Supplementary material

Methods

Item selection

The main objective was to develop a clinical tool consisting of only a few items to capture a wide range of respiratory, sleep and NIV treatment experiences, but with valid and reliable measurement properties. For this purpose, we selected all items pertaining to "respiratory complaints" and "attendant symptoms and sleep" from the French translation and cultural adaptation of the SRI questionnaire [1] for further psychometric validation. The need for an additional assessment of a "comfort" dimension was agreed on by all NIV experts in our group. To this end, issues related to the experience of NIV treatment (mainly comfort and side-effects) were investigated in depth during 15 qualitative interviews with patients (D.A.). Items were also selected from two NIV comfort scales used in previous studies by our group, but with no formal psychometric validation [2, 3]. All items were rated on a 5-point Likert scale from "strongly disagree" to "strongly agree" and were related to patient's status during the "last 4 weeks". Selected core items consisted of 22 questions. Of these, 8 questions were related to the "respiratory complaints" domain, 7 to "attendant symptoms and sleep", and 7 to NIV-related side-effects.

Item analysis and reduction

Concepts of psychometric analysis are summarised in table S1. The Mokken scale analysis (non-parametric Item Response Theory (IRT) analysis) was first performed to check the assumptions of unidimensionality, local independence, and latent monotonicity that are required by parametric IRT models. To evaluate whether the unidimensionality assumption was plausible, an exploratory factorial analysis using the polychoric correlations and the number of underlying dimensions was performed as a parallel analysis [4]. Factor loadings were then used to investigate the meaning of each dimension. At this stage, numerous IRT models (graded rating scale model, the graded response model, and the generalized partial credit mod) were tested on each unidimensional subscale for item reduction. The Akaike information criterion was applied to choose the IRT model that best fitted the data. The graded rating scale model was used to assess item information (i.e., item precision based on difficulty and discrimination for all values of the total scale). Items with a low information function or with information that was redundant with another item were removed [5]. Item fit statistics (signed chi-square test) were

checked for each item. Item reduction also took into consideration expert opinion from investigators and other clinical experts in the field. Mean values for each subscore were transformed accordingly with a total score ranging from 0 to 10. The lowest possible score (0) corresponded to the highest impact of disease and treatment on quality of life, while the highest possible score (10) corresponded to the lowest impact of disease and treatment on quality of life.

Reliability

Internal consistency of each subscale was calculated by Cronbach's alpha coefficient using data from all patients. Values >0.7 are considered acceptable for group comparison and values >0.8 are considered good or excellent [6]. Test-retest reliability of the final instrument was assessed by the intra-class correlation coefficient (ICC) in a subset of stable patients (n=43) who completed the questionnaire 4 weeks after initial testing. A value of 0.7 indicates good reliability. To assess the validity of the factor structure, the fit of a confirmatory item factor model was evaluated using the M2 model fit statistic [7], the root mean square error of approximation (RMSEA) and the standardised root mean square residual (SRMR). Differential item functioning analysis (DIF) was used to assess whether the selected items were perceived differently by patients in differing disease categories.

Construct and discriminant validity

Construct and discriminant validity was assessed by evaluating correlations between the final instrument and the French version of the St. George's Respiratory Questionnaire (SGRQ) [8] and the Quebec Sleep Questionnaire [9].

All analyses were performed with R version 3.4.4 [10] and the mirt[11] and psych packages[12].

Online supplementary table S1: Summary of psychometric concepts used in this study

Loading: measurement of the correlational relationship between the observed and the latent variable/s (also called "factor").

Undimensionality assumption: a scale is said to be unidimensional if it measures only one underlying construct (or one latent variable).

Local independence: this assumption implies that the response to any items is unrelated to any other item when the latent variable level is controlled.

Monotonicity (latent monotonicity): this means that the item step response functions are non-decreasing functions of the underlying construct, i.e. patients at a higher level of the underlying construct have a higher probability of scoring higher for an item.

Item response theory (IRT) is a group of specialised methods for the development and validation of rating scales with items of ordinal or nominal type. IRT considers a class of latent variable models that link mainly dichotomous and polytomous observed variables to a single latent variable.

Mokken scale analysis: an analytical method for ordinal data to assess the dimensionality and scalability of psychometric measures. It can be viewed as a non-parametric IRT model.

Differential item functioning occurs when subgroups of a sample respond in different ways to a particular item, despite having the same level of the underlying trait being measured.

Legend to online supplementary figures

Figure S1: Visual illustration of the two-factor model in 20 of 22 selected core items.

Ovals indicate latent factors MR1 and MR2 regressed on all items (boxes) in the exploratory factor analysis. The value linking MR1 to MR2 is the correlation between the two factors. The values in the single-headed arrows are the factor loadings standardised to both the latent factor and the variances observed among the variables of the factor analysis model. The 22 cores items were the following ones:

Q1: My ventilator doesn't inflate my lungs enough.

Q2: I am disturbed by leaks.

Q3: My mask is uncomfortable.

Q4: My ventilator is too fast-paced.

Q5: I receive too much air from my ventilator.

Q6: My ventilator is too noisy.

Q7: I suffer from nasal or oral dryness

Q8: I suffer from breathing problems when I eat.

Q9: I suffer from breathing problems even without physical exertion.

Q10: I often have a headache.

Q11: I sometimes feel dizzy.

Q12: I wake up at night with breathing difficulties.

Q13: I often have neck pain.

Q14: I often wake up at night.

Q15: I am often short of breath.

Q16: I have trouble breathing when I speak.

Q17: I cough a lot.

Q18: There is often mucus in my airways.

Q19: I have difficulties breathing during physical exertion.

Q20: I am tired during the day.

Q21: I go to sleep easily (not included in factor analysis).

Q22: I sleep through the night easily (not included in factor analysis).

Figure S2: Panel A. Correlation between the "respiratory symptoms" S^3-NIV questionnaire subscore and the SGRQ symptoms score. Panel B. Correlation between the "sleep and NIV side-effects" S^3-NIV questionnaire subscore and the Quebec Sleep Questionnaire Score.

SGRQ: St. George's Respiratory Questionnaire.

Supplementary references:

1. Cuvelier A, Lamia B, Molano LC, Muir JF, Windisch W: **[The French translation and cultural adaptation of the SRI questionnaire. A questionnaire to assess health-related quality of life in patients with chronic respiratory failure and domiciliary ventilation]**. *Rev Mal Respir* 2012, **29**(5):705-713.
2. Adler D, Perrig S, Takahashi H, Espa F, Rodenstein D, Pepin JL, Janssens JP: **Polysomnography in stable COPD under non-invasive ventilation to reduce patient-ventilator asynchrony and morning breathlessness**. *Sleep Breath* 2012, **16**(4):1081-1090.
3. Janssens JP, Metzger M, Sforza E: **Impact of volume targeting on efficacy of bi-level non-invasive ventilation and sleep in obesity-hypoventilation**. *Respir Med* 2009, **103**(2):165-172.
4. Horn JL: **A rationale and test for the number of factors in factor analysis**. *Psychometrika* 1965, **30**(2):179-185.
5. Nering ML, Ostini R: **Handbook of polytomous item response theory models**: Taylor & Francis; 2011.
6. Cronbach LJ: **Coefficient alpha and the internal structure of tests**. *psychometrika* 1951, **16**(3):297-334.
7. Cai L, Hansen M: **Limited-information goodness-of-fit testing of hierarchical item factor models**. *British Journal of Mathematical and Statistical Psychology* 2013, **66**(2):245-276.
8. Bourbeau J, Maltais F, Rouleau M, Guimont C: **French-Canadian version of the Chronic Respiratory and St George's Respiratory questionnaires: an assessment of their psychometric properties in patients with chronic obstructive pulmonary disease**. *Can Respir J* 2004, **11**(7):480-486.
9. Lacasse Y, Bureau MP, Series F: **A new standardised and self-administered quality of life questionnaire specific to obstructive sleep apnoea**. *Thorax* 2004, **59**(6):494-499.
10. Core Team R: **R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016**. In.; 2017.
11. Chalmers RP: **mirt: A multidimensional item response theory package for the R environment**. *Journal of Statistical Software* 2012, **48**(6):1-29.
12. Revelle W: **psych: Procedures for Personality and Psychological Research**. *Northwestern University, Evanston, Illinois, USA* 2018.