

RESEARCH

Open Access



Mepolizumab improvements in health-related quality of life and disease symptoms in a patient population with very severe chronic rhinosinusitis with nasal polyps: psychometric and efficacy analyses from the SYNAPSE study

Wyske Fokkens¹, Andrew Trigg², Stella E. Lee³, Robert H. Chan^{4*}, Zuzana Diamant^{5,6,7}, Claire Hopkins⁸, Peter Howarth⁹, Valerie Lund¹⁰, Bhabita Mayer¹¹, Ana R. Sousa⁴, Steve Yancey¹² and Maggie Tabberer¹³ on behalf of SYNAPSE study group

Abstract

Background Although the psychometric properties of patient-reported outcome measures (e.g. the 22-item Sino-nasal Outcomes Test [SNOT-22]) in chronic rhinosinusitis with nasal polyps (CRSwNP) have been defined, these definitions have not been extensively studied in patients with very severe CRSwNP, as defined by recurrent disease despite ≥ 1 previous surgery and a current need for further surgery. Therefore, the psychometric properties of the symptoms visual analogue scales (VAS) were evaluated, and meaningful within-patient change thresholds were calculated for VAS and SNOT-22.

Methods SYNAPSE (NCT03085797), a randomized, double-blind, placebo-controlled, 52-week trial, assessed the efficacy and safety of 4-weekly mepolizumab 100 mg subcutaneously added to standard of care in very severe CRSwNP. Enrolled patients ($n = 407$) completed symptom VAS (six items) daily and SNOT-22 every 4 weeks from baseline until Week 52. Blinded psychometric assessment of individual and composite VAS was performed post hoc, including anchor-based thresholds for meaningful within-patient changes for VAS and SNOT-22, supported by cumulative distribution function and probability density function plots. The effect of mepolizumab versus placebo for 52 weeks on VAS and SNOT-22 scores was then determined using these thresholds using unblinded data.

Results Internal consistency was acceptable for VAS and SNOT-22 scores (Cronbach's α -coefficients ≥ 0.70). Test-retest reliability was demonstrated for all symptom VAS (Intra-Class Correlation coefficients > 0.75). Construct validity was acceptable between individual and composite VAS and SNOT-22 total score ($r = 0.461$ – 0.598) and between individual symptom VAS and corresponding SNOT-22 items ($r = 0.560$ – 0.780), based upon pre-specified ranges. Known-groups validity assessment demonstrated generally acceptable validity based on factors associated with respiratory health, with all VAS responsive to change. Mepolizumab treatment was associated with significantly increased odds

*Correspondence:

Robert H. Chan
robert.h.chan@gsk.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

of meeting or exceeding meaningful within-patient change thresholds, derived for this very severe cohort using six anchor groups for individual VAS (odds ratio [OR] 2.19–2.68) at Weeks 49–52, and SNOT-22 (OR 1.61–2.96) throughout the study.

Conclusions Symptoms VAS and SNOT-22 had acceptable psychometric properties for use in very severe CRSwNP. Mepolizumab provided meaningful within-patient improvements in symptom severity and health-related quality of life versus placebo, indicating mepolizumab provides substantial clinical benefits in very severe CRSwNP.

Keywords SNOT-22, VAS, Chronic rhinosinusitis with nasal polyps, Psychometric, Efficacy, Quality of life, Severity, Mepolizumab, Patient-reported outcome measures (PROMs)

Plain English summary

Patients with chronic rhinosinusitis (CRS) often have blocked or runny noses, and loss of sense of smell. They can also have sac-like growths in their nose called nasal polyps, which often require surgical removal. The symptoms of CRS with nasal polyps can affect quality of life. In a clinical study named SYNAPSE, a new treatment option called mepolizumab reduced the size and severity of nasal polyps in patients suffering from very severe CRS with nasal polyps, compared with placebo. Mepolizumab also reduced the need for nasal polyp surgery. The SYNAPSE study also measured if 1 year of mepolizumab treatment improved patients' symptoms and quality of life. This was evaluated by asking patients to complete two separate tasks. These tasks were rating symptoms on a visual analogue scale (VAS) and completing a quality of life questionnaire called SNOT-22. The objective of this analysis was to see if these questionnaires accurately assessed a patient's quality of life. The analysis also assessed how many patients had major improvements in their symptoms with mepolizumab. Overall, data from 407 patients in the SYNAPSE study was analyzed. Results showed that both the VAS and SNOT-22 questionnaires accurately captured CRS symptoms and quality of life. In addition, patients treated with mepolizumab for 1 year had improvements in quality of life compared with placebo. In conclusion, these findings suggest that the VAS and SNOT-22 questionnaires are appropriate evaluation tools for patients with very severe CRS with nasal polyps. The findings also show that mepolizumab treatment is beneficial for these patients.

Background

Chronic rhinosinusitis with nasal polyps (CRSwNP) is a subtype of CRS, characterized by persistent inflammation of the paranasal sinuses [1–3]. Inflammation is typically eosinophilic in nature, driven by type 2 cytokines such as interleukin (IL)-4, IL-5, and IL-13 [4, 5], resulting in symptoms of nasal blockage, loss of sense of smell, nasal discharge and/or facial pain [1, 6]. The current standard of care (SoC) for CRSwNP includes intranasal corticosteroids, saline nasal irrigation, short courses of systemic corticosteroids (SCS) for exacerbations, and sinus surgery when appropriate medical therapy fails [1]. However, these treatments have substantial limitations, including adverse events associated with SCS use, and nasal polyp (NP) recurrence following surgery [7, 8], highlighting the need for new, effective treatment options.

The anti-IL-5 humanized monoclonal antibody mepolizumab is approved for the treatment of CRSwNP in the US and EU [9–11]. In patients with CRSwNP in previous Phase II trials, mepolizumab has been shown to improve symptoms, reduce NP size and reduce the need for sinus surgery compared with placebo [12, 13]. In addition, results from SYNAPSE, a Phase III trial have shown that in adults with very severe CRSwNP in need of revision

surgery, 4-weekly mepolizumab 100 mg subcutaneously (SC) plus SoC significantly improved NP size and nasal obstruction versus placebo, with no new safety findings found in addition to those previously reported in patients with CRSwNP treated with mepolizumab [12–14].

Since CRSwNP significantly impacts health-related quality of life (HRQoL) [12, 15, 16], clinical trials assessing the efficacy of novel treatments in patients with CRSwNP have included patient-reported outcome measures (PROMs) to evaluate symptom severity (HRQoL via the use of, for example, the visual analogue scale (VAS) and Sino-nasal Outcome Test (SNOT)-22, respectively, as supported by the European Position Paper on Rhinosinusitis and Nasal Polyposis (EPOS 2020) [1]. The psychometric properties of the VAS were assessed in a Phase II study assessing mepolizumab clinical efficacy in CRSwNP [17]. The analyses indicated that although overall patient comprehension of the VAS was good, there was room for improvement. Consequently, to improve clarity (and therefore data quality) of the VAS, modifications were made to the item names and the verbal descriptors, the facial pain/pressure VAS item was added, and a 24-h recall period was implemented for use in the Phase III SYNAPSE study [14, 18]. As VAS is a key endpoint in

CRSwNP research, it was necessary to repeat the psychometric analysis of the modified VAS using Phase III data. Furthermore, as the score interpretation thresholds available to assess treatment response in CRSwNP trials are based on a CRSwNP population of mixed severity [19, 20], information on thresholds required for patients with the most severe disease was needed as the burden of symptoms in patients with severe CRSwNP is high and has a significant impact on HRQoL [21]. Indeed, the need for different thresholds for patients with the most severe disease has been demonstrated in the asthma field [19, 22]. As such, the objectives of this post hoc analysis were two-fold. First, to evaluate the psychometric properties of VAS scores and derive meaningful within-patient change thresholds for VAS and SNOT-22 in patients with very severe CRSwNP using blinded data from SYNAPSE. Second, patients were classified as responders or non-responders using meaningful within-patient change thresholds thereby assessing the efficacy of mepolizumab 100 mg SC administered every 4 weeks on symptom severity and HRQoL in adults with very severe CRSwNP.

Methods

Study design and patients

SYNAPSE was a Phase III randomized, double-blind, placebo-controlled, parallel-group trial (GSK205687; NCT03085797; Additional file 1: Figure S1) [14]. Patients (N = 407) were randomized (1:1) to mepolizumab 100 mg SC (n = 206) or placebo (n = 201) every 4 weeks, for 52 weeks in addition to SoC, including daily mometasone furoate nasal spray throughout the study period, saline nasal irrigations, and courses of SCS and/or antibiotics, as required [14].

Patient eligibility criteria included having a diagnosis of very severe CRS, characterized by (1) the presence of recurrent, refractory severe bilateral NP symptoms, (2) ≥ 1 prior surgery for NPs (a procedure involving instruments with resulting tissue removal) in the past 10 years, and 3) a current need for surgery (defined as an overall symptoms VAS score > 7 and endoscopic NP score of ≥ 5 [maximum 8], with a score ≥ 2 in each nasal cavity). Inclusion criteria for this study resulted in patients with very severe CRSwNP [14].

VAS and SNOT-22 assessments

Patients completed six individual VAS assessments daily (nasal obstruction, nasal discharge, mucus in the throat, loss of sense of smell, facial pain, and overall symptoms) using an electronic diary (eDiary) and a recall period of 24 h. For each VAS, patients rated their symptom on a scale from 'none' (0) to 'as bad as you can imagine' (100). Results were transformed linearly to a 0–10 range for reporting (higher scores indicated

greater symptom severity); the psychometric properties were unaffected by this transformation.

Patients also completed SNOT-22 assessments every 4 weeks using the same electronic device and a recall period of 2 weeks. A total of 22 symptoms were rated on a 0–5 scale, with a final total score of 0–110 (higher scores indicated greater disease impact on HRQoL). Endpoints from SYNAPSE relevant to this analysis are listed in Additional file 1: Supplementary Methods.

Psychometric assessments

Blinded assessments of VAS and SNOT-22 were performed post hoc using data from the intent-to-treat population. Each VAS item was assessed separately, in addition to proposed four-item composite (nasal symptoms) and five-item composite (nasal symptoms and facial pain) scores (Additional file 1: Figure S2). Daily VAS scores at the Week 20 visit were used for confirmatory factor analysis (CFA) and internal consistency analyses; 4-week average VAS scores were used for analyses of the VAS over time, for example, test–retest (assessed at Week 20 and Week 24). Key analysis time points for efficacy assessments were Baseline, Weeks 20, 24, and 52. Where available, Week 20 data were prioritized for cross-sectional psychometric analyses as this time point was expected to have maximal change from baseline in individual patient scores while maximizing the number of patients included in the analyses (minimal patient drop out at this time point). Floor and ceiling effects, test–retest reliability, known-group validity, and ability to detect change were not assessed for SNOT-22 as these have been previously demonstrated [14]. No imputation of missing data was performed; 4-week averages were derived based on all available data within that timeframe.

Item characteristics

VAS ceiling and floor effects were explored, defined as $> 15\%$ of patients selecting the most severe health state ('worst imaginable' [ceiling]) or least severe health state ('none' [floor]) [23]. However, owing to the severity of the CRSwNP in patients included in this study, it was anticipated that scores would be distributed towards the higher end of the scale at baseline and ceiling effects may have been present at this time point. Inter-item correlations were therefore assessed for the individual symptom VAS scores (Spearman's correlations to allow for violation of normality) and the SNOT-22 items (polychoric) at Week 20. Items that correlated very highly with one another ($r \geq 0.90$) were flagged as potentially indicating redundancy.

Domain structure

CFA was performed to summarize the shared variance between items for the nasal symptoms composite VAS score, nasal symptoms and facial pain composite VAS score and SNOT-22 hypothesized domain structure [24, 25] using Week 20 data. Maximum likelihood estimator was used for continuous VAS items. A standardized loading of >0.40 was considered indicative of an item as an adequate indicator of the factor [26]. Global model fit was assessed using Chi-square (values $P \geq 0.05$ indicate acceptable fit), comparative fit index (CFI; values ≥ 0.95 indicate acceptable fit), standardized root mean residual (SRMR; values <0.10 are considered acceptable), and root mean square error of approximation (RMSEA) models (values <0.10 are considered acceptable) [24, 27, 28].

Reliability

Internal consistency, the extent to which responses to individual items within a score are interrelated, was investigated for both VAS composite scores and the SNOT-22 by calculating Cronbach's α coefficient for each score, with a coefficient ≥ 0.70 considered acceptable [29]. Test-retest reliability, the degree to which scores are similar between two time points in patients who are stable, was evaluated for the individual symptoms VAS, overall symptoms VAS and both VAS composite scores by calculating the Intra-Class Correlation (ICC) coefficient for VAS and assessing stability between Weeks 20 and 24 (Additional file 1: Table S1). The ICC was based on a multiple measurement, absolute agreement, two-way random effects model (equivalent to ICC 2,k) [30]. Samples of all stable patients (between Weeks 20 and 24) were defined as patients that had no improvement or worsening according to other measures (detailed in Additional file 1: Supplementary Methods). Test-retest reliability was interpreted as follows: $ICC < 0.5$ poor reliability, $0.75 > ICC \geq 0.5$ moderate reliability, $ICC \geq 0.75 - 0.9$ good reliability, and $ICC > 0.90$ excellent reliability [30] (See Additional file 1: Supplementary Methods for the formula).

Validity

Construct validity, whether similar concepts are more correlated than dissimilar concepts, was assessed at Week 20 for individual and overall symptoms VAS and both composite VAS scores using Spearman's correlations. A priori hypothesized construct validity correlations were assessed with the following measures: The University of Pennsylvania Smell Identification Test (UPSIT) [31], endoscopic NP score [32], Peak Nasal Inspiratory Flow (PNIF) [33], SNOT-22 total score and individual items [34, 35], 36-item Short Form Health Survey (SF-36v2; Physical and Mental Component Summary scores [PCS

and MCS]) [36], and Work Productivity and Activity Impairment (WPAI) Questionnaire (absenteeism, presenteeism, work productivity loss and activity impairment) (Additional file 1: Table S2) [37]. Correlations were based on predefined ranges (high: ≥ 0.50 , medium: ≥ 0.30 to <0.50 , low: <0.30) [38].

The known-groups method (to differentiate between clinically distinct groups [29]) was used to evaluate the construct validity of the individual and overall symptoms VAS and both composite VAS scores. VAS scores were compared in patients grouped according to ranges of key baseline characteristics: comorbid asthma, baseline blood eosinophil count categories, non-steroidal anti-inflammatory drug-exacerbated respiratory disease (N-ERD), Asthma Control Questionnaire (ACQ)-5 score in patients with comorbid asthma and number of prior surgeries [39]. Between-group effect size estimates were calculated as per Hedges 1981 [39], and effect sizes (ES) were interpreted as small ($ES = 0.20$), moderate ($ES = 0.50$), and large ($ES = 0.80$) [38]. The statistical significance ($P \leq 0.05$) of differences in scores between groups was calculated using the F-test of one-way analysis of variances.

Responsiveness

Ability to detect change was determined by assessing change between baseline to Week 52 in individual and overall symptoms VAS and both composite VAS scores in patients believed to have experienced change (improvement or worsening) versus patients understood to be stable. Improved, stable and worsened scores, respectively, were defined as ≤ -1 , 0 , ≥ 1 -point changes for endoscopic NP score, ≥ 20 , < -20 – <20 , ≤ -20 L/min changes for PNIF, ≤ -2 , > -2 – <2 , ≥ 2 -point changes for overall VAS symptom score, ≤ -8.9 , -8.9 – <8.9 , ≥ 8.9 -point changes for SNOT-22 total score, and ≤ -2 , -2 – <2 , ≥ 2 -point changes for SNOT-22 domain scores (nasal obstruction, loss of taste or smell, thick nasal discharge, facial pain/pressure, post-nasal discharge). Mean change scores were compared within and between groups and interpreted as described above for known-groups validity assessments.

Interpretation of scores: meaningful within-patient improvements

Thresholds for meaningful within-patient change were derived to determine the proportion of responders for subsequent unblinded analyses. Anchor-based analyses were performed for individual and overall VAS symptoms and both composite VAS scores and the SNOT-22 total score using data from baseline to Week 52. Potential anchors included in this exploratory analysis were endoscopic NP score, PNIF, overall VAS and SNOT-22, none

of which have been established as verified anchors and no meaningful changes have been reported for overall VAS and endoscopic NP scores. Polyserial correlations were used to assess the relationship between potential anchors and change in VAS scores; anchors with a coefficient of ≥ 0.3 were selected [40]. Anchors that were deemed to have a sufficient relationship with the PROM scores were used to define patients as minimally improved or stable as described in Additional file 1: Table S3. Definition justifications are also provided in Additional file 1: Table S3 and were based on clinical insight, published literature and patient input. As recommended by the Food and Drug Administration, [41, 42] descriptive statistics within each anchor category were supplemented with cumulative distribution function (CDF) and probability density function (PDF) plots. These were split by anchor categories to compare estimated thresholds for meaningful within-patient changes in VAS scores and SNOT-22 scores from baseline to Week 52 in this very severe CRSwNP population. The use of group-level statistics to estimate within-individual change thresholds has been contested owing to individual patients varying in their own personal threshold for meaningful improvement; however, when the objective of an analysis is to use these thresholds to estimate the proportion of responders in a population, this estimated responder rate should still be valid [43]. In addition, it has been recommended that the statistical significance of individual-level change should inform use of within-patient thresholds; [44] this was assessed by the 95% coefficient of repeatability ($1.96 \cdot \sqrt{2} \cdot \text{standard error of measurement}$).

Response to treatment

Mepolizumab efficacy versus placebo on VAS and SNOT-22 was determined using the meaningful within-patient change thresholds for individual and overall symptoms VAS and for the SNOT-22 total score. Meaningful within-patient changes were analyzed using a logistic regression model with covariates of treatment group, geographic region, baseline score, and log(e) baseline blood eosinophil count. Additional analysis details are included in Additional file 1: Supplementary Methods.

Results

Patient population

Of the 407 patients included in SYNAPSE, 206 received mepolizumab and 201 received placebo. Overall, patients were mostly male (65%) with a mean (SD) age of 48.8 (13.0) years. The median baseline VAS and SNOT-22 total scores were similar between treatment groups (Table 1).

VAS psychometric analysis

Item characteristics

The individual symptom VAS scores were clustered at the high end of the scale at baseline, owing to the severity of symptoms present in the population. VAS responses were distributed across the full response scale at Weeks 20, 24 and 52. Ceiling effects were present at all time points for the loss of smell VAS, and a floor effect was observed for the facial pain or pressure VAS at Week 52. While no other VAS demonstrated substantial floor or ceiling effects, a higher proportion of responses were clustered at the upper versus lower end of the scale for all VAS scores at baseline (as was expected owing to the eligibility criteria), but not during the treatment period (Additional file 1: Figure S3).

All correlations between single VAS items were acceptable (> 0.3), with correlations > 0.9 observed only between two pairs of VAS: nasal obstruction VAS and nasal discharge VAS ($r = 0.929$), and nasal discharge VAS and mucus in throat VAS ($r = 0.900$). While a correlation of > 0.9 may show proof of redundancy, all items were included in the final model, as these item pairs include symptoms highly important to patients [21].

Domain structure

CFA using Week 20 data found that standardized loadings were ≥ 0.40 for the unidimensional nasal symptoms composite VAS score, and nasal symptoms and facial pain composite VAS score. Further details of model fit are included in Additional file 1: Supplementary Results. Residual correlations were not suggested (by modification indices > 10 , to capture a critical value of 10.83 corresponding to $P < 0.001$ [45]) between the pairs of items with high inter-item correlations but were suggested between nasal obstruction and loss of smell.

Reliability

Cronbach's α coefficients indicated acceptable internal consistency and reliability. For both nasal symptoms composite VAS score, and nasal symptoms and facial pain composite VAS scores, coefficients exceeded the predefined acceptable threshold of ≥ 0.70 at Week 20 (0.910 and 0.926) and Week 52 (0.904 and 0.926). To evaluate test-retest reliability, ICC coefficients between scores at Weeks 20 and 24 were assessed in a subset of participants defined as stable (Additional file 1: Table S1). All ICCs for individual symptom VAS, Overall VAS and both composite scores exceeded the prespecified threshold of > 0.75 , with the lower 95% confidence intervals (CIs) of all estimates > 0.90 (indicating 'excellent' reliability).

Table 1 Baseline demographics and clinical characteristics

	Placebo (n = 201)	Mepolizumab 100 mg SC (n = 206)	Total (n = 407)
Age, years, mean (SD)	48.9 (12.5)	48.6 (13.6)	48.8 (13.0)
Male, n (%)	125 (62)	139 (67)	264 (65)
Blood eosinophil count, cells/ μ L, geometric mean (standard logs)	400 (0.775)	390 (0.755)	–
Patients with asthma, n (%)	149 (74)	140 (68)	289 (71)
Patients with N-ERD, n (%)	63 (31)	45 (22)	108 (27)
Nasal surgeries in previous 10 years			
1	81 (40)	108 (52)	189 (46)
2	47 (23)	47 (23)	94 (23)
> 2	73 (36)	51 (25)	124 (30)
Duration of NP, years, mean (SD)	11.46 (8.27)	11.36 (8.52)	11.41 (8.39)
VAS score*, median (range)			
Nasal obstruction	9.1 (5.3–10.0)	9.0 (6.5–10.0)	9.1 (5.3–10.0)
Loss of sense of smell	10.0 (6.7–10.0)	10.0 (0.9–10.0)	10.0 (0.9–10.0)
Overall symptoms	9.2 (7.2–10.0)	9.1 (7.2–10.0)	9.2 (7.2–10.0)
Nasal discharge	9.0 (1.4–10.0)	8.9 (1.0–10.0)	–
Mucus in throat	9.1 (0.5–10.0)	8.9 (0.2–10.0)	–
Facial pain	8.9 (0.0–10.0)	8.5 (0.0–10.0)	–
SNOT-22 total score, median (range)*	64.0 (19–110)	64.0 (17–105)	64.0 (17–110)
Total endoscopic score, median (range) (scale: 0–8)	6.0 (0–8)	5.0 (2–8)	–
PNIF, median (range)	90.0 (0–380)	92.5 (0–350)	–
ACQ-5 score, mean (SD) [†]	2.2 (1.4)	2.4 (1.4)	–

ACQ-5 Asthma Control Questionnaire-5, N-ERD Non-steroidal anti-inflammatory drug exacerbated respiratory disease, NP Nasal polyp, PNIF peak nasal inspiratory flow, SC Subcutaneous, SD Standard deviation, SNOT-22 Sino-nasal outcome test-22, VAS Visual analogue scale

*Higher scores indicate greater disease severity or worse health-related quality of life

[†] In patients with asthma only, placebo n = 144, mepolizumab n = 138

Validity

Construct validity was acceptable between individual and composite VAS and the SNOT-22 total score ($r = 0.461$ – 0.598), and between individual symptom VAS and corresponding SNOT-22 items ($r = 0.560$ – 0.780), exceeding a priori hypothesized correlations (Additional file 1: Table S2). A priori hypothesized correlations between loss of smell VAS and UPSIT ($r = -0.494$), and other non-loss of smell VAS scores and UPSIT ($r = 0.211$ to -0.239), endoscopic NP ($r = 0.199$ – 0.279) and PNIF ($r = -0.216$ to -0.243) were not met. Hypothesized weak associations between VAS scores and WPAI work missed ($r = 0.129$ – 0.167) and SF-36 MCS ($r = -0.149$ to -0.216) were met, but associations between VAS scores and WPAI impairment-based scores ($r = 0.370$ – 0.553) and SF-36 PCS ($r = -0.308$ to -0.367 ; except loss of smell VAS) exceeded predictions (Additional file 1: Table S2).

Known-groups validity assessment demonstrated generally acceptable validity based on comorbid asthma, blood eosinophil count and N-ERD. Statistically significant between-group differences (F-test $P < 0.05$), with

small to large effect sizes (ES range 0.20 to 0.80), on several VAS were noted for ACQ-5 score and number of prior surgeries at Week 20 (Additional file 1: Table S4).

Responsiveness

In the ability to detect change analysis, effect sizes (-3.61 to -8.84) reported using all anchors indicated very large changes in the improved groups across all VAS and both composite scores. Changes in VAS scores were also large within the stable groups (ES -0.48 to -5.02), although always smaller than those in the improved groups.

Interpretation of scores

Anchors based on the SNOT-22 (polyserial correlation coefficient range: -0.436 to -0.599) and overall symptoms VAS (polyserial correlation coefficient range: -0.768 to -0.974) were sufficiently correlated ($r \geq 0.3$) with change in VAS scores and thus suitable for use. Mean anchor-based changes from baseline to Week 52 in VAS scores for patients categorized as stable or minimally improved (Table 2), CDF plots (Additional file 1: Figures S4–S10) and PDF plots informed meaningful

Table 2 Anchor-based change from baseline to Week 52 for individual VAS scores

Anchor	Nasal obstruction	Loss of sense of smell	Overall symptom	Nasal discharge	Mucus in throat	Facial pain
	Mean (95% CIs)					
<i>Overall VAS score*</i>						
Minimal improvement n = 49	-3.16 (-3.42, -2.90)	-1.11 (-1.58, -0.65)	-	-3.25 (-3.65, -2.86)	-3.25 (-3.77, -2.73)	-2.99 (-3.46, -2.52)
Stable n = 85	-0.45 (-0.63, -0.26)	-0.33 (-0.50, -0.16)	-	-0.55 (-0.87, -0.23)	-0.54 (-0.91, -0.16)	-0.49 (-0.88, -0.11)
<i>SNOT-22 total score†</i>						
Minimal improvement, n = 31	-2.80 (-3.70, -1.90)	-1.51 (-2.46, -0.57)	-2.94 (-3.92, -1.97)	-2.58 (-3.51, -1.65)	-2.33 (-3.37, -1.29)	-2.62 (-3.64, -1.59)
Stable, n = 38	-1.77 (-2.61, -0.93)	-1.05 (-1.83, -0.27)	-1.84 (-2.69, -0.99)	-1.80 (-2.64, -0.96)	-2.07 (-2.94, -1.20)	-1.58 (-2.50, -0.66)
<i>SNOT-22 nasal obstruction‡</i>						
Minimal improvement, n = 86	-4.66, (-5.23, -4.09)	-	-	-	-	-
Stable, n = 113	-2.62 (-3.12, -2.13)	-	-	-	-	-
<i>SNOT-22 loss of taste or smell‡</i>						
Minimal improvement, n = 43	-	-4.35 (-5.24, -3.46)	-	-	-	-
Stable, n = 187	-	-1.05 (-1.34, -0.75)	-	-	-	-
<i>SNOT-22 thick nasal discharge‡</i>						
Minimal improvement, n = 81	-	-	-	-5.08 (-5.76, -4.39)	-	-
Stable, n = 113	-	-	-	-2.75 (-3.27, -2.23)	-	-
<i>SNOT-22 post-nasal discharge‡</i>						
Minimal improvement, n = 75	-	-	-	-	-4.83 (-5.49, -4.17)	-
Stable, n = 137	-	-	-	-	-2.96 (-3.47, -2.44)	-
<i>SNOT-22 facial pain/pressure‡</i>						
Minimal improvement, n = 78	-	-	-	-	-	-4.82 (-5.51, -4.14)
Stable, n = 130	-	-	-	-	-	-2.66 (-3.18, -2.13)

CI, confidence interval, SNOT-22 Sino-nasal outcome test-22, VAS Visual analogue scale

*Minimal improvement - 2 ≥ change score > -4, stable -2 < change score < 2

† Minimal improvement - 8.9 ≥ change score > -17.8, stable -8.9 < change score < 8.9

‡ Minimal improvement - 2, stable -2 < change score < 2

within-patient change improvement thresholds of -2.5 points for the overall symptoms, nasal discharge, and facial pain VAS, and -3.0 points for the nasal obstruction, loss of sense of smell, and mucus in throat VAS. The absolute magnitude of these thresholds (threshold divided by SD) ranged from 0.76 to 1.14 in standardized units. The selection of these thresholds was largely driven by the CDF plots, where each was generally deemed sensitive enough to capture most minimally improved patients while specific enough to exclude the majority of stable patients across the different anchors. Estimates for

the 95% coefficient of repeatability ranged from 1.00 to 2.13, suggesting that all thresholds represent statistically significant within-individual changes [44].

SNOT-22 psychometric analysis

Item characteristics

Inter-item correlations between each possible pairing of the 22 individual items of the SNOT-22 were medium to high (r > 0.3) and there was no item redundancy (i.e. r ≥ 0.9).

Reliability and validity

Cronbach’s α was ≥ 0.70 at Weeks 20 (0.963) and 52 (0.961), further supporting acceptable internal consistency reliability of the SNOT-22 total score. The CFA of SNOT-22 data supported the six-domain model with second order symptoms and impact factors (Fig. 1). All standardized loadings were ≥ 0.40 , indicating that all SNOT-22 items were adequate indicators of their respective factors [26], and the validity of the SNOT-22 total

score was also supported due to the high correlation between the second order symptoms and impact factors ($r = 0.792$). Global fit statistics were acceptable for RMSEA (0.089) and SRMR (0.064) but not CFI (0.897) and Chi-square ($P < 0.001$).

Interpretation of scores

Although overall symptoms VAS is not a verified anchor, it was identified as the overall measure of stability and

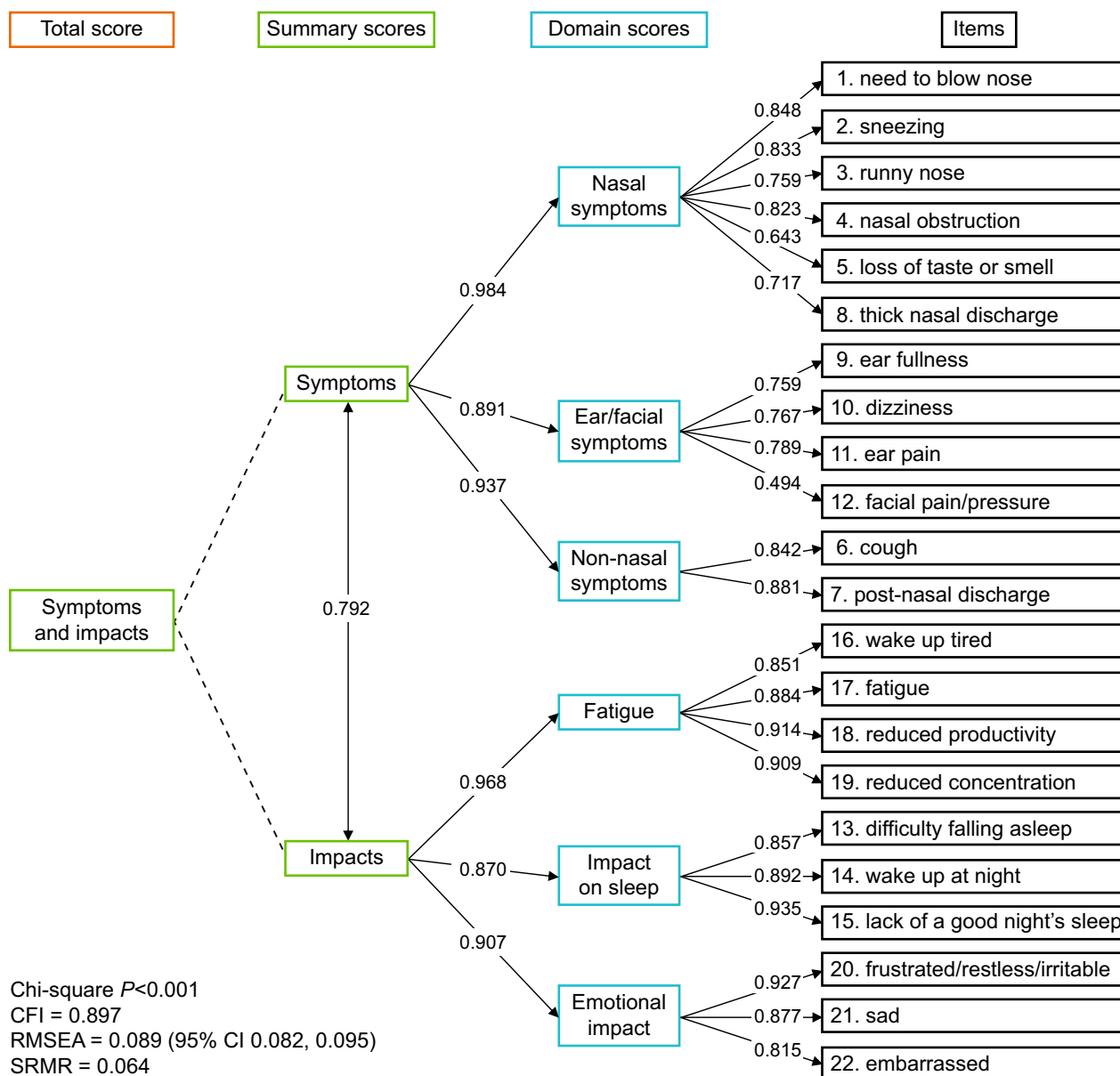


Fig. 1 The six-domain SNOT-22 model with standardized factor loadings. The CFA model, including the standardized factor loadings (numbers on arrows) from each hypothesized domain to the SNOT-22 items and Chi-squared, CFI, RMSEA, and SRMR fit statistics. Standardized factor loadings represent the correlation coefficients between factors. CFA Confirmatory factor analysis, CFI Comparative fit index, RMSEA Root mean square error of approximation, SNOT-22 Sino-nasal outcome test-22, SRMR Standardized root mean square residual

improvement in patients with severe CRSwNP, as this was the only anchor sufficiently correlated ($r \geq 0.3$) with SNOT-22 scores (polyserial correlation coefficient: -0.558). Using the overall symptoms VAS as an anchor to classify patients as stable or minimally improved (minimal improvement was arbitrarily defined as $-2 \geq \text{change score} \geq -4$, based on EPOS guidance for severity [1]), the mean (95% CI) within-patient change in SNOT-22 total

score (analyzed using observed data from 301 patients) for patients reporting minimal improvement was -28.52 ($-33.42, -23.62$) points (Fig. 2). CDF (Fig. 3) and PDF plots suggested that a -28 -point change was sensitive enough to capture most improved patients while being specific enough to exclude the majority of stable patients, within this very severe population. The absolute magnitude of this threshold (threshold divided by SD) was 1.25

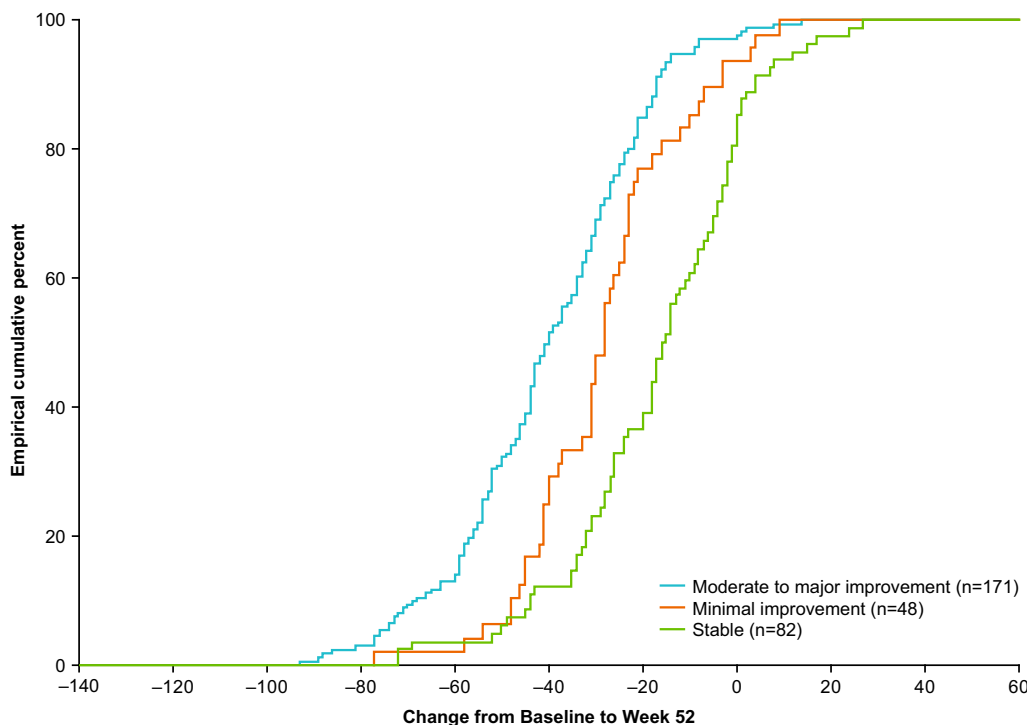


Fig. 2 Mean change in SNOT-22 score by VAS anchor in total ITT population. Overall VAS anchor groups were defined by change in overall VAS symptom score from baseline to Week 52: moderate to major improvement: ≤ -4 -point change in VAS score; minimal meaningful improvement: ≤ -2 to > -4 -point change in VAS score; limited improvement or worsening (stable): > -2 to < 2 -point change in VAS score; CDF Cumulative distribution function, CI Confidence interval, ITT Intent-to-treat, SNOT-22 Sino-nasal outcome test-22, VAS Visual analogue scale

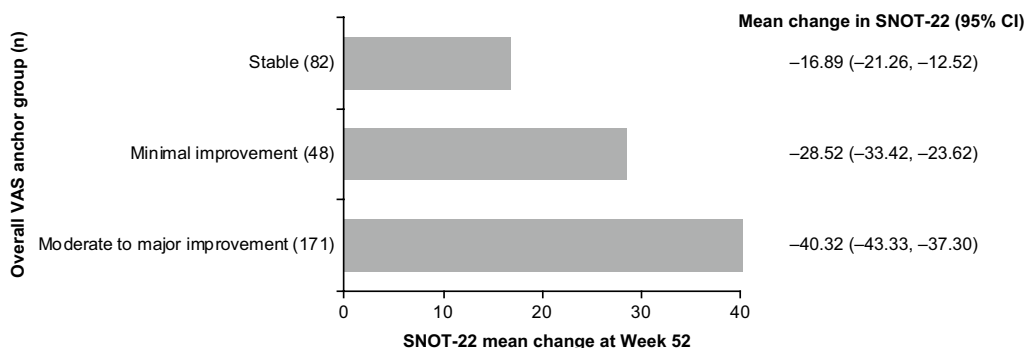


Fig. 3 CDF plot: change from baseline in SNOT-22 total score by overall VAS anchor. Overall VAS anchor groups were defined by change in overall VAS symptom score from baseline to Week 52: moderate to major improvement: ≤ -4 -point change in VAS score; minimal meaningful improvement: ≤ -2 to > -4 -point change in VAS score; limited improvement or worsening (stable): > -2 to < 2 -point change in VAS score. CDF Cumulative distribution function, SNOT-22 Sino-nasal outcome test-22, VAS Visual analogue scale

in standardized units. This 28-point threshold is close to 50% of the mean change from baseline in SNOT-22 total score observed in SYNAPSE. Therefore, 50% of the baseline score may be a plausible general threshold to use beyond this particular, very severe CRSwNP population. The 95% coefficient of repeatability was 11.89, suggesting a threshold of ≥ 28 points represents a statistically significant within-patient change [44].

VAS response to treatment

Compared with placebo, patients receiving mepolizumab had significantly increased odds (odds ratio [OR] 2.19–2.68) of achieving a meaningful improvement in individual VAS scores, based on meeting or exceeding the meaningful within-patient change thresholds at Weeks 49–52 (Fig. 4). Mepolizumab treatment was associated with a significantly greater change from baseline in

individual VAS scores compared with placebo at Weeks 49–52 ($P < 0.001$) (Fig. 5). Patients who had undergone ≥ 2 surgeries versus 1 surgery prior to mepolizumab treatment showed less improvement in median change from baseline in loss of sense of smell VAS (Fig. 6).

SNOT-22 response to treatment

A total of 54% and 32% of patients treated with mepolizumab and placebo achieved a ≥ 28 -point improvement in SNOT-22 total scores, respectively (Additional file 1: Table S5). Patients receiving mepolizumab had significantly increased odds (OR 1.61–2.96) of achieving a ≥ 28 -point improvement in SNOT-22 total score at each 4-weekly time points from Week 4–52 (Fig. 7). There were also significant improvements ($P < 0.001$) in the mean change from baseline in SNOT-22 total and

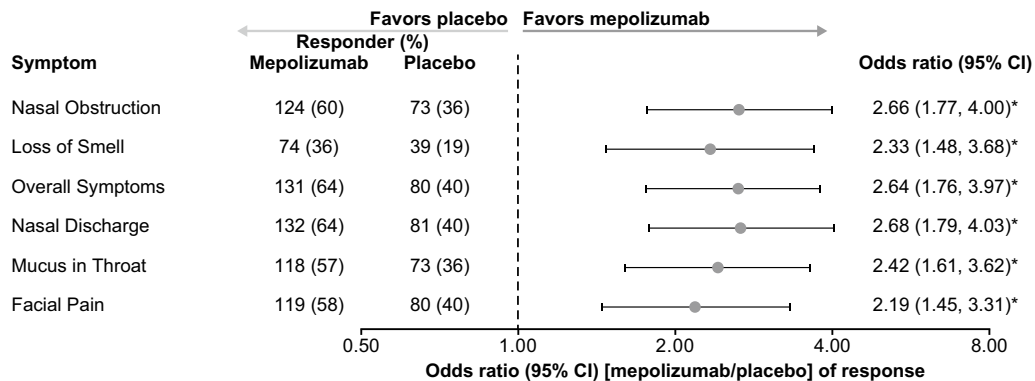


Fig. 4 Probability of meaningful within-patient change[†] in VAS score (Weeks 49–52). * $P < 0.001$; [†] ≥ 2.5 -point change (improvement) for overall symptoms, nasal discharge, and facial pain VAS scores, and a ≥ 3 -point change (improvement) for nasal obstruction, loss of sense of smell, and mucus in throat VAS scores. CI Confidence interval, VAS Visual analogue scale

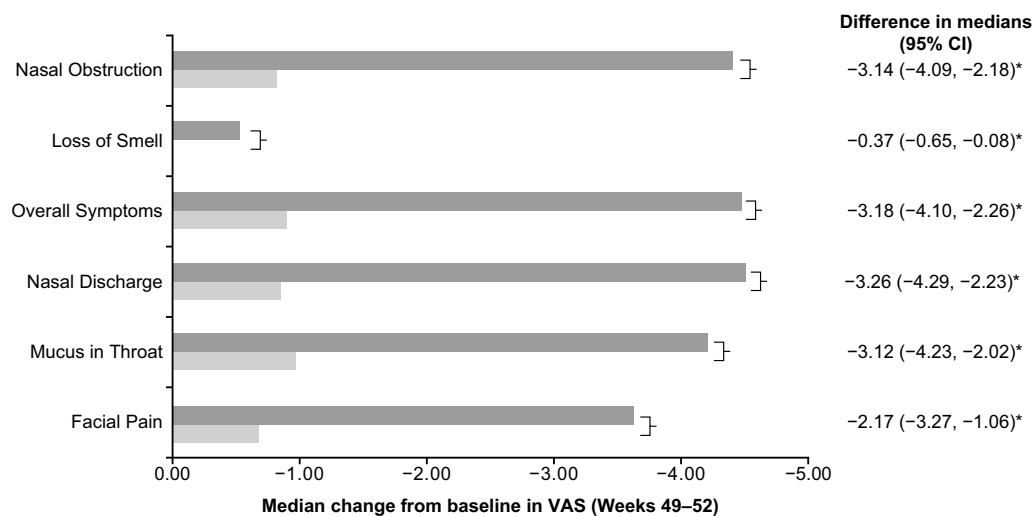


Fig. 5 Change from baseline at Weeks 49–52 in individual VAS scores. * $P < 0.001$. CI Confidence interval, VAS Visual analogue scale

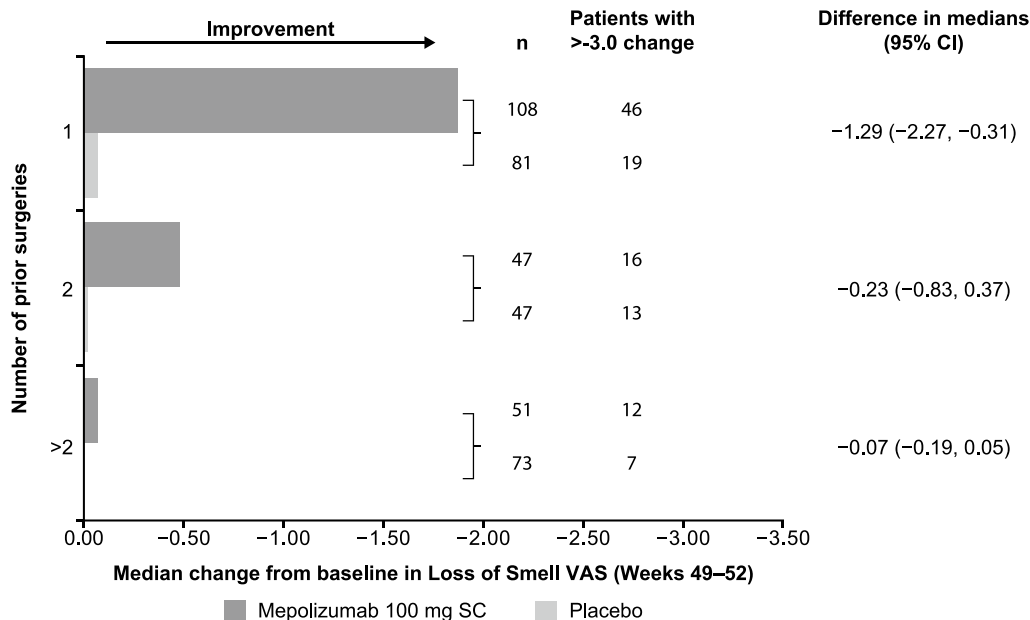


Fig. 6 Change from baseline in loss of sense of smell VAS score by surgery number. CI Confidence interval, SC Subcutaneous, VAS visual analogue scale

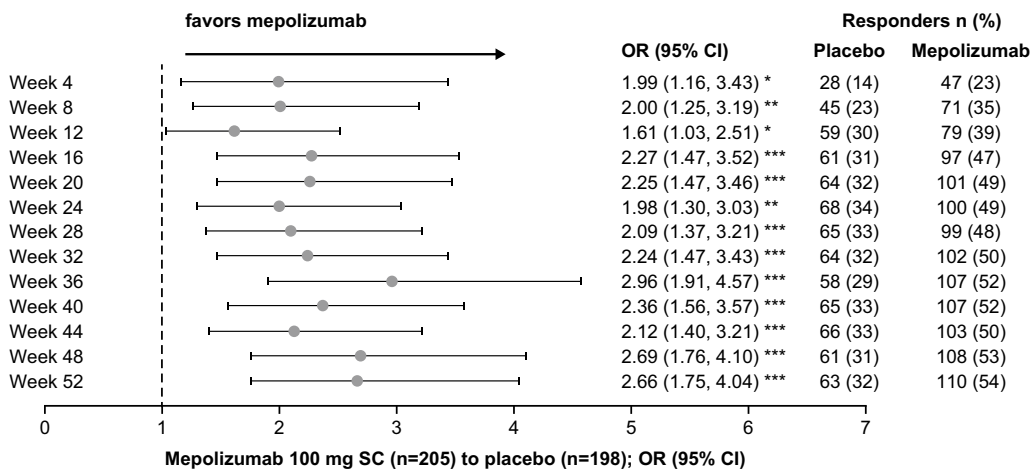


Fig. 7 Probability of patients reporting ≥ 28 -point improvement from baseline in SNOT-22 total score over time. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$. OR (mepolizumab vs placebo) of percentage of patients reporting ≥ 28 point improvement from baseline in SNOT-22 total score. OR > 1 indicates greater efficacy of mepolizumab. CI Confidence interval, OR Odds ratio, SC subcutaneous, SNOT-22 Sino-nasal outcome test-22

item scores following mepolizumab treatment compared with placebo at Week 52 (Fig. 8).

Discussion

Using data from SYNAPSE, this analysis evaluated the psychometric properties of the VAS and SNOT-22 in a blinded manner and determined the proportion of responders to mepolizumab using meaningful within-patient change thresholds for these measures,

relevant to this population with very severe CRSwNP. Findings indicate that the VAS and SNOT-22 perform well and have acceptable psychometric properties in this very severe CRSwNP population. They also indicate that mepolizumab provided meaningful within-patient improvements in HRQoL versus placebo when added to standard of care. This indicates mepolizumab provides substantial clinical benefits in very severe CRSwNP.

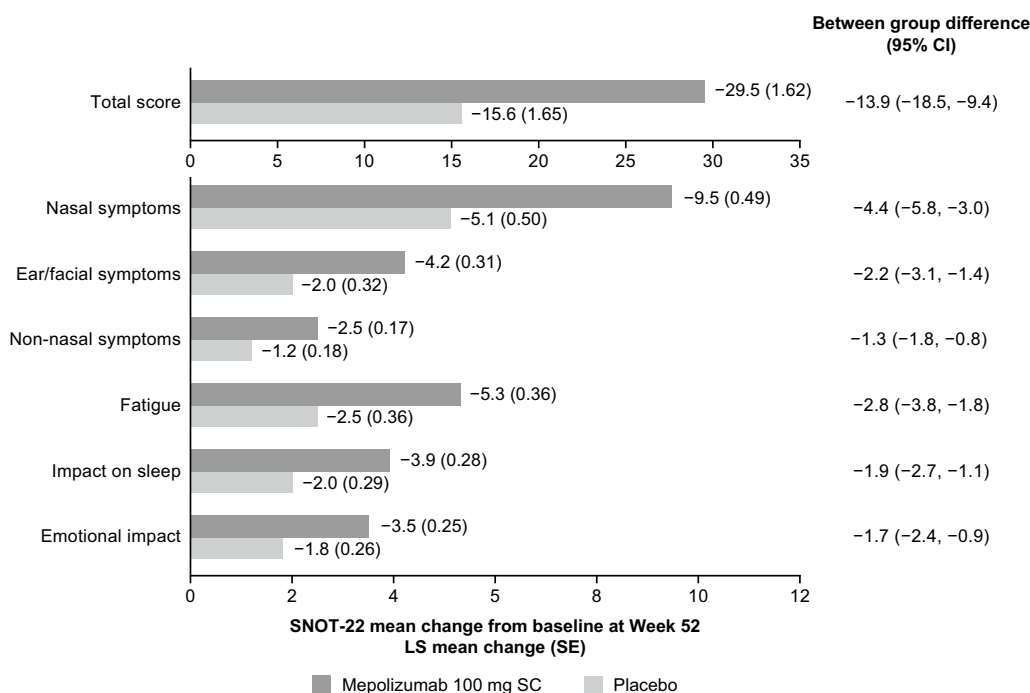


Fig. 8 Change from baseline in SNOT-22 total and domain scores at Week 52. Estimates are based on weighting applied to each level of class variable determined from observed proportions. *CI* confidence interval, *LS* Least squares, *SC* Subcutaneous, *SE* Standard error, *SNOT-22* Sino-nasal outcome test-22

VAS item responses were selected across the range at Weeks 20, 24, and 52. While a higher proportion of responses were clustered at the upper versus lower end of the scale for all symptom VAS scores at baseline, this was expected owing to the very severe CRSwNP SYNAPSE eligibility criteria. No substantial floor or ceiling effects were seen during the treatment period except for a floor effect for loss of sense of smell across all time points. This was likely a result of the CRSwNP and the history of repeat nasal surgery in SYNAPSE patients, which can lead to permanent sense of smell impairment [46, 47]. A floor effect was noted in the facial pain or pressure VAS at Week 52, likely because not all patients experience this symptom [21], and successful mepolizumab treatment will have potentially further decreased this proportion by the end of the study.

All correlations between single VAS items were acceptable, supporting the use of two VAS composite scores (symptoms composite, and nasal symptoms and facial pain composite). Factor structure of the VAS composite scores, as determined by CFA, was acceptable with the CFI and SRMR statistics. Although Chi-square and RMSEA model fit were initially not acceptable, model fit improved following the incorporation of conceptually justifiable residual correlations. Although the residual correlation suggested some redundancy in the VAS items, which can lead to overweighting in composite

scoring, these items were still included in the VAS composite scores based on their believed importance to patients. Regarding the Chi-square test, the improvement in model fit may have been due to the model's tendency to reject the null hypothesis in large samples, such as the 407 patients included in this analysis, even when the hypothesized model shows trivial misfit.

The VAS scores demonstrated acceptable internal consistency and excellent test-retest reliability. The VAS also demonstrated construct validity between corresponding VAS and SNOT-22 items, consistent with a priori hypothesized correlations. Correlations between the VAS and UPSIT, endoscopic NP and PNIF were weaker than a priori hypothesized correlations, which may be indicative of differences between objective or clinician-reported measures and daily patient reports of CRSwNP disease severity. In contrast, correlations between VAS scores and WPAI impairment-based scores and SF-36 PCS exceeded predictions.

The VAS demonstrated acceptable validity based on several known-groups assessments and detected large effect sizes for all VAS scores using all anchors in patients determined to have improved. Psychometric analyses indicated a meaningful within-patient change threshold of -2.5 points for overall symptoms, nasal discharge, and facial pain VAS and a threshold of -3.0 points for nasal obstruction, loss of sense of smell, and mucus in throat

VAS. This is consistent with previous studies indicating that a change of 2–3 points on a 0–10 VAS is considered a meaningful within-patient change [48, 49], suggesting similar thresholds are appropriate across the spectrum of disease severity as reported elsewhere [50]. Interestingly, while it has been previously assumed that anchor-based mean estimates will be lower than the 95% coefficient of repeatability (underestimating the amount of change needed to be statistically significant at the individual level) [44], our empirical mean estimates were higher. This is likely due to the high reliability estimates observed, resulting in a small degree of measurement error. We also note that the recommended thresholds (guided by CDF and PDF plots in addition to the mean changes) were primarily between the 95% coefficient of repeatability and the anchor-based mean changes.

Psychometric analyses of SYNAPSE confirm the validity of the SNOT-22 total score, as previously reported [14] and support a six-domain structure for reporting SNOT-22 results. Other studies have suggested four [51] or five [52] domains underlying the total score; however, the six-domain structure in this study was indicated by a previous analysis of patients with very severe CRSwNP in a clinical trial setting and thus may be most appropriate for analyzing SYNAPSE data [53]. These psychometric analyses also suggested that a ≥ 28 -point improvement ($\sim 50\%$ change from baseline) in SNOT-22 was an appropriate threshold to determine meaningful within-patient change within this very severe CRSwNP trial population. Different within-patient meaningful change thresholds may be more appropriate in less severe populations or clinical practice; this subject warrants future research. However, a 2010 study by Browne et al., which generated anchor-based values for within-patient meaningful changes for a number of commonly used PROMs suggested little association between baseline severity and within-patient meaningful change values as mathematical coupling can lead to an artificially inflated association between initial value and change score when correlation or regression is used [50]. Such a limitation should be considered when interpreting the results of the present study. In addition, Browne et al., advocated that an MCID should be calculated using a wide range of baseline severities and a single value applied across cohorts; [50] in the case of the SNOT-22 a value of 8.9 points might be appropriate regardless of baseline symptom burden.

The analyses of patient responses to treatment using the VAS and SNOT-22 demonstrated that patients with CRSwNP experienced significant clinical benefits in symptom severity and HRQoL with mepolizumab versus placebo. Furthermore, the effect of mepolizumab on loss of sense of smell measured by the VAS was greatest in patients who had undergone one versus multiple

prior surgeries. This may be related to the increased scarring and nerve damage associated with repeat surgeries, limiting the recovery of sense of smell [46, 47]. These results are consistent with results from SYNAPSE, which showed significant improvements in total endoscopic NP score, nasal obstruction VAS and SNOT-22 total score with mepolizumab versus placebo [14].

There were several limitations in the current analyses. Firstly, the psychometric analysis would have ideally used tailored ‘global impression’ anchor measures, specifically designed to evaluate meaningful change thresholds for the VAS and SNOT-22, rather than using the anchoring and triangulation of other outcome measures that were available at the time of this analysis. Use of an arbitrary overall symptoms VAS anchor may have overestimated changes in SNOT-22 score. Given this limitation, care was taken to justify the levels of change on each anchor considered a minimal improvement, where justification was based on clinical insight, published literature and patient input. However, we appreciate that our assumptions on what constitutes a minimal improvement are not guaranteed to precisely capture this level of change, and our chosen definitions may not be universally agreed upon. The suggested exploratory thresholds should therefore be confirmed in future research. Secondly, the same data were used to derive the meaningful within-patient thresholds and to determine the treatment response, which could be considered ‘overfitting’ the data. However, the psychometric analyses used data specific to the SYNAPSE trial in patients with very severe CRSwNP, therefore, the absolute values for respiratory threshold are likely not applicable in a less severe population. Furthermore, blinding the data to establish the thresholds mitigates the potential for overfitting. Thirdly, although the developer-recommended scoring of the SF-36v2 PCS and MCS has been contested, we did not explore alternate scoring options in detail [36, 54]. Finally, the analysis is limited by the SYNAPSE study population of patients with at least one previous nasal surgery [14]. As such, it is important to note that these analyses and their outcomes apply only to the very severe CRSwNP population included in SYNAPSE.

Conclusions

These psychometric analyses demonstrate that both the VAS and SNOT-22 have acceptable psychometric properties for outcome assessment in patients with very severe CRSwNP. Scores derived from both measures exhibited acceptable internal consistency and construct validity, in addition to test–retest reliability and known-groups validity (assessed only for the VAS). Overall, our analysis demonstrated that patients receiving mepolizumab were more likely to experience improvements

in symptom severity and HRQoL versus placebo, when added to standard of care.

Abbreviations

ACQ	Asthma Control Questionnaire
AERD	Aspirin-exacerbated respiratory disease
ANOVA	Analysis of variance
CDF	Cumulative distribution function
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Confidence interval
CRSwNP	Chronic rhinosinusitis with nasal polyps
EPOS	European Position Paper on Rhinosinusitis and Nasal Polyposis
ES	Effect size
HRQoL	Health-related quality of life
ICC	Intra-class correlation
IL	Interleukin
ITT	Intent-to-treat
LS	Least squares
MCID	Minimal clinically important difference
MCS	Mental component summary
N-ERD	Non-steroidal anti-inflammatory drug-exacerbated respiratory disease
NP	Nasal polyps
OR	Odds ratio
PCS	Physical component summary
PDF	Probability density function
PNIF	Peak nasal inspiratory flow
PROMs	Patient-reported outcome measures
RMSEA	Root mean square error of approximation
SC	Subcutaneously
SCS	Systemic corticosteroids
SD	Standard deviation
SE	Standard error
SEM	Standard error of measurement
SF-36	36-Item Short Form Health Survey
SNOT	Sino-nasal outcome test
SoC	Standard of care
SRMR	Standardized root mean residual
UPSIT	University of Pennsylvania Smell Identification Test
VAS	Visual analogue scale
WPAl	Work productivity and impairment

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41687-023-00543-5>.

Additional file 1. Supplementary methods and results.

Acknowledgements

We thank the patients who took part in the SYNAPSE study. Editorial support (in the form of writing assistance, including preparation of the draft manuscript under the direction and guidance of the authors, collating, and incorporating authors' comments for each draft, assembling tables and figures, grammatical editing and referencing) was provided by Rachael Baylie PhD, of Fishawack Indicia Ltd, UK, part of Fishawack Health, and funded by GSK. The authors would like to thank the principal investigators and patients of the SYNAPSE study. Participating SYNAPSE investigators: *Argentina*: Ledit Arduoso, Miguel Bergna, María De Salvo, Pedro Elias, Gabriel García, Jorge Maspero, Ramón Rojas, Pablo Saez Scherbovsky, Alberto Tolcachier, Luis Wehbe, Anahí Yañez; *Australia*: Philip Bardin, Sara Barnes, Andrew Gillman, Richard Harvey, Chady Sader, Narinder Singh; *Canada*: Jaime Del Carpio, Marie-Noëlle Corriveau, Martin Desrosiers, Arif Janjua, Shaun Kilty, Doron Sommer, Leigh Sowerby, Peter Spafford; *Germany*: Christian Betz, Achim Beule, Adam Chaker, Mandy Cuevas, Moritz Groeger, Ludger Klimek, Heidi Olze, Carolina van Schaik, Martin Wagenmann, Barbara Wollenberg, Yury Yarin; *Republic of Korea*: Hyung-Ju Cho, Hun-Jong Dhong, Chang-Hoon Kim, Seontae Kim, Chae-Seo Rhee, Soo

Whan Kim, Hyo Yeol Kim; *the Netherlands*: Wytse J Fokkens; *Romania*: Valeriu Bronescu, Corina Mella, Adriana Neagos, Doinel Radeanu, Catalin Stefan; *Russian Federation*: Anton Edin, Sergey Karpischenko, Fatimat Khanova, Ekaterina Mirzabekyan, Andrey Ovchinnikov, Dmitriy Polyakov, Sergei Ryazantsev, Valeriy Svistushkin, Galina Tarasova, Vladimir Yakusevich; *Sweden*: Cecilia Ahlström Emanuelsson, Johan Hellgren, Mattias Jangard, Anders Mårtensson, Karin Toll; *UK*: Sean Carrie, Stephen Durham, Simon Gane, Jonathan Hobson, Claire Hopkins, Naveed Kara, Samuel Leong, Neil Massey, Guy Scadding; *USA*: Michael Armstrong, James Blotter, Matthew Brown, Timothy Courville, Cecelia Damask, Adam DeConde, Dale Ehmer Jr, Adil Fataki, Christine Franzese, Joseph Han, Thomas Higgins, Edward Kerwin, Craig LaForce, Stella Lee, Bradley Marple, Jonathan Matz, Chad McDuffie, Steven Miller, Jonathan Moss, Nayla Mumneh, Robert Nathan, Randall Ow, Jeffrey Rosenbloom, Rodney Schlosser, Heena Shah-Patel, Ronald Shealy, Ayesha Siddiqi, Stacey Silvers, Weily Soong, Richard Sterling, Neetu Talreja, Martha Tarpay, Luke Webb, H James Wedner, Simon Wright, David Yen.

Author contributions

WF contributed to the acquisition of study data, and data analysis and interpretation. RHC, BM and ARS contributed to the conception and design of the study, acquisition of study data, and data analysis and interpretation. SEL, ZD and CH contributed to the acquisition of study data. PH, SY and MT contributed to data analysis and interpretation. AT was responsible for developing the psychometric analysis plan and contributed to data analysis and interpretation. VL contributed to the conception and design of the study. All authors were involved in development of all stages of the manuscript, revising it critically for important intellectual content. All authors provided final approval of the version submitted for publication. All authors are accountable for the accuracy and integrity of the work. All authors read and approved the final manuscript.

Funding

This analysis and the parent study were funded by GlaxoSmithKline (GSK ID: 205687; NCT03085797).

Availability of data and materials

Anonymized individual participant data and study documents for the parent study can be requested for further research from www.clinicalstudydataquest.com.

Declarations

Ethics approval and consent to participate

All patients provided written informed consent prior to participation in the SYNAPSE study. The study was approved by the local institutional review board or ethics committee at each study center. The approving ethics committees were as follows: *Argentina*: Comite de Etica Dr. Claude Bernard, San Lorenzo 2555, Rosario, Santa Fe, S2000KPG; Comite de etica en Investigacion CEMER, Esmeralda 1550, Florida, Buenos Aires, B1602DQD; Comité Independiente de ética para ensayos en Farmacología Clínica del centro Medico Dra. De Salvo, Avenida Cabildo 1536 5° B, Ciudad Autonoma de Buenos Aires, Buenos Aires, C1426ABP; Comité de Ética en Investigación. INAER, Arenales 3146 1°A, Ciudad Autonoma de Buenos Aires, Buenos Aires, C1425BEN; Framingham Centro Médico, Calle 9 Numero 431, La Plata, Buenos Aires, B1902COS; Comite de Etica en Farmacologia Clinica de la Fundacion CIDEA, Paraguay 2041, Ciudad Autonoma de Buenos Aires, CP1121ABE; Investigaciones en Alergia y Enfermedades Respiratorias, Arenales 3146, Buenos Aires, C1425BEN; Comite de Etica en Investigacion, Instituto Ave Pulmo, Carlos M. Alvear 3345, Mar del Plata, Buenos Aires, B7602DCK; *Australia*: Saint Vincent's Hospital, Research Office, Level 6, DeLacy Building, Darlinghurst, New South Wales, 2010; Bellberry Limited, 123 Glen Osmond Road, Eastwood, South Australia, 5063; *Canada*: Comité d'éthique de la recherche du CHU de Québec, A0-124, Comité d'éthique de la recherche du CHU de Québec, Québec, Québec, G1L 3L5; Advarra, Suite 300, 372 Hollandview Trail, Aurora, Ontario, L4G 0A5; Comité d'éthique de la Recherche du CHUM, Tour Viger, 3e étage, 900, rue St-Denis, Montréal, Québec, H2X 0A9; UBC CREB Office, 210, 828 West 10th Avenue, Vancouver, British Columbia, V5Z 1L8; Hamilton Integrated Research Ethics Board, Suite 102, 293 Wellington Street North, Hamilton, Ontario, L8L 8E7; Western University, Office of Research Ethics, Support Services Building Room 5150, 1151 Richmond Street, London, Ontario, N6G 1G9; University

of Saskatchewan, Thorvaldson Building, Room 223, Research Services and Ethics Office, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5C9; **Germany:** Ethik-Kommission der Landesärztekammer Hessen, Im Vogelsge-sang 3, Frankfurt, Hessen, 60488; **Republic of Korea:** Severance Hospital Clinic Trial Centre, Yonsei University Health System, 50-1 Yonsei-Ro Seodaemun-Gu, Seoul, 03722; Samsung Medical Centre, 81 Irwon-Ro Gangnam-gu, Seoul, 06351; Gachon University of Gil Medical Centre, 21, Namdong-daero 774 bean-gil, Namdong-gu, 21565; The Catholic University of Korea Seoul Saint Mary's Hospital, 222 Banpo-Daero, Seocho-gu, Seoul, 06591; Korea University Guro Hospital, 148, Gurodong-ro, Guro-gu, Seoul, 08308; Seoul National University Bundang Hospital, 82 Gumi-ro 173 beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, 13620; **Netherlands:** Medisch Ethische Commissie, Meibergdreef 9, Amsterdam, 1105 AZ; **Romania:** Comisia Nationala de Bioetica a medicamentelor si a Dispozitivelor Medicale, Pavilion K, Spitalul Clinic Colentina, Sos. Stefan cel Mare near 19-21, Bucuresti, 020125; **Russia:** Independent Ethics Committee PharmNadzor, office 618, 4, Komendantskiy Prospect, Lit.A, Saint-Petersburg, 197227; GOU VPO Saint.Petersburg State Medical University "I.P.Pavlova", 6\8,Lva Tolstogo, Saint Petersburg, 197022; Federal medical biophysical center n.a. A.I. Burnazyan of FMBA of Russia, 23, Marshala Novikova Strasse., Moscow, 123098; Hospital of Russian Academy of Science, Troitsk city, Oktyabrskiy prospect, 3, Moscow, 142190; Inter-Institutional ETHICS COM-MITTEE, per. Gagarinskiy, d. 37, Strasse. 2, Moscow, 119002; Scientific Clinical Center of Otorhinolaryngology, 30/2, Volokolamskoe shosse, Moscow, 123182; ENT Research Institute, 9, Bronnitskaya Street, Saint-Petersburg, 190013; I.M. Sechenov First Moscow State Medical University, Local Ethics Committee of Federal State Autonomous Educational Institution of Higher Education, 8, ul. Trubetskaya, Moscow, 119991; I.M.Sechenov First Moscow State Medical University of Ministry of health and social development, Local Ethics Committee of State budget educational institution of higher professional education, 8 Trubetskaya street, Moscow; Municipal Health Care Institution Clinical Hospital for Emergency Medical Care, 11 Zagorodny Sad., Yaroslavl, 150003; **Sweden:** Regionala Etikprövningsnämnden i Göteborg, Guldhedsgatan 5A, Göteborg, SE-413 20; **United Kingdom:** NRES Committee East Midlands-Leicester, The Old Chapel, Royal Standard Place, Nottingham, NG1 6FS; **United States of America:** University of Utah Institutional Review Board, Research Administration Building, 75 South 2000 East, Salt Lake City, Utah, 84112; Advarra Institutional Review Board, Suite 110, 6940 Columbia Gateway Drive, Columbia, Maryland, 21046; Penn State college of Medicine Institutional Review Board/Human Subjects Protection Office, Penn State Hershey Medical Center, Mailcode A115, 90 Hope Drive, Hershey, Pennsylvania, 17033, 855; University of California, San Diego (UCSD) Human Research Protection Program Office (HRPP), 9452 Medical Center drive, La Jolla, California, 92093; University of Texas Southwestern Medical Center Institutional Review Board, Room BL9.100, 5323 Harry Hines Boulevard, Dallas, Texas, 75390-8843; Office of Research Integrity, Institutional Review for Human Research, 19 Hagwood Avenue, Suite 601, Charleston, South California, 29425.

Consent for publication

Not applicable.

Competing interests

WF has received clinical trial funding from Sanofi, Mylan, ALK, Allergy Therapeutics, Novartis, and Chordate, personal fees from GSK, Sanofi and Novartis, has participated on an advisory board for Lyra, and is Secretary General of the European Rhinologic Society. AT was employed by Adelphi Values during the conduct of the study and is currently an employee of Bayer Plc. SEL has participated in advisory boards and received clinical trial funding from Sanofi Genzyme, Regeneron, Genentech, AstraZeneca, and GSK. RHC, PH, BM, ARS and SY are all employees of GSK and own stocks/shares in GSK. MT was an employee of GSK at the time of the study and owns stocks/shares in GSK. ZD has acted as Research Director at QPS-NL, an institution which has received research support from HAL Allergy, Foresee Pharmaceuticals, Patara Pharma (now Respiant) and Novartis. ZD has also received honoraria or speaker fees serving on advisory boards or as a consultant for ALK, Antibio, AstraZeneca, Boehringer Ingelheim, Merck Sharp & Dohme, QPS-NL and Sanofi-Genzyme-Regeneron. CH has received advisory board fees from Sanofi, AstraZeneca, Olympus, Dianosis, and GSK, and has received honoraria for lectures from GSK, Sanofi, Intersect, and Mylan. VL reports advisory board and lecture fees from GSK, Novartis, and Sanofi.

Author details

¹Department of Otolaryngology, University of Amsterdam, Amsterdam, The Netherlands. ²Patient-Centred Outcomes, Adelphi Values, Bollington, Cheshire, UK. ³Division of Otolaryngology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ⁴Clinical Sciences, Respiratory, GSK, GSK House, Brentford, Middlesex, UK. ⁵Department of Microbiology Immunology and Transplantation, KU Leuven, Catholic University of Leuven, Leuven, Belgium. ⁶Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, Groningen, The Netherlands. ⁷Respiratory Medicine and Allergology, Skane University, Lund, Sweden. ⁸ENT Department, Guys and St Thomas's Hospital, King's College, London, UK. ⁹Global Medical Affairs, GSK, Brentford, Middlesex, UK. ¹⁰Royal National Throat, Nose and Ear Hospital, University College London Hospitals, London, UK. ¹¹Clinical Statistics, GSK, GSK House, Brentford, Middlesex, UK. ¹²Respiratory Therapeutic Area Unit, GSK, Research Triangle Park, NC, USA. ¹³Respiratory Patient Centered Outcomes, Value Evidence and Outcomes GSK, GSK House, Brentford, Middlesex, UK.

Received: 4 January 2022 Accepted: 4 January 2023

Published online: 20 January 2023

References

- Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, Reitsma S et al (2020) European position paper on rhinosinusitis and nasal polyps 2020. *Rhinology* 58(Suppl 29):1–464
- Schleimer RP (2017) Immunopathogenesis of chronic rhinosinusitis and nasal polyposis. *Annu Rev Pathol* 12:331–357
- Stevens WW, Schleimer RP, Kern RC (2016) Chronic rhinosinusitis with nasal polyps. *J Allergy Clin Immunol Pract* 4(4):565–572
- Bachert C, Wagenmann M, Hauser U, Rudack C (1997) IL-5 synthesis is upregulated in human nasal polyp tissue. *J Allergy Clin Immunol* 99(6 Pt 1):837–842
- Bachert C, Zhang N, Holtappels G, De Lobel L, van Cauwenberge P, Liu S et al (2010) Presence of IL-5 protein and IgE antibodies to staphylococcal enterotoxins in nasal polyps is associated with comorbid asthma. *J Allergy Clin Immunol* 126(5):962–968
- Alobid I, Cardelus S, Benítez P, Guilemany JM, Roca-Ferrer J, Picado C et al (2011) Persistent asthma has an accumulative impact on the loss of smell in patients with nasal polyposis. *Rhinology* 49(5):519–524
- DeConde AS, Mace JC, Levy JM, Rudmik L, Alt JA, Smith TL (2017) Prevalence of polyp recurrence after endoscopic sinus surgery for chronic rhinosinusitis with nasal polyposis. *Laryngoscope* 127(3):550–555
- Hox V, Lourijzen E, Jordens A, Aasbjerg K, Agache I, Alobid I et al (2020) Benefits and harm of systemic steroids for short- and long-term use in rhinitis and rhinosinusitis: an EAACI position paper. *Clin Transl Allergy* 10:1
- GSK. Nucala (mepolizumab) Highlights of prescribing information https://gskpro.com/content/dam/global/hcpportal/en_US/Prescribing_Information/Nucala/pdf/NUCALA-PI-PIL-IFU-COMBINED.PDF. Accessed Aug 2021
- EMA. Nucala (mepolizumab) Summary of product characteristics. https://www.ema.europa.eu/en/documents/product-information/nucala-epar-product-information_en.pdf. Accessed Nov 2021
- GSK. European Commission approves Nucala (mepolizumab) in three additional eosinophil-driven diseases. <https://www.gsk.com/en-gb/media/press-releases/european-commission-approves-nucala-mepolizumab-in-three-additional-eosinophil-driven-diseases/>. Accessed Nov 2021
- Bachert C, Sousa AR, Lund VJ, Scadding GK, Gevaert P, Nasser S et al (2017) Reduced need for surgery in severe nasal polyposis with mepolizumab: randomized trial. *J Allergy Clin Immunol* 140(4):1024–31.e14
- Gevaert P, Van Bruaene N, Cattaert T, Van Steen K, Van Zele T, Acke F et al (2011) Mepolizumab, a humanized anti-IL-5 mAb, as a treatment option for severe nasal polyposis. *J Allergy Clin Immunol* 128(5):989–995
- Han JK, Bachert C, Fokkens W, Desrosiers M, Wagenmann M, Lee SE et al (2021) Mepolizumab for chronic rhinosinusitis with nasal polyps (SYNAPSE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Respir Med* 9(10):1141–1153
- Khan A, Vandeplas G, Huynh TMT, Joish VN, Mannent L, Tomassen P et al (2019) The global allergy and asthma European network (GALEN) rhinosinusitis cohort: a large European cross-sectional study of chronic

- rhinosinusitis patients with and without nasal polyps. *Rhinology* 57(1):32–42
16. Kim DH, Han K, Kim SW (2016) Effect of chronic rhinosinusitis with or without nasal polyp on quality of life in South Korea: 5th Korea national health and nutrition examination survey Korean. *Clin Exp Otorhinolaryngol* 9(2):150–156
 17. Hall R, Trennery C, Bradley H, Gater A, Sikirica M, Nelsen L et al (2019) PRS59 qualitative assessment of the content validity of the SNOT-22 in patients with nasal polyps. *Value Health* 22:S360
 18. Tabberer M, Trigg A, Busse W, Lund V, Lee J, Bachert C et al (2021) Mepolizumab reduces disease symptoms for patients with chronic rhinosinusitis with nasal polyps: data from the SYNAPSE study. *J Allergy Clin Immunol* 147(2):AB126
 19. Chowdhury NI, Mace JC, Bodner TE, Alt JA, Deconde AS, Levy JM et al (2017) Investigating the minimal clinically important difference for SNOT-22 symptom domains in surgically managed chronic rhinosinusitis. *Int Forum Allergy Rhinol* 7(12):1149–1155
 20. Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP (2009) Psychometric validity of the 22-item sinonasal outcome test. *Clin Otolaryngol: Off J ENT-UK, Off J Neth Soc Oto-Rhino-Laryngol Cervico-Fac Surg* 34(5):447–54
 21. Hall R, Trennery C, Chan R, Gater A, Bradley H, Sikirica MV et al (2020) Understanding the patient experience of severe, recurrent, bilateral nasal polyps: a qualitative interview study in the United States and Germany. *Value Health: J Int Soc Pharmacoecon Outcomes Res* 23(5):632–641
 22. Juniper EF, Bousquet J, Abetz L, Bateman ED, Committee G (2006) Identifying “well-controlled” and “not well-controlled” asthma using the Asthma Control Questionnaire. *Respir Med* 100(4):616–621
 23. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
 24. Kline R (2015) Principles and practice of structural equation modeling. Fourth Edition ed: Guilford Publications
 25. Jöreskog KG (1967) A General approach to confirmatory maximum likelihood factor analysis. *ETS Res Bull Ser* 1967(2):183–202
 26. Brown T (2015) Confirmatory factor analysis for applied research. Second Edition ed: Guilford Publications
 27. Lt Hu, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Model: Multidiscip J* 6(1):1–55
 28. Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66(4):507–514
 29. Nunnally J. (2010) Psychometric theory. Third Edition ed: Tata McGraw-Hill Education
 30. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155–163
 31. Doty RL, Shaman P, Kimmelman CP, Dann MS (1984) University of Pennsylvania smell identification test: a rapid quantitative olfactory function test for the clinic. *Laryngoscope* 94(2 Pt 1):176–178
 32. Lund VJ, Mackay IS (1993) Staging in rhinosinusitis. *Rhinology* 31(4):183–184
 33. Timperley D, Srubisky A, Stow N, Marcellis GN, Harvey RJ (2011) Minimal clinically important differences in nasal peak inspiratory flow. *Rhinology* 49(1):37–40
 34. Browne JP, Hopkins C, Slack R, Topham J, Reeves B, Lund V et al (2006) Health-related quality of life after polypectomy with and without additional surgery. *Laryngoscope* 116(2):297–302
 35. Buckland JR, Thomas S, Harries PG (2003) Can the sino-nasal outcome test (SNOT-22) be used as a reliable outcome measure for successful septal surgery? *Clin Otolaryngol Allied Sci* 28(1):43–47
 36. Maruish M, Kosinski M, Bjorner JB, Gandek B, Turner-Bowker DM, Ware JE (2011) User’s manual for the SF36v2 health survey. Quality Metric Incorporated. 3rd ed
 37. Reilly MC, Zbrozek AS, Dukes EM (1993) The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 4(5):353–365
 38. Cohen J (2013) Statistical power analysis for the behavioral sciences, Taylor & Francis
 39. Hedges LV (1981) Distribution theory for Glass’s estimator of effect size and related estimators. *J Educ Stat* 6(2):107–128
 40. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61(2):102–109
 41. Fda U (2018) Patient-focused drug development guidance public workshop: methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments [Workshop]. <https://www.fda.gov/media/116277/download>
 42. FDA (2019) Patient-Focused Drug Development Guidance Public Workshop: Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. <https://www.fda.gov/media/132505/download>. Accessed Feb 2021
 43. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D et al (2021) Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Qual Life Res: Int J Qual Life Aspect Treat Care Rehabil* 30(10):2729–2754
 44. Hays RD, Peipert JD (2021) Between-group minimally important change versus individual treatment responders. *Qual Life Res: Int J Qual Life Asp Treat Care Rehabil* 30(10):2765–2772
 45. Muthén LK, Muthén BO (2017) Mplus user’s guide, 8th edn. Muthén & Muthén, Los Angeles
 46. Nguyen DT, Bey A, Arous F, Nguyen-Thi PL, Felix-Ravelo M, Jankowski R (2015) Can surgeons predict the olfactory outcomes after endoscopic surgery for nasal polyposis? *Laryngoscope* 125(7):1535–1540
 47. Seys SF, De Bont S, Fokkens WJ, Bachert C, Alobid I, Bernal-Sprekelsen M et al (2020) Real-life assessment of chronic rhinosinusitis patients using mobile technology: the mySinusitisCoach project by EUFOREA. *Allergy* 75(11):2867–2878
 48. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL (2000) Defining the clinically important difference in pain outcome measures. *Pain* 88(3):287–294
 49. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole MR (2001) Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 94(2):149–158
 50. Browne J, van der Meulen J, Lewsey J, Lamping D, Black N (2010) Mathematical coupling may account for the association between baseline severity and minimally important difference values. *J Clin Epidemiol* 63:865–874
 51. DeJaco D, Riedl D, Huber A, Moschen R, Giotakis AI, Bektic-Tadic L et al (2019) The SNOT-22 factorial structure in European patients with chronic rhinosinusitis: new clinical insights. *Eur Arch Otorhinolaryngol* 276(5):1355–1365
 52. Lidder AK, Detwiller KY, Price CP, Kern RC, Conley DB, Shintani-Smith S et al (2017) Evaluating metrics of responsiveness using patient-reported outcome measures in chronic rhinosinusitis. *Int Forum Allergy Rhinol* 7(2):128–134
 53. Lee S, Tabberer M, Trigg A, Han J, Fokkens W, Naclerio R et al (2021) Mepolizumab improves health related quality of life for patients with chronic rhinosinusitis with nasal polyps: data from the SYNAPSE study. *J Allergy Clin Immunol* 147(2):AB125
 54. Tucker G, Adams R, Wilson D (2013) Observed agreement problems between sub-scales and summary components of the SF-36 version 2: an alternative scoring method can correct the problem. *PLoS ONE* 8(4):e61191

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.