

# BMJ Open Study protocol to develop a patient-reported outcome measuring disability associated with unilateral vocal fold paralysis: a mixed-methods approach with the CoPE collaborative

Sara Fernandes-Taylor,<sup>1</sup> Cara Damico Smith,<sup>1</sup> Natalia Arroyo,<sup>1</sup> Kemberlee Bonnet,<sup>2</sup> David Schlundt,<sup>2</sup> Margarete Wichmann,<sup>3</sup> Irene Feurer,<sup>4</sup> David O Francis<sup>5</sup>

**To cite:** Fernandes-Taylor S, Damico Smith C, Arroyo N, *et al.* Study protocol to develop a patient-reported outcome measuring disability associated with unilateral vocal fold paralysis: a mixed-methods approach with the CoPE collaborative. *BMJ Open* 2019;**9**:e030151. doi:10.1136/bmjopen-2019-030151

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-030151>).

Received 07 March 2019

Revised 13 September 2019

Accepted 18 September 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Sara Fernandes-Taylor; [fernandes-taylor@surgery.wisc.edu](mailto:fernandes-taylor@surgery.wisc.edu)

## ABSTRACT

**Introduction** Patient-reported outcome (PRO) measures are increasingly developed with multisite, representative patient populations so that they can serve as a primary endpoint in clinical trials and longitudinal studies. Creating multisite infrastructure during PRO measure development can facilitate future comparative effectiveness trials. We describe our protocol to simultaneously develop a PRO measure and create a collaborative of tertiary care centres to address the needs of patients with unilateral vocal fold paralysis (UVFP). We describe the stakeholder engagement, information technology and regulatory foundations for PRO measure development and how the process enables plans for multisite trials comparing treatments for this largely iatrogenic condition.

**Methods and analysis** The study has three phases: systematic review, measure development and measure validation. Systematic reviews and qualitative interviews (n=75) will inform the development of a conceptual framework. Qualitative interviews with patients with UVFP will characterise the lived experience of the condition. Candidate PRO measure items will be derived *verbatim* from patient interviews and refined using cognitive interviews and expert input. The PRO measure will be administered to a large, multisite cohort of adult patients with UVFP via the CoPE (vocal Cord Paralysis Experience) Collaborative. We will establish CoPE to facilitate measure development and to create preliminary infrastructure for future trials, including online data capture, stakeholder engagement, and the identification of barriers and facilitators to participation. Classical test theory psychometrics and grounded theory characterise our approach, and validation includes assessment of latent structure, reliability and validity.

**Ethics and dissemination** Our study is approved by the University of Wisconsin Health Sciences Institutional Review Board. Findings from this project will be published in open-access journals and presented at international conferences. Subsequent use of the PRO measure will include comparative effectiveness trials of treatments for UVFP at CoPE Collaborative sites.

## Strengths and limitations of this study

- We will develop a patient-reported outcome (PRO) measuring unilateral vocal fold paralysis-related disability using a patient-centred, iterative, mixed-methods approach that is consistent with published guidelines for PRO measure development.
- We will ensure a representative patient population is engaged in the development of our PRO measure by leveraging the vocal CoPE (Cord Paralysis Experience) collaborative, which includes 50 tertiary care institutions that treat unilateral vocal fold paralysis in a specialty otolaryngology setting.
- Reliability and validity will be rigorously tested to ensure that the PRO measure is properly developed and appropriately measures constructs related to unilateral vocal fold paralysis.
- Engagement of centres nationwide will be time intensive and will require meticulous study co-ordination and survey administration.
- Submission to the FDA's Clinical Outcome Assessment Program may take several years as their guidelines are currently under revision.

## BACKGROUND

Patient-reported outcome (PRO) measures are increasingly developed with multisite, representative patient populations with the express purpose of serving as a primary endpoint in clinical trials and longitudinal observational studies.<sup>1</sup> The most recent FDA guidance emphasises the importance of representative patient involvement in instrument development and validation to support the use of a PRO measure as a clinical trial endpoint.<sup>2</sup> Creating a multisite infrastructure has dual purpose: (1) it ensures representative patient involvement and (2) allows the comparison of treatment alternatives during PRO measure development. This approach has the potential to streamline the process by

which evidence is generated to improve patient-centred outcomes. We propose one such model for simultaneously developing a PRO measure while creating a collaborative of tertiary care centres to address the needs of patients with unilateral vocal fold paralysis (UVFP).

UVFP is an increasingly common and debilitating neurological condition caused by injury to one recurrent laryngeal nerve. Increased prevalence of head, neck, spine and cardiothoracic surgeries has increased the population at risk for UVFP; these procedures account for 50% of UVFP cases.<sup>3–6</sup> Specifically, UVFP complicates up to 15%<sup>7–9</sup> and 11%<sup>10–11</sup> of thyroidectomies and anterior spine procedures, respectively. In recent decades, these procedures have increased threefold and eightfold,<sup>12–14</sup> with a corresponding rise in UVFP incidence.<sup>6–15–17</sup> UVFP has debilitating quality of life and health sequelae that include disordered communication, swallowing and breathing with substantial associated work productivity losses. For comparison, studies of *non-UVFP-attributable* voice disorders report health-related quality of life (HRQoL) implications and work productivity losses comparable with patients with asthma, acute coronary syndrome and depression, with 1:10 affected individuals filing short-term disability claims.<sup>18</sup> Patients with *UVFP-attributable* voice disorders have substantially worse HRQoL<sup>19–20</sup> and even greater productivity losses. From a health perspective, 60% of patients suffer dysphagia<sup>21</sup> (23% with aspiration)<sup>22</sup> and 75% new-onset dyspnoea.<sup>21</sup> Treatments for UVFP vary widely, most likely because the degree of spontaneous recovery and timing of intervention vary with the severity of neurological injury. Patients benefit significantly from some interventions, but the optimal type(s) and timing(s) remain undetermined because of insufficient high-quality comparative evidence.

With the exception of two randomised trials,<sup>23–24</sup> studies are limited to case series and single-centre observational studies using only voice outcome measures. Few studies consider treatment effects on swallowing or breathing dysfunction.<sup>25–26</sup> In 1999, Gliklich *et al* created the Voice Outcome Survey (VOS) and validated it among 56 patients with UVFP.<sup>27</sup> Subsequent evaluations of VOS's psychometric characteristics raised questions about its validity and clinical applicability.<sup>28</sup> Moreover, it was intended to measure UVFP handicap related to *voice only* and overlooked other laryngeal dysfunctions associated with UVFP. A psychometrically reliable, validated, clinically applicable and generalisable measure of UVFP-attributable disability would enable clinicians and researchers to assess affected patients' disability in all laryngeal domains, determine which treatment algorithms minimise disability while maximising HRQoL and allow interventions' effectiveness to be compared in clinical trials.

Although physiological measures are routinely used to evaluate clinical endpoints, these measures often do not correlate with patient-reported disability and may not fully reflect all components of important clinical outcomes. In everyday practice, clinicians primarily rely on patient

self-report to evaluate treatment effectiveness. Many clinicians use existing PRO measures to gauge disability. However, systematic review of available voice, swallowing and upper airway breathing-specific PRO measures show significant methodological limitations including lack of patient involvement in item development, lack of robust content and construct validity, and lack of clear interpretability and scaling characteristics.<sup>28–34</sup> The current consensus is that all PRO measures used as *primary endpoints* in clinical trials should be condition specific.<sup>35–38</sup> The development of disease-specific PRO measures that reflect these self-reports is essential to evaluating the full impact of treatment on patient function.

Our goal is to develop the first comprehensive PRO measure that can quantify the disability specific to patients with UVFP and perform as the primary endpoint in future trials comparing the effectiveness and durability of rehabilitative treatments including behavioural (speech therapy), and temporary (injectables) and permanent surgical treatments for UVFP. Validating the PRO measure in a large, nationally representative sample will ensure its generalisability to the US population and will provide data to establish a minimal clinically important difference for interpreting scores. Recruitment at multiple sites will facilitate the establishment of a multicentre consortium of 50 voice centres that treat UVFP, called the CoPE (vocal Cord Paralysis Experience) Collaborative, thus laying the groundwork for comparative effectiveness trials of treatments. We describe the stakeholder engagement, information technology, and regulatory foundations for PRO measure development and how the process enables plans for multisite trials comparing treatments for this largely iatrogenic condition.

## METHOD

Consistent with published guidelines for PRO measure development,<sup>39–43</sup> we will develop our UVFP PRO measure using an iterative, mixed-methods approach. Our quantitative approach draws primarily from classical test theory psychometrics, and our qualitative approach is characterised by an iterative inductive–deductive approach.<sup>44</sup>

## Patient and public involvement

Content of the PRO measure will derive *verbatim* from patient interviews describing their lived experience with UVFP and defining outcomes that matter to them. All patient-facing recruitment materials will be twice reviewed by a panel of Community Advisors on Research Design and Strategies (CARDS) maintained by University of Wisconsin–Madison in partnership with community centres. These partners are recruited 'from centre programs such as senior meals, women's groups, food pantries, and parenting programs. All of the CARDS complete an orientation with university staff to prepare them for effective meetings with researchers. The CARDS bring valuable perspectives from diverse racial, socioeconomic, and educational backgrounds'.<sup>45</sup>

In our study, patients will complete the PRO measure items retained after preliminary analysis. Results will be disseminated to academic audiences through publication in an open-access journal and to participants in a lay summary available on the study website.

## PHASE I: SYSTEMATIC REVIEW

We first performed systematic reviews using PRISMA criteria evaluating developmental characteristics of all available PRO measures related to voice,<sup>30</sup> swallowing,<sup>31</sup> globus pharyngeus and laryngopharyngeal reflux,<sup>29</sup> and upper airway-related dyspnoea.<sup>32–46</sup> We briefly report on these reviews here to explain their importance in informing our conceptual framework. We found wide variability in measurement properties of published PRO measures. Although several high-quality PRO measures exist in some domains (eg, swallowing), none had adequate measurement properties to serve as a primary endpoint in comparative effectiveness trials of UVFP treatments. Most PRO measures did not document content validity, longitudinal validity/responsiveness to change, construct validity and necessary scaling characteristics.<sup>29–32 46</sup> As a result, we undertook a round of qualitative interviews to develop a preliminary conceptual framework for patients' lived experience with UVFP. The qualitative interviews served as a platform for developing a content-valid item pool.

## PHASE II: MEASURE DEVELOPMENT

### Qualitative interviews

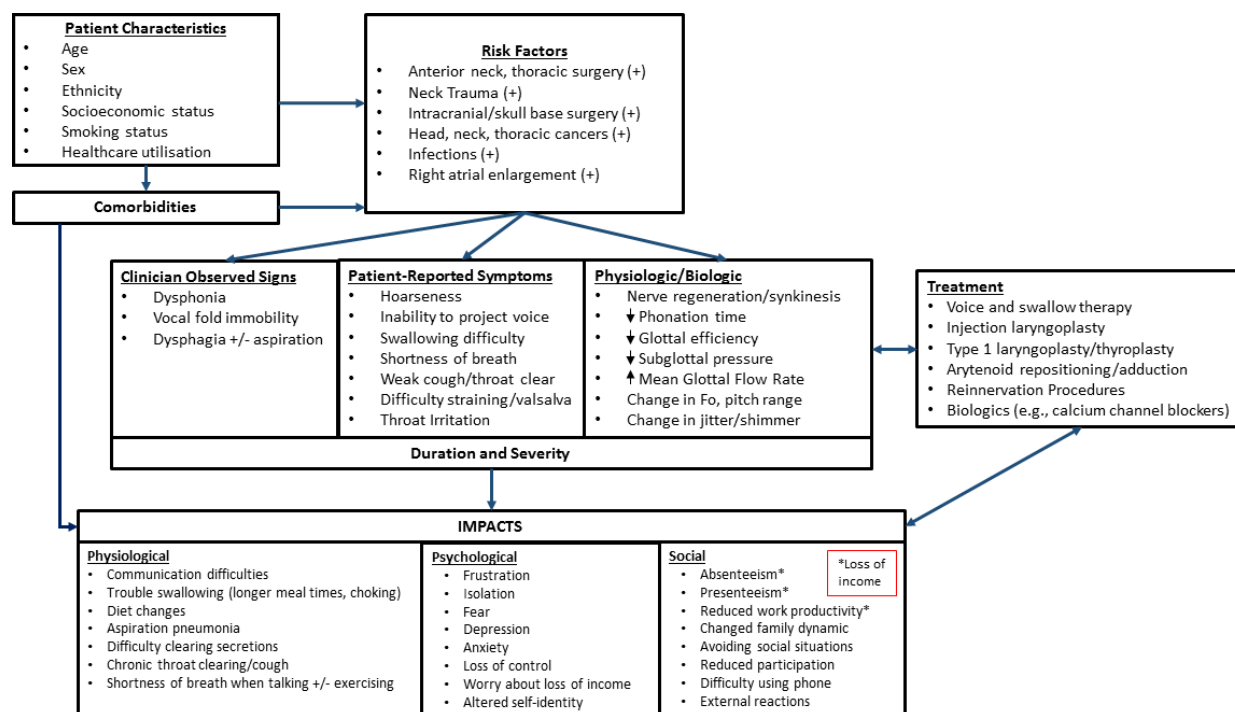
Patient and clinician interviews are reviewed briefly since they underpin the development of our conceptual framework and elicitation of PRO measure items. Patients were recruited from the Vanderbilt Voice Center and John S. Odess Head and Neck Surgery Clinic between July 2012 and July 2016. Our goal was to achieve patient heterogeneity with respect to UVFP-related symptoms, age, race, sex, employment, voice demands and socioeconomic status. In all, we conducted 75 semistructured interviews with English-speaking patients to best assess experienced UVFP-associated disability. A semistructured approach allowed participants to discuss their unique perspectives related to UVFP effects on quality of life and disability; detailed discussion of concept elicitation is described in related work.<sup>47</sup> All interviews were audio recorded and transcribed *verbatim* for data analysis. The principal investigator (DOF) and coding team subsequently identified saturation points in the patient interview transcripts to ensure all themes were adequately identified and described.

Interview data from both patients and treating clinicians helped inform the conceptual framework. In all, more than 90% of patients developed symptoms after undergoing neck or cardiothoracic surgery. All interviewees noticed an immediate voice change. The extent to which dysphonia affected HRQoL depended significantly on the

person's pre-morbid voice demands. Dysphagia affected 70% of patients with UVFP. Although liquid dysphagia is the traditional hallmark of UVFP-attributable swallowing dysfunction,<sup>48 49</sup> only 40% of dysphagic patients with UVFP identified liquids as a primary difficulty. The remainder identified other consistencies (30% solids, 10% all consistencies, 10% 'dry foods'). Interestingly, 70% of interviewees had overt, new-onset 'breathing difficulties',<sup>21</sup> and slightly more (74%) had 'shortness of breath' when talking. Respiratory complaints were diverse. Nearly 30% had difficulty with the Valsalva manoeuvre, described as new difficulty 'bearing down' or 'carrying heavy objects'. Inability to effectively clear secretions was also prevalent. Increased 'post-nasal drainage' or 'congestion' were described by 65% of patients and 55% could not 'efficiently clear secretions' from their throat. UVFP changed their ability to function in society, in their vocation, and had a greater than expected emotional and psychological toll.

In addition to the semistructured patient interviews, we facilitated a focus group of 10 voice-oriented, speech-language pathologists and completed 12 additional semistructured interviews with laryngologists dedicated to the care of this population. Participating laryngologists and speech-language pathologists had median years in practice of 15.9 years and 11.5 years, respectively. Physical, psychological, emotional and functional limitations identified by patients correlated poorly with those described in interviews with treating surgeons. Providers agreed that "lack of understanding" and poor education about the condition and its prognosis were a basis for fear in this patient population. One participating clinician remarked that patients with UVFP have "issues with identity, [particularly those] for whom their voice is an essential part of what they do and who they see themselves as". Several clinicians discussed this issue, with one summing up that patients' reaction to this injury "has a lot to do with what somebody brings to the table, what their coping style tends to be, which is obviously true of pretty much any event in life". Despite clinicians' ability to recognise these issues, clinicians and surgeons in particular tend to focus on correcting anatomical and physiological consequences of this injury. Thus, they spend a great deal of time in clinic and at professional meetings discussing the technical aspects of the physiological deficits and corrective surgical approaches. It is therefore not surprising that physical symptoms (dysphonia, dysphagia, dyspnoea) and pathophysiological considerations are well documented.<sup>21 22</sup> Due to time constraints and practice patterns, clinicians rarely prioritise discussion of seemingly peripheral sequelae (eg, psychological and social implications).<sup>18</sup> One clinician summarised his perspective: "it took [me] many years listening to patients to realize what they were saying" and continued stating: "we minimize the effects of unilateral paralysis in society because we see these people and they come and speak to us in their nice quiet voice, in our offices, and they can communicate to us; but if we just ran around the streets





**Figure 1** Conceptual framework.

and their work with them, some of them are really devastated with it”.

### Conceptual framework

Because no existing PRO measure assesses the full range of UVFP symptoms, including swallowing, voice and breathing problems, we used the results of our systematic review combined with interview data to develop a patient-centred conceptual framework that represents the broad range of UVFP symptoms and impacts (figure 1). Our framework includes the range of symptoms and the impact of the condition on HRQoL, mental health and interpersonal functioning. The framework accounts for the moderating effects of sociodemographic characteristics and comorbid conditions. We identify the variety of symptoms of UVFP and describe the complex interplay between symptoms and life circumstances, including HRQoL, employment, mental health and interpersonal functioning. We also include the scope of treatment options available to people with UVFP. This model serves as a working schematic that will evolve with our PRO measure development.

### Preliminary analysis and item development

#### Interview coding

We will develop a hierarchical codebook based on the proposed conceptual model that will be used to assess all interview transcripts. Transcripts will be compiled and systematically coded. Each statement by the interviewer or interviewee will be treated as a separate quote, and each interviewee quote will receive up to five content codes. Two trained analysts will code a subset of transcripts independently. Any coding discrepancies will be reconciled through a consensus process. If two coders

cannot agree, a third trained investigator will reconcile the difference. Final coded transcripts will be combined into one data file and sorted by code. Analysis will involve an inductive/deductive process of identifying common themes and relationships among themes. Deductively, we will be guided by the model in figure 1. Inductively, we will add to existing themes that are identified from specific quotes.

#### Item development

Candidate items for inclusion in our PRO measure will derive from the *verbatim* patient interview transcripts. Content validity will be established according to FDA criteria for PRO measure development by selecting representative quotes from within the hierarchically coded categories for item wording and evaluation for comprehensiveness, clarity and meaningful content by the patient population. Saturation will be confirmed when no new or important information arises that could contribute to our understanding of the patients’ perspectives and items.<sup>2</sup> Each item will use a 1-week recall period, will have a single target domain, will be written at less than a 6th grade reading level,<sup>50</sup> and will be clear, concise and worded to be understood by the patient population. We will use a Likert scale for all items, and include clear, consistent instructions and formatting.<sup>51</sup> A panel of content, instrument development and experts in psychometrics will assess items for appropriateness, content and clarity. Evaluation and standardisation of the recall period will be based on interview data and expert consensus. Candidate items will be compiled into an item bank and will undergo initial winnowing using PROMIS criteria: (1) content inconsistent; (2) semantically redundant with previous item(s);

(3) content too narrow for universal applicability; or (4) if confusing (eg, double barrelled).<sup>51</sup> Three investigators will review winnowing decisions to ensure a high level of consensus and impose process standardisation.

### Cognitive interviews

Retained items will be administered to 40 English-speaking participants with UVFP.<sup>52</sup> Participants will be asked to 'think out loud' about how they would answer the item, what they are thinking about as they answer the item, what is confusing about the item and what trouble, if any, they experienced in answering. The sample will be recruited to include a diverse group of this small population based on age, sex, education and race/ethnicity. The expert panel will use these cognitive interview findings to refine and further winnow items into those that will be included in *phase III: measure validation*.

## PHASE III: MEASURE VALIDATION

### Establishing a collaborative of tertiary care centres

To ensure that a representative patient population is engaged in the development of our PRO measure, we will form the vocal CoPE collaborative that includes tertiary care institutions that treat UVFP in a specialty otolaryngology setting. To develop the collaborative, the principal investigator (DOF) will invite Otolaryngology Department or Laryngology Division Chiefs at 50 medical centres in the USA to participate. Because this specialty care community is relatively small, the majority of physicians that treat UVFP know each other through training and professional associations; all 50 centres have agreed to participate. We will ask participating centres to designate primary point(s) of contact who can help identify barriers to enrolment and be involved in inter-site trial co-ordination. Patients will be incentivised to enrol, including US\$50 for completion of the first time point, and for each subsequent time point, a US\$5 pre-incentive and a US\$25 post-incentive if they complete the time point. The first survey will take 45 min to complete and the following surveys will take 5 to 10 min. Thus, the total possible incentive is planned to be US\$140 for most patients. Over 6-month follow-up, the cumulative time for completing all surveys will be about 90 min. For a subset of patients who will be invited to complete an additional time point, the total possible incentive is planned to be US\$170.

In managing the CoPE collaborative, the coordinating site (UW) opted to enrol patients and manage survey distribution centrally. Providers at collaborating sites will share informational brochures with affected patients, which directs them to a customised online platform housed and managed at UW. In this way, the protocol employs multicentre recruitment and provider engagement, but uses single-site enrolment. Thus, we focus on multisite stakeholder engagement at this stage and avoid a larger administrative role and oversight of sites,

including facilitation of institutional review board/regulatory activities, education and communication.

### Patient sampling and enrolment

In order to develop a well-constructed PRO measure assessing UVFP-attributable disability, we will test new PRO measure items in a representative, nationwide cohort of participants with this condition. These items will be evaluated to determine their latent structure. Administration to a large and diverse population of patients with UVFP will ensure its generalisability across socioeconomic and racial groups, geographical regions, healthcare systems, aetiologies and treatment status (eg, not treated, spontaneous recovered, surgically treated).

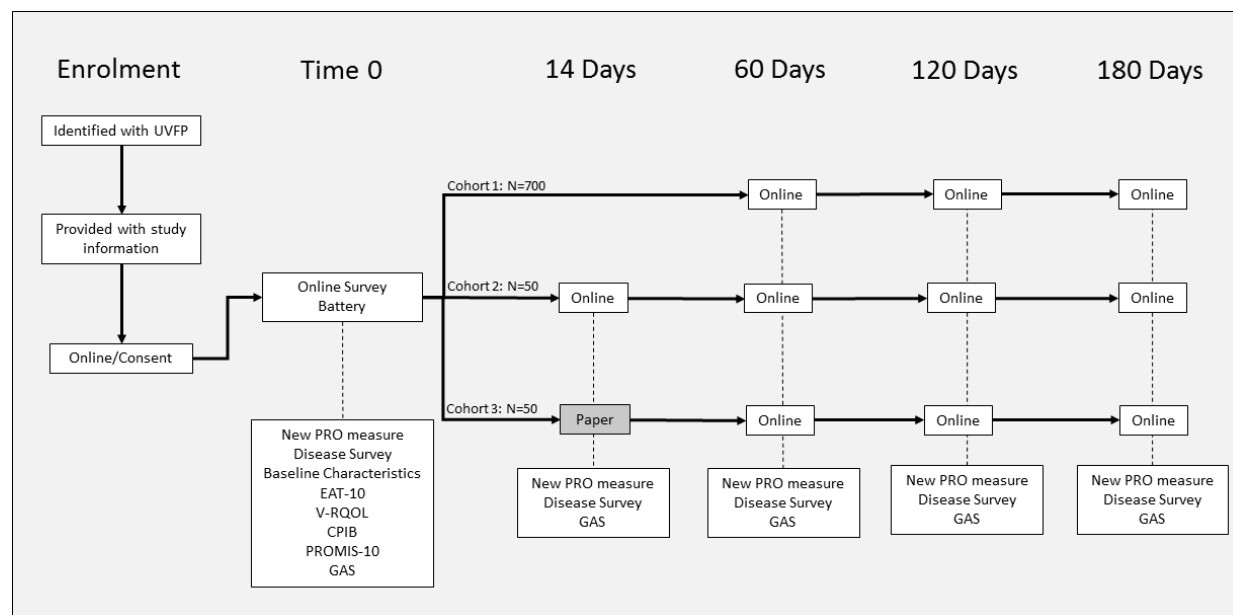
The goal is to enrol 800 English-speaking adult participants (≥18 years of age) with UVFP of varied aetiology (eg, iatrogenic, idiopathic) from 50 US centres. Direct visualisation by flexible laryngoscopy and otolaryngologist confirmation is required for diagnosis of UVFP (*standard of care*) and therefore study inclusion. Patients will be eligible for inclusion regardless of UVFP duration and treatment status in order to maximise symptom variability. Excluded will be patients with bilateral vocal fold movement abnormalities or if they have chronic voice disorders, objective dysphagia or objective pulmonary disease that predates UVFP onset.

### Data collection

Software programmers at the UW Survey Center will develop the Health Insurance Portability and Accountability Act (HIPAA)-compliant online patient-facing portal. Potential enrollees will be directed to the secure, online portal via the study brochure shared by their participating CoPE centre. Those interested will complete electronic consent, be administered the new PRO measure items and a battery of previously validated symptom-related and HRQoL-related questionnaires on a predetermined schedule (figure 2). University of Wisconsin Surgical Outcomes Research Center will serve as the Data Coordinating Center. Data collection will occur within a 2-year period from April 2019 to April 2021.

### PRO measure administration schedule

Each participant will complete questionnaires at multiple time points in English. There are three administration schedule cohorts (figure 2). At time 0, all enrollees will complete the new PRO measure items and a questionnaire battery online. The battery will include (1) a survey of baseline characteristics (eg, age, gender, race) and comorbidities (based on Charlson Comorbidity Index),<sup>53</sup> (2) a disease survey (eg, aetiology, date of UVFP diagnosis, any treatments/dates) and several existing PRO measures related to (3) general quality of life (PROMIS-10),<sup>51</sup> (4) voice-related quality of life (VRQOL),<sup>54</sup> (5) dysphagia-related quality of life (EAT-10),<sup>55</sup> (6) communication and participation (CPIB),<sup>56</sup> and a Global Assessment Scale (GAS)<sup>57</sup> (online supplementary appendix A–E).



**Figure 2** Cohorts involved, number of participants (n) and timeline of administration involved in PRO measure development. GAS, Global Assessment Scale; PRO, patient-reported outcome; UVFP, unilateral vocal fold paralysis.

After time 0, all participants regardless of cohort will complete the new PRO measure items and the disease survey every 2 months for 6 months (three administrations). These data are necessary for validity testing and for evaluating responsiveness to change. In addition, 100 randomly selected participants will be allocated into cohorts 2 (n=50) and 3 (n=50) and will undergo an additional administration to evaluate test-retest and alternative form reliability testing (figure 2). Two weeks after baseline survey completion, cohort 2 will be administered an online version of the new PRO measure items and the disease survey and cohort 3 will be mailed and complete paper versions. Automated email reminders and follow-up surveys will be automatically sent via our online interface to enrolled participants who completed time 0 activities.

## QUANTITATIVE ANALYSIS

### Exploratory and confirmatory factor analyses

All quantitative analyses and potential results are summarised in table 1, which is based on published quality criteria for assessing the measurement properties of health questionnaires.<sup>58</sup> Participant data derived from all cohorts will be randomised 1:1 (400:400) for exploratory principal component analysis (PCA) and confirmatory factor analysis (CFA), each analysis using data from time 0. PCA identifies latent variables/constructs and facilitates combining items into the final PRO measure and is an ideal approach when no a priori validated conceptual model exists.

We will ensure sample size adequacy for PCA based on (1) recommendations for participant:item ratios over varying strengths of inter-item correlations<sup>59</sup> and (2) how sampling and loadings interact in 'typical' samples of at

least 50 participants.<sup>60</sup> We anticipate that with 30 items in the preliminary tool, enrolling 400 unique participants meeting inclusion criteria will be sufficient for PCA. Approximately 50 high-volume laryngology practices that each see approximately 100–200 unique (new and returning) patients with UVFP per year will refer patients to this study. Conservative estimates, assuming a stable patient with UVFP volume, indicate that we should complete enrolment (n=800) within 2 years if we recruit and retain 30% of patients at each centre. The sample will result in a participant:item ratio of 5:1 or greater to support both PCA and CFA (even if inter-item correlations are modest).<sup>59</sup>

We will avoid retaining too few (underextraction) or too many (overextraction) components.<sup>60 61</sup> A simple structure solution will be sought such that each item loads substantively on one component. Rather than relying on a single criterion (eg, eigenvalues >1), we will use multiple criteria. These include (1) maximising the cumulative proportion of item covariance (and the related Scree plot), (2) minimum average partial correlation criterion,<sup>62</sup> (3) parallel analysis,<sup>63</sup> (4) well-saturated retained components with average loadings of about 0.6 and minimum loadings of about 0.4,<sup>60</sup> and (5) well-identified components, comprising at least four salient items.<sup>60</sup> Randomly generating a unique variable(s) will protect against 'factor splitting' if a single component solution is warranted.<sup>61</sup>

We will apply an orthogonal rotation initially, but we do not expect our choice of rotation method, orthogonal or oblique, to affect our decisions regarding the number of components. We will instead consider whether components are conceptually valid and interpretable. Clinical experts will review and characterise latent components. Finally, we will determine item scale characteristics (eg, endorsement rates, means,

**Table 1** Psychometric characteristics and quality scoring system to evaluate PRO measures (reprinted with adaptation from Terwee *et al*.<sup>58</sup>)

Characteristic	Description	Positive	Intermediate	Poor
Content Validity	Is the domain of interest comprehensively represented by items in instrument	Clear description of: 1. Measurement aim 2. Target population 3. Concepts measured 4. Item selection in target population	Clear description is lacking or only target population involved or doubtful method	No target population involved
Internal consistency	Are the items in scale measuring same construct	Factor analysis on adequate sample and Cronbach's alpha between 0.70 and 0.95 per dimension	No factor analysis or doubtful method	Cronbach's alpha(s) <0.70 or >0.95, despite sound methods
Criterion validity	Are scores on questionnaire related to gold standard	Convincing argument for a gold standard and correlation with gold standard $\geq 0.70$	No convincing arguments for gold standard or doubtful method	Correlation with gold standard <0.70, despite sound method
Construct validity	Do scores on questionnaire relate to other measures in a manner consistent with theoretically derived hypotheses about concepts measured; principal and confirmatory factor analyses	Specific hypotheses and at least 75% of results in support of hypotheses	Doubtful method (eg, no hypotheses)	<75% of hypotheses supported, despite sound methods
Agreement	Is the score on repeated measure close (absolute measurement error)	MIC <SDC or MIC outside of LOA or convincing argument that agreement acceptable	Doubtful method or MIC undefined and agreement unacceptable	MIC $\geq$ SDC or MIC outside of LOA, despite sound methods
Reliability	Are patients distinguished from each other, despite relative measurement errors	ICC or weighted Kappa $\geq 0.70$	Doubtful method (eg, time interval not mentioned)	ICC or weighted Kappa <0.70, despite sound methods
Responsiveness	Ability of questionnaire to detect clinically important changes over time	SDC or SDC >MIC or MIC outside LOA or RR >1.96 or AUC $\geq 0.70$	Doubtful design/method	SDC or SDC $\geq$ MIC or MIC equals or inside LOA or RR $\leq 1.96$ or AUC <0.70, despite sound methods
Floor and ceiling effects	No of respondents who achieved lowest and highest possible score	$\leq 15\%$ respondents achieved highest or lowest possible scores	Doubtful design/method	>15% respondents achieved highest or lowest scores despite sound methods
Interpretability	Can one assign qualitative meaning to quantitative scores	Mean/SD scores presented to >4 relevant patient subgroups and MIC defined	Doubtful method or <4 subgroups or MIC undefined	No information found on interpretation

Doubtful design or method=lacking clear description of study design or methods, sample <50 subjects or important methodological design weakness.

AUC, area under the curve; ICC, intraclass correlation; LOA, limits of agreement; MIC, minimally important change; PRO, patient-reported outcome; RR, relative risk; SD, Standard Deviation; SDC, smallest detectable change.



coefficient alpha) and develop a scoring algorithm. Although we anticipate the new PRO measure will include more than one component (domain), a 'general factor' or single component solution may be indicated. If so, we will adopt a summation scoring strategy and test its reliability and validity.

We will test the reproducibility of the latent variable structure identified in the exploratory PCA using CFA in a random sample of 400 participants. This approach will compare solutions with attention to the composition and proportion of covariance represented by each component in both exploratory and confirmatory samples. Standard fit indices (eg,  $\chi^2$ , adjusted goodness-of-fit index) will test overall model fit of the confirmatory sample to the underlying data and compare it with the exploratory sample.

### Reliability testing

We will assess two forms of reliability using participant data from time 0 and the 2-week administration (figure 2). A 2-week interval was selected to minimise potential carryover effect while retaining stable clinical symptomatology and severity. We will confirm any change in disease status (eg, treatment, spontaneous improvement) by concomitant administration of the disease survey. We will examine *test-retest reliability* (cohort 2, n=50) and *alternate form reliability* (cohort 3, n=50) using time 0 and 2-week PRO measure item data. For alternative form reliability, we expect that hardcopy and online administration modes will be statistically interchangeable. No enrollee will participate in both forms of reliability testing.

Both test-retest and alternate form reliability will be assessed using interclass correlation coefficients and by calculating per cent agreement. We will assume high within-subject correlation (0.9) between repeated administrations over the short term. Therefore, a sample of 50 will detect a small difference (0.2 SD) in scores with 81% power at 0.05 alpha level. We will determine the adequacy of correlation coefficient estimates relative to an accepted threshold ( $r \geq 0.8$ ).

We will calculate three forms of per cent agreement for each participant and across the sample: 'perfect' (same score for each item on repeat testing) and 'within 1' agreement. 'Within 1' agreement is used because items will likely be scored on a 5-point Likert scale. Participants who answer 'within 1' score of their first administration will have 'within 1' agreement. We expect 80% perfect and 90% 'within 1' agreement. Using a sample of 50 in both reliability forms, the SD for these estimates is 11.3% and 8.3%, respectively. We will also calculate a kappa coefficient to account for agreement by chance.

### Validity testing

Validity will be tested in several ways. Following the 2009 FDA guidelines, *content validity* is addressed via the generation of the item pool from qualitative patient interview transcripts, evaluation of item relevance and completeness, and clinical expert review of quantitative results. It will be reinforced by performing cognitive interviews with patients with UVFP. Item assessment

will be based on patient understanding and feedback derived from the transcripts.<sup>2</sup> All validity coefficients will be corrected for attenuation by incorporating estimates of PRO measure reliability.<sup>64</sup> *Construct validity* will be evaluated by time 0 co-administration of existing PRO measures directed at specific symptom domains (eg, V-RQOL<sup>54</sup>→voice). We will use a multi-trait approach to analyse data which will yield validity coefficients expected to vary from low to high in clinically meaningful ways that will be specified a priori based on our interpretations of the latent dimension(s)/scales(s) within the new PRO measure.

*Concurrent criterion-related validity* will be assessed by testing the association of scores obtained at initial PRO measure administration with physiological measures including maximum phonatory time (MPT), flexible laryngoscopy findings, modified barium swallow study (MBSS) and laryngeal electromyography (LEMG) results. Symptoms and thus PRO measure scores should correlate with MPT, which is a surrogate for ease of phonatory vocal fold closure known to vary significantly in patients with UVFP.<sup>65 66</sup> Relevant scale scores should also correlate with degree of phonatory vocal fold closure directly visualised on flexible laryngoscopy examination. Patients with associated dysphagia routinely undergo MBSS. Objective dysphagia findings on MBSS should correlate with relevant PRO measure scores. Finally, LEMG findings are associated with degree of nerve injury and permanency of UVFP.<sup>67</sup> Physiological results should correlate with scores to provide additional measures of concurrent criterion-related validity.

We will also evaluate *responsiveness to time and intervention*. We will record the timing and type of intervention(s) (eg, surgery). Longitudinal survey data should be sensitive to spontaneous improvement (responsive to time). In general, affected patients experience spontaneous recovery within 6 to 9 months of injury (if it occurs).<sup>68</sup> We chose a 6-month follow-up period because patients with UVFP have a median 3-month delay from symptom onset to presentation.<sup>68–70</sup> Thus, newly diagnosed participants will likely be surveyed between 3 and 9 months after symptom onset; this is the period during which spontaneous recovery of the affected recurrent laryngeal nerve is likely to occur.<sup>68</sup> Based on published data,<sup>69 71</sup> we expect that the majority of participants will have some spontaneous vocal recovery within this period with ~45% experiencing 'complete' and ~25% partial recovery.<sup>68</sup> Approximately 80% will have an intervention within that period (eg, speech therapy and/or procedure/surgery).<sup>69</sup> Thus, the anticipated sample, n=800, should include  $\geq 600$  unique participants with *acute* UVFP (<12 months since inception) who will have 6 months of longitudinal PRO measure data. This sample size is adequate (power  $\geq 80\%$ ,  $\alpha=0.05$ ) to conduct planned statistical tests. We will assess the magnitude of observed validity coefficients relative to small thresholds for meaningful differences. We will use longitudinal disability scores to assess 'recovery' and the effects of interventions on the PRO measure score.



These preliminary data will also establish a minimally clinically important difference in score for the PRO measure. We expect to detect a meaningful difference (0.5 SD) among participants with acute or chronic UVFP.<sup>72</sup> Participants without a meaningful effect size are unlikely to demonstrate responsiveness to change. Instead, these data would reinforce the stability (*reliability*) of the PRO measure over time and provide additional evidence of *discriminant validity*. We expect 75% of participants (n=600) will have had UVFP for <12 months (acute) and have a meaningful improvement over the 6-month follow-up period. The remainder will be characterised as having stable UVFP.

### Scoring manual

We will use our analyses to develop a scoring manual to guide future users of the new PRO measure. Traditional psychometric characteristics will be reported. For example, determining the number of respondents who achieved lowest and highest possible scores will assess floor and ceiling effects. Scoring methodology will be adjusted if  $\geq 15\%$  of respondents' score at the floor (lowest) and ceiling (highest) as this would impact the PRO measure's discriminative ability. We will also report distinct latent variables/ domains identified in PCA and confirmed in CFA. Detailed item analyses and all evidence of model fit, reliability and validity will be reported in the scoring manual.

### Future study

Our multisite approach to PRO measure development facilitates stakeholder engagement and information technology infrastructure for future multisite trials. We will leverage the PRO measure developmental process to engage clinical stakeholders via the CoPE collaborative. These 50 participating sites will refer patients to our PRO measure development study, allowing us to identify and address any site-specific barriers to recruitment prior to initiating a large-scale trial. In addition, creating a HIPAA-compliant central data collection interface and repository during measure development will lay the groundwork to collect similar data during planned future trials. In so doing, our PRO measure developmental activities will generate both essential preliminary data and multisite collaboration for future trials comparing treatments for UVFP.

### Author affiliations

<sup>1</sup>Department of Surgery, University of Wisconsin, Madison, Wisconsin, USA

<sup>2</sup>Department of Psychology, Vanderbilt University, Nashville, Tennessee, USA

<sup>3</sup>University of Wisconsin Survey Center, University of Wisconsin, Madison, Wisconsin, USA

<sup>4</sup>Department of Biostatistics, Vanderbilt University, Nashville, Tennessee, USA

<sup>5</sup>Division of Otolaryngology, Department of Surgery, University of Wisconsin, Madison, Wisconsin, USA

**Contributors** SF-T drafted and revised the manuscript with input from all other authors. CDS and NA provided project support, administration and performed qualitative analyses. KB and DS designed and performed the qualitative methods. MW directed design for survey administration for the online platform. IF provided survey and biostatistical guidance. DOF obtained funding, designed the study, performed systematic reviews, provided oversight of the study, and drafted and

revised the manuscript. All authors reviewed, contributed to and approved the final manuscript.

**Funding** This paper is supported by National Institutes of Health grants 5K23DC013559-07 and 1R21DC016724-01.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The University of Wisconsin Institutional Review Board approved the study protocol as minimal risk (IRB no. 140075).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### REFERENCES

- 1 Mercieca-Bebber R, King MT, Calvert MJ, *et al.* The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas* 2018;9:353–67.
- 2 US Department of Health and Human Services Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims 2009.
- 3 Merati AL, Shemirani N, Smith TL, *et al.* Changing trends in the nature of vocal fold motion impairment. *Am J Otolaryngol* 2006;27:106–8.
- 4 Ko H-C, Lee L-A, Li H-Y, *et al.* Etiologic features in patients with unilateral vocal fold paralysis in Taiwan. *Chang Gung Med J* 2009;32:290–6.
- 5 Chen H-C, Jen Y-M, Wang C-H, *et al.* Etiology of vocal cord paralysis. *ORL J Otorhinolaryngol Relat Spec* 2007;69:167–71.
- 6 Rosenthal LHS, Benninger MS, Deeb RH. Vocal fold immobility: a longitudinal analysis of etiology over 20 years. *Laryngoscope* 2007;117:1864–70.
- 7 Bhattacharyya N, Fried MP. Assessment of the morbidity and complications of total thyroidectomy. *Arch Otolaryngol Head Neck Surg* 2002;128:389–92.
- 8 Jeannon J-P, Orabi AA, Bruch GA, *et al.* Diagnosis of recurrent laryngeal nerve palsy after thyroidectomy: a systematic review. *Int J Clin Pract* 2009;63:624–9.
- 9 Zakaria HM, Al Awad NA, Al Kreedes AS, *et al.* Recurrent laryngeal nerve injury in thyroid surgery. *Oman Med J* 2011;26:34–8.
- 10 Kriskovich MD, Apfelbaum RI, Haller JR. Vocal fold paralysis after anterior cervical spine surgery: incidence, mechanism, and prevention of injury. *Laryngoscope* 2000;110:1467–73.
- 11 Jung A, Schramm J, Lehnardt K, *et al.* Recurrent laryngeal nerve palsy during anterior cervical spine surgery: a prospective study. *J Neurosurg Spine* 2005;2:123–7.
- 12 Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA* 2006;295:2164–7.
- 13 Marawar S, Girardi FP, Sama AA, *et al.* National trends in anterior cervical fusion procedures. *Spine* 2010;35:1454–9.
- 14 National Cancer Institute. *SEER Cancer Statistics Review 1975–2007*, 2012.
- 15 Benninger MS, Gillen JB, Altman JS. Changing etiology of vocal fold immobility. *Laryngoscope* 1998;108:1346–50.
- 16 Cohen SM, Kim J, Roy N, *et al.* Prevalence and causes of dysphonia in a large treatment-seeking population. *Laryngoscope* 2012;122:343–8.
- 17 Cantarella G, Dejonckere P, Galli A, *et al.* A retrospective evaluation of the etiology of unilateral vocal fold paralysis over the last 25 years. *Eur Arch Otorhinolaryngol* 2017;274:347–53.
- 18 Cohen SM, Kim J, Roy N, *et al.* The impact of laryngeal disorders on work-related dysfunction. *Laryngoscope* 2012;122:1589–94.
- 19 Cohen SM, Dupont WD, Courey MS. Quality-of-life impact of non-neoplastic voice disorders: a meta-analysis. *Ann Otol Rhinol Laryngol* 2006;115:128–34.
- 20 Fang T-J, Li H-Y, Gliklich RE, *et al.* Quality of life measures and predictors for adults with unilateral vocal cord paralysis. *Laryngoscope* 2008;118:1837–41.
- 21 Brunner E, Friedrich G, Kiesler K, *et al.* Subjective breathing impairment in unilateral vocal fold paralysis. *Folia Phoniatr Logop* 2011;63:142–6.

- 22 Bhattacharyya N, Kotz T, Shapiro J. Dysphagia and aspiration with unilateral vocal cord immobility: incidence, characterization, and response to surgical treatment. *Ann Otol Rhinol Laryngol* 2002;111:672–9.
- 23 Paniello RC, Edgar JD, Kallogjeri D, et al. Medialization versus reinnervation for unilateral vocal fold paralysis: a multicenter randomized clinical trial. *Laryngoscope* 2011;121:2172–9.
- 24 Pei Y-C, Fang T-J, Hsin L-J, et al. Early hyaluronate injection improves quality of life but not neural recovery in unilateral vocal fold paralysis: an open-label randomized controlled study. *Restor Neurol Neurosci* 2015;33:121–30.
- 25 Périé S, Roubeau B, Liesenfelt I, et al. Role of medialization in the improvement of breath control in unilateral vocal fold paralysis. *Ann Otol Rhinol Laryngol* 2002;111:1026–33.
- 26 Cates DJ, Venkatesan NN, Strong B, et al. Effect of vocal fold medialization on dysphagia in patients with unilateral vocal fold immobility. *Otolaryngol Head Neck Surg* 2016;155:454–7.
- 27 Gliklich RE, Glovsky RM, Montgomery WW. Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngol Head Neck Surg* 1999;120:153–8.
- 28 Branski RC, Cukier-Blaj S, Pusic A, et al. Measuring quality of life in dysphonic patients: a systematic review of content development in patient-reported outcomes measures. *J Voice* 2010;24:193–8.
- 29 Francis DO, Patel DA, Sharda R, et al. Patient-reported outcome measures related to laryngopharyngeal reflux: a systematic review of instrument development and validation. *Otolaryngol Head Neck Surg* 2016;155:923–35.
- 30 Francis DO, Daniero JJ, Hovis KL, et al. Voice-related patient-reported outcome measures: a systematic review of instrument development and validation. *J Speech Lang Hear Res* 2017;60:62–88.
- 31 Patel DA, Sharda R, Hovis KL, et al. Patient-reported outcome measures in dysphagia: a systematic review of instrument development and validation. *Dis Esophagus* 2017;30:1–23.
- 32 Noud M, Hovis K, Gelbard A, et al. Patient-reported outcome measures in upper airway-related dyspnea: a systematic review. *JAMA Otolaryngol Head Neck Surg* 2017;143:824–31.
- 33 Walton C, Carding P, Conway E, et al. Voice outcome measures for adult patients with unilateral vocal fold paralysis: a systematic review. *Laryngoscope* 2019;129:187–97.
- 34 Speyer R, Kim J-H, Doma K, et al. Measurement properties of self-report questionnaires on health-related quality of life and functional health status in dysphonia: a systematic review using the COSMIN taxonomy. *Qual Life Res* 2019;28:283–96.
- 35 McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Med* 2011;9:86.
- 36 McTaggart-Cowan HM, Marra CA, Yang Y, et al. The validity of generic and condition-specific preference-based instruments: the ability to discriminate asthma control status. *Qual Life Res* 2008;17:453–62.
- 37 Canestaro WJ, Edwards TC, Patrick DL. Systematic review: patient-reported outcome measures in coeliac disease for regulatory submissions. *Aliment Pharmacol Ther* 2016;44:313–31.
- 38 Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
- 39 Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health* 2008;11:967–77.
- 40 Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health* 2011;14:978–88.
- 41 Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
- 42 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- 43 Reeve BB, Wyrrich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889–905.
- 44 Morse JM. *Essentials of qualitatively-drive mixed-method designs*. New York: Routledge, 2016.
- 45 Wisconsin Network for Research Support. Cards: community advisors on research design and strategies. Available: <https://winrs.nursing.wisc.edu/services/cards/>
- 46 Francis DO, McPheeters ML, Noud M, et al. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Syst Rev* 2016;5:129.
- 47 Francis DO, Sherman AE, Hovis KL, et al. Life experience of patients with unilateral vocal fold paralysis. *JAMA Otolaryngol Head Neck Surg* 2018;144:433–9.
- 48 Bhattacharyya N, Kotz T, Shapiro J. The effect of bolus consistency on dysphagia in unilateral vocal cord paralysis. *Otolaryngol Head Neck Surg* 2003;129:632–6.
- 49 Leder SB, Ross DA. Incidence of vocal fold immobility in patients with dysphagia. *Dysphagia* 2005;20:163–7. discussion 168–169.
- 50 Wilkinson GS, Robertson GJ. *Wide Range Achievement Test 4 Professional Manual*. Lutz, FL: Psychological Assessment Resources, 2006.
- 51 Ader DN. Developing the patient-reported outcomes measurement information system (PROMIS). *Med Care* 2007;45(Suppl 1):S1–2.
- 52 Krueger RA, Casey MA. *Focus groups : a practical guide for applied research*. 4th edn. Los Angeles: SAGE, 2009.
- 53 Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
- 54 Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice* 1999;13:557–69.
- 55 Belafsky PC, Mouadeb DA, Rees CJ, et al. Validity and reliability of the eating assessment tool (EAT-10). *Ann Otol Rhinol Laryngol* 2008;117:919–24.
- 56 Baylor C, Yorkston K, Eadie T, et al. The communicative participation item bank (CPIB): item bank calibration and development of a disorder-generic short form. *J Speech Lang Hear Res* 2013;56:1190–208.
- 57 Endicott J, Spitzer RL, Fleiss JL, et al. The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 1976;33:766–71.
- 58 Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- 59 Baggaley AR. Deciding on the ratio of the number of subjects to the number of variables in factor analysis. *Multivar Exp Clin Res* 1982;6:81–5.
- 60 Velicer WF, Fava JL. Affects of variable and subject sampling on factor pattern recovery. *Psychol Methods* 1998;3:231–51.
- 61 Wood JM, Tataryn DJ, Gorsuch RL. Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychol Methods* 1996;1:354–65.
- 62 Velicer WF. Determining the number of components from the matrix of partial correlations. *Psychometrika* 1976;41:321–7.
- 63 Lautenschlager GJ. A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behav Res* 1989;24:365–95.
- 64 Allen MJ, Yen WM. *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company, 1979.
- 65 Kent RD, Kent JF, Rosenbek JC. Maximum performance tests of speech production. *J Speech Hear Disord* 1987;52:367–87.
- 66 Bielamowicz S, Kapoor R, Schwartz J, et al. Relationship among glottal area, static supraglottic compression, and laryngeal function studies in unilateral vocal fold paresis and paralysis. *J Voice* 2004;18:138–45.
- 67 Rickert SM, Childs LF, Carey BT, et al. Laryngeal electromyography for prognosis of vocal fold palsy: a meta-analysis. *Laryngoscope* 2012;122:158–61.
- 68 Sulica L. The natural history of idiopathic unilateral vocal fold paralysis: evidence and problems. *Laryngoscope* 2008;118:1303–7.
- 69 Francis DO, Williamson K, Hovis K, et al. Effect of injection augmentation on need for framework surgery in unilateral vocal fold paralysis. *Laryngoscope* 2016;126:128–34.
- 70 Spataro EA, Grindler DJ, Paniello RC. Etiology and time to presentation of unilateral vocal fold paralysis. *Otolaryngol Head Neck Surg* 2014;151:286–93.
- 71 Husain S, Sadoughi B, Mor N, et al. Time course of recovery of idiopathic vocal fold paralysis. *Laryngoscope* 2018;128:148–52.
- 72 Norman GR, Sridhar FG, Guyatt GH, et al. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Med Care* 2001;39:1039–47.