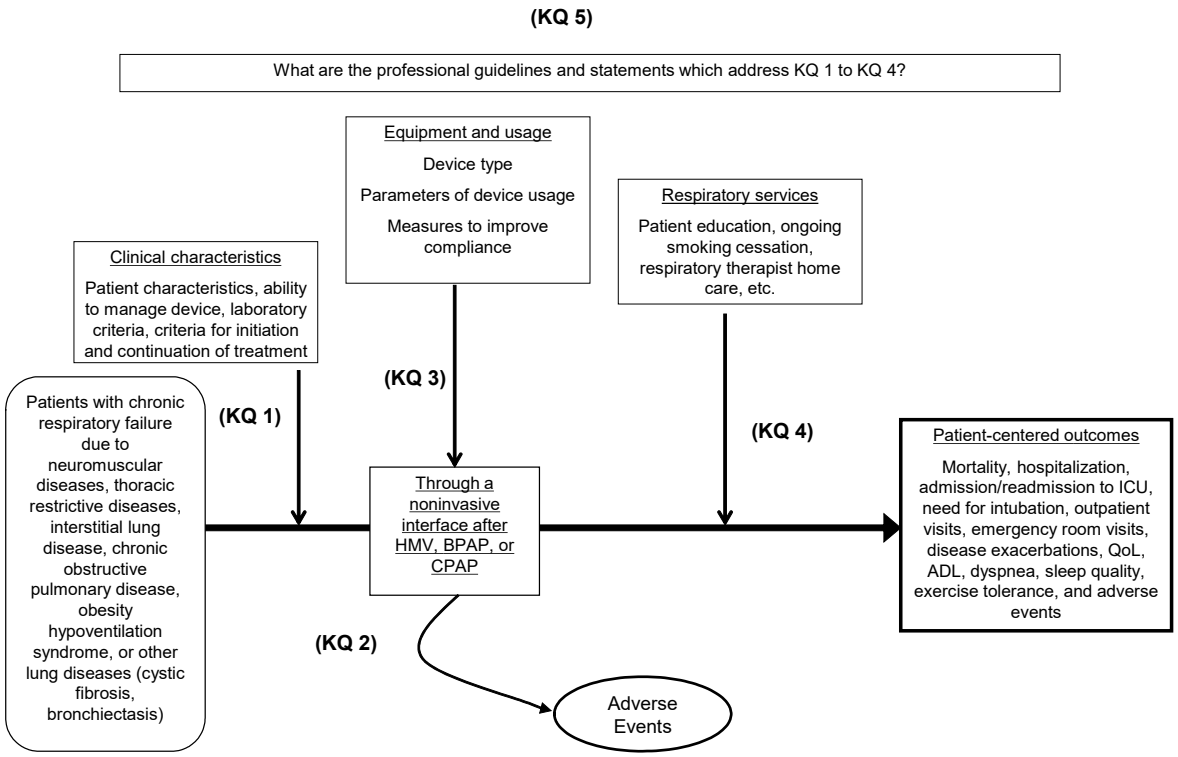
Methods

We developed an analytic framework to guide the process of the systematic review (Figure 1). We followed the established methodologies of systematic reviews as outlined in the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Comparative Effectiveness Reviews.[15] The reporting complies with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements.[98] The study protocol is registered in the international prospective register of systematic reviews (PROSPERO #: CRD42018085676) and published on the AHRQ Web site (https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/ta/topicrefinement/hmv-protocol.pdf).

Figure 1. Analytic framework



Literature Search Strategy

Search Strategy

We conducted a comprehensive literature search of eight databases, including National Guideline Clearinghouse, Embase, Epub Ahead of Print, In-Process & Other Non-Indexed Citations, MEDLINE Daily, MEDLINE, Cochrane Central Registrar of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus from January 1, 1995 to June 26, 2018. We also searched Food and Drug Administration (FDA) Establishment Registration & Device Listing, ClinicalTrials.gov, Health Canada, Medicines and Healthcare Products Regulatory Agency (MHRA), AHRQ's Horizon Scanning System, conference proceedings, patient advocate group websites, and medical society websites. Relevant clinical guidelines, systematic reviews, and meta-analysis, as well as reference mining of relevant publications, were used to identify additional literature. An experienced librarian, with the help of the study investigators, developed the search strategy (Appendix B) and conducted the search. An independent information specialist peer reviewed the search strategy.

Inclusion and Exclusion Criteria

The eligible studies had to meet all the following criteria: 1) Adults 18 years and older with chronic respiratory failure due to neuromuscular diseases, thoracic restrictive diseases, chronic obstructive pulmonary diseases (COPD), obesity hypoventilation syndrome , or other lung diseases (cystic fibrosis, bronchiectasis); 2) received noninvasive positive pressure ventilation supplied by a Home Mechanical Ventilator (HMV), Bi-level Positive Airway Pressure device (BPAP), or Continuous Positive Airway Pressure device (CPAP) through noninvasive interface; 3) received at least 1 month of treatment at home or assisted living; 4) compared with usual care; different type of noninvasive mechanical ventilation, different modes of same equipment, or other noninvasive ventilation; 5) reported patient-centered outcomes, and 6) published after 1995 and in English only. We included randomized controlled trials (RCTs), nonrandomized comparative studies (prospective and retrospective), and clinical guidelines. We did not restrict study location, or sample size. The detailed inclusion and exclusion criteria can be found in Table 1.

Study Selection

Independent reviewers, working in pairs, screened the titles and abstracts of all citations using pre-specified inclusion and exclusion criteria. Studies included by either reviewer were retrieved for full-text screening. Independent reviewers, again working in pairs, screened the full-text version of eligible references. Discrepancies between the reviewers were resolved through discussions and consensus. If consensus could not be reached, a third reviewer resolved the difference.

Data Extraction

We developed a standardized data extraction form to extract study characteristics (author, study design, inclusion and exclusion criteria, patient characteristics, laboratory criteria, intervention, comparisons, outcomes, equipment parameters, respiratory services, and related items for assessing study quality and applicability). The standardized form was pilot-tested by all

study team members using 10 randomly selected studies. We iteratively continued testing the form until no additional items or unresolved questions existed. After we finalized the form, reviewers worked independently to extract study details. A second reviewer reviewed data extraction and resolved conflicts.

Assessment of Risk of Bias of Individual Studies

We evaluated the risk of bias of the included study using predefined criteria. For RCTs, we used the Cochrane Collaboration's Risk of Bias tool to assess sequence generation; allocation concealment; participant, personnel, and outcome assessor blinding; attrition bias; incomplete outcome data; selective outcome reporting; and other sources of bias (e.g. conflict of interest, imbalance of baseline characteristics).[99] Each domain was rated as high, low, or unclear risk. For observational studies, we selected appropriate items from the Newcastle-Ottawa Scale, including representativeness of the patients, ascertainment of exposure and outcomes, adequacy of follow-up, and possible conflicts of interest.[100] Each item was rated as high, low, or unclear risk. Finally, we gave an overall risk of bias for each study with focus on sequence generation, allocation concealment, and other sources of bias for RCTs and representativeness and ascertainment of exposure and outcomes for observational studies.

Data Synthesis

We qualitatively summarized key features/characteristics (e.g. study populations, design, intervention, outcomes, device model, equipment parameters, and conclusions) of the included studies and presented them in evidence tables by each disease and device.

Table 2 lists rules we used to categorize HMV device, BPAP device, or CPAP device.

Table 2. Rules used to categorize HMV, BPAP, and CPAP

Device	Rules
HMV	1) The study reported the device model/manufacturer, and the device was classified as a life support ventilator by either the FDA or the manufacturer listed information, or 2) The study reported the device to be a life support device, or 3) The study reported the device was also able to be used interchangeably with invasive mechanical ventilation through a tracheostomy or endotracheal tube, or 4) The study reported the mode to be (or the device was capable of) continuous mandatory ventilation (CMV) in either a pressure controlled PC-CMV (AC-PC) or volume controlled VC- CMV (AC-VC) configuration.
BPAP	1) The study reported the device model/manufacturer, and the device was classified as a BPAP machine or respiratory assist device (RAD) by either the FDA or the manufacturer listed information, or 2) The study reported the device to be an exclusive BPAP machine. 3) Devices were categorized as BIPAP ST if the mode utilized intermittent mandatory ventilation IMV (back up rate) with pressure support (IPAP) PC-IMV. BIPAP S if breath delivery was continuous spontaneous ventilation CSV with pressure support (IPAP) PC-CSV.
CPAP	1) The study reported the device model/manufacturer, and the device was classified as a CPAP machine either by the FDA or the manufacturer listed information, or 2) The study reported the device to be a CPAP machine.

BPAP/BIPAP: bi-level positive airway pressure, CMV: continuous mandatory ventilation, CPAP: continuous positive airway pressure, CSV: continuous spontaneous ventilation, FDA: Food and Drug Administration, HMV: home mechanical ventilation, IMV: intermittent mandatory ventilation, IPAP: inspiratory positive airway pressure, PC-CMV: pressure controlled continuous mandatory ventilation, PC-CSV: pressure controlled continuous spontaneous ventilation, PC-IMV: pressure controlled intermittent mandatory ventilation, RAD: respiratory assist device, S: spontaneous mode, ST: spontaneous/timed mode, VC-CMV: volume controlled continuous mandatory ventilation

Adverse events were grouped into adverse events likely due to device use, including 1) mask, tubing (interface) related problems, 2) problems related to nasal route; and 3) pressure, airflow related problems. All adverse events that were likely not linked to device use were grouped as other adverse events (Table 3).

Table 3. Categories of adverse events

Type of adverse events	Example
Serious adverse events	Death, hospitalization, and need for intubation were reported as primary efficacy outcomes. Acute respiratory failure Any life-threatening event/illness Any disability or permanent damage Any required intervention to prevent impairment (such as pacemaker) Any congenital anomaly/birth defect
Non serious adverse events	Skin symptoms (e.g. facial rash, nasal ulceration) Eye symptoms (e.g. dry eyes, conjunctivitis) Nose/mouth symptoms (e.g. nasal stuffiness, rhinorrhea, nosebleed, mucosal dryness, oral air leak) Gastrointestinal symptoms (e.g. gastric distension, aerophagia) Device/mask intolerance (e.g. claustrophobia, discomfort, noncompliance) Other

We conducted meta-analyses to quantitatively combine study findings. All analyses were conducted based on the intention-to-treat principle for RCTs and the number of patients initially assigned to the intervention for observational studies. We calculated odds ratio (OR) and corresponding 95-percent confidence intervals for binary outcomes. For continuous outcomes, we extracted or calculated the difference between post intervention and baseline for each group for all observational studies and for RCTs (whenever possible). When the difference between post intervention and baseline was not presented in RCTs, we extracted post intervention data instead as baseline between groups was typically balanced. When studies used different measures for the same outcome (e.g. Epworth Sleepiness Scale and Pittsburgh Sleep Quality Index for sleep quality), we calculated standardized mean difference (SMD). When studies used the same outcome measure, we used the original scale. For count data (i.e. a patient may have more than one event, e.g. number of hospital admissions), we calculated rate ratio (ratio of the incidence rate of events within a given time between the intervention and the control). For adverse events, we calculated incidence rate by adverse events and type of device. The DerSimonian and Laird random effect method was used except when the number of studies included in the comparison was less than three. The fixed effect model based on the Mantel and Haenszel method was used in that case because of concern about instability of between study variance [101] We evaluated heterogeneity between studies using I^2 indicator. Subgroup analysis was only possible when BPAP was compared with no device. We conducted ad-hoc analyses (stable COPD vs. COPD with recent exacerbation). Per peer reviewers' suggestions, we added post-hoc subgroup analyses on different levels of hypercapnia (PaCO2) used as an initiation criterion for initiation of NIPPV. The cutoffs (PaCO2>=45 to 49 mmHg, PaCO2 >=50 to 51 mmHg, PaCO2 >=52 mmHg or greater) were selected to reflect those commonly reported by the clinical guidelines and to investigate a dose response of PaCO2 and clinical outcomes. The details of the post-hoc analyses are listed in Appendix I. We compared the effect sizes from the post-hoc subgroup analyses to the pooled effect sizes of all data to evaluate reporting bias. We

were unable to use statistical methods (e.g. funnel plots, Egger's regression test, etc.) to assess publication bias because the number of studies included in the analysis was small (n<20).[102] All statistical analyses were conducted using Stata/SE version 15.1 (StataCorp LLC, College Station, TX).

Grading the Strength of Evidence

We graded the strength of the body of evidence (SOE) as per the EPC methods guide on assessing the strength of evidence.[15] We designated four outcomes to be most critical to patients and conducted SOE rating for these major outcomes (mortality, need for intubation, quality of life and all-cause hospital admissions). We produced summary of evidence tables for the major outcomes that include data source, effect size, SOE rating; and rationale for judgments made on each domain of evidence rating. Other outcomes were either encompassed in these constructs (e.g., symptoms or functional test as a part of quality of life) or were not well ascertained (e.g., cause specific hospitalization). These other outcomes are summarized in tables showing the data source, study type and the effect size.

RCTs start as high SOE and observational studies start as low SOE. We considered the following SOE domains: the methodological limitations of the studies (i.e., risk of bias); precision (based on the size of the body of evidence, number of events, and confidence intervals); directness of the evidence to the KQs (focusing on whether the outcomes were important to patients vs surrogates); consistency of results (based on qualitative and statistical approaches to evaluate for heterogeneity); and the likelihood of reporting and publication bias. When confidence intervals were very wide showing substantial benefit and harm and the number of patients was small; we rated SOE as insufficient due to severe imprecision.

Based on this assessment and the initial study design, we assigned a SOE rating as high, moderate, low, or 'insufficient evidence to estimate an effect.'

Assessing Applicability

We followed the procedures outlined in the EPC Methods Guide for Comparative Effectiveness Reviews to assess the applicability of the findings within and across studies.[15] We focused on whether the populations and interventions in existing studies are representative of current practice. For studies to have good applicability, the devices used in research need to have similar parameters and characteristics to those available in the US at the present time. The characteristics of individuals enrolled in the studies should be similar to typical patients with the targeted conditions described in the PICOTS in terms of disease severity and comorbidities and threshold for being prescribed HMV, BIPAP and CPAP. Patients in the studies should not have excessive home support than what is feasible in real life; otherwise, applicability was judged as limited.

This congruence between research and practice as it relates to applicability was evaluated qualitatively and reported narratively. Research gaps in the topic area were reported in the Discussion.

Peer Review and Public Commentary

A draft report was posted for peer review and public comments between September 10 and October 1, 2018. We revised and finalized the draft report in response to comments. However,

the findings and conclusions are those of the authors, who are responsible for the contents of the report.