

Review

Not peer-reviewed version

From Subjective to Objective: Validating Patient Satisfaction in Facial Surgery Through Psychometrics

[Maarten J. Ottenhof](#)*

Posted Date: 13 January 2026

doi: 10.20944/preprints202601.0861.v1

Keywords: patient-reported outcome measures; FACE-Q; facial aesthetic surgery; quality of life; computerized adaptive testing; psychometrics; value-based healthcare



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

From Subjective to Objective: Validating Patient Satisfaction in Facial Surgery Through Psychometrics

Maarten J. Ottenhof^{1,2,*}

¹ Clinique Rebelle, Amsterdam 1071 DJ, The Netherlands

² Maastricht University, Maastricht, The Netherlands

* Correspondence: maartenottenhof@gmail.com

Abstract

Patient satisfaction is crucial to aesthetic surgery, yet measuring how well outcomes meet patient expectations has always been challenging. Rather than relying on the surgeon's impression, we've synthesized research on Patient-Reported Outcome Measures (PROMs) in facial aesthetics. Our work zeroes in on the FACE-Q instrument and explores newer technological applications. We conducted a comprehensive literature review of studies on facelifts (563 patients across 10 studies), injectable treatments (2292 patients in 23 studies), and rhinoplasty (937 patients across 10 studies). Our original data came from a Dutch cohort of Clinique Rebelle in Amsterdam—259 patients undergoing facial procedures, supplemented by Computerized Adaptive Testing (CAT) simulation research. The FACE-Q scales demonstrated strong psychometric properties—Cronbach's alpha between 0.885 and 0.951—and successfully captured differences between patients that traditional photos miss. CAT methods reduced questionnaire length by roughly 71% without sacrificing measurement accuracy ($r = 0.98$ with complete surveys). Looking ahead, machine learning shows real potential for forecasting patient satisfaction outcomes. Implementing routine PROM collection in aesthetic practice makes sense on multiple fronts: better patient selection, benchmarking quality across surgeons, protecting against medicolegal concerns, and aligning with value-based healthcare models. We also discuss how AI and 3D imaging might reshape outcome assessment going forward.

Keywords: patient-reported outcome measures; FACE-Q; facial aesthetic surgery; quality of life; computerized adaptive testing; psychometrics; value-based healthcare

1. Introduction

What makes an aesthetic procedure successful? The obvious answer—patient satisfaction—turns out to be surprisingly hard to pin down. Complication rates tell us something. Symmetry measurements tell us something else. Photos document change. But none of this captures whether the person looking in the mirror actually feels better about what they see [1]. That gap between technical achievement and lived experience is both a clinical problem and, if we're honest, a measurement problem we've been slow to address [2].

Healthcare more broadly has been moving toward patient-centred models for decades now [3]. The shift to value-based care—paying for outcomes rather than procedures—makes systematic outcome measurement less optional than it once was [4]. For aesthetic surgery specifically, Patient-Reported Outcome Measures offer something photographs cannot: direct access to how patients experience their results. The question is no longer whether PROMs have a place in practice, but how quickly we can integrate them.

The FACE-Q emerged from this context. Developed with rigorous methodology—extensive patient interviews, proper psychometric testing—it has become the reference standard for facial procedures [5–7]. A growing body of research documents its performance across populations and procedure types. At the same time, technological developments are pushing the field forward. Computerized Adaptive Testing promises shorter questionnaires without sacrificing accuracy [8].

Machine learning applications are beginning to show what prediction models might eventually offer [9].

This review sets out to do several things: synthesize what we know about PROM use in facial aesthetics, share validation data from our Dutch cohorts and CAT simulation work, consider where emerging technologies fit in, and—perhaps most practically—outline how implementation actually works in a clinical setting. The literature is full of advocacy for PROMs. What seems less common is honest discussion of what adoption looks like on the ground.

2. Historical Development of Outcome Measurement

2.1. *The Surgeon's Eye Versus the Patient's Experience*

For most of the specialty's history, outcomes meant what surgeons said they meant. We took photographs, judged our work, tracked complications [10]. Standardized photography was—and remains—useful for documentation, but a before-and-after image shows morphology, not psychology. Studies have repeatedly found that surgeon satisfaction correlates poorly with patient satisfaction [11]. This shouldn't surprise us. Surgeons evaluate technical execution. Patients live with emotional and social consequences.

Early attempts to bring patient perspectives into outcomes research relied on generic instruments. The SF-36, introduced in the 1980s, measured health-related quality of life across domains like physical function and mental health [12]. The problem was sensitivity. A patient experiencing significant appearance-related distress might score normally on scales designed to detect conditions like chronic pain or depression. Generic measures simply weren't built for aesthetic concerns [13]. This limitation drove development of specialty-specific tools.

2.2. *How the FACE-Q Came About*

Klassen, Pusic, and their team at Memorial Sloan Kettering developed the FACE-Q between 2008 and 2010, and the approach mattered as much as the result [5,6]. Rather than adapting an existing questionnaire—the usual shortcut—they started from scratch. Patient interviews and focus groups generated the content. Cognitive debriefing tested whether items made sense to respondents. The resulting scales address what patients actually care about, not what clinicians assume they care about.

The instrument covers four domains: appearance satisfaction (both overall and feature-specific), quality of life (psychological function, social confidence, perceived age), adverse effects (symptoms and recovery), and process of care (satisfaction with surgeon, staff, information). Each scale works independently—you can administer just the modules relevant to a given patient. Perhaps more importantly, the scales were built using Rasch Measurement Theory, which ensures interval-level measurement [14,15]. This technical detail matters because it's what allows us to calculate meaningful change scores and make valid statistical comparisons.

3. What the Evidence Shows Across Procedure Types

3.1. *Facelifts and Lower Face Work*

Ten studies covering 563 facelift patients give us a reasonable picture of FACE-Q performance in this population [16]. By 12 months, satisfaction scores are high and psychological function has improved substantially. The Age Appraisal scale—which asks how old patients think they look—shows perceived rejuvenation of 2.5 to 7 years, depending on the study. Adding blepharoplasty seems to boost composite scores more than fat grafting does. These aren't earth-shattering findings, but they provide the kind of data that helps with patient counselling. Expectations can be grounded in something other than anecdote.

3.2. *Injectables*

The injectable literature is larger—23 studies, 2292 patients [16]. Minimally invasive treatments produce measurable improvements across FACE-Q domains. Botulinum toxin and dermal fillers both work, though combination approaches tend to outperform single-modality treatment. Peak satisfaction arrives at different times: around two weeks for neurotoxin, four weeks for fillers.

What's particularly striking is the FACE-Q's ability to discriminate between products and techniques that look similar in photographs. Clinical photography might show comparable results from two different fillers; patient-reported scores can reveal that one performs better. This sensitivity makes the instrument useful not just for practice quality but for product evaluation research.

3.3. *Rhinoplasty*

The FACE-Q Rhinoplasty module has been used in 10 studies with 937 patients [16,17]. Satisfaction at six months is generally strong, but the picture is more nuanced when you break it down. Male and female patients show different score patterns. Ethnic minorities may require more careful goal-setting—a finding that demands thoughtful interpretation rather than algorithmic application. Younger patients with higher socioeconomic status tend to show larger improvements, though whether this reflects different starting points or genuinely different responses remains unclear.

3.4. *Our Dutch Validation Work*

We've contributed to the evidence base through a validation cohort of 259 Dutch patients undergoing facial surgery [18]. Scale reliability was excellent—Cronbach's alpha ranged from 0.885 to 0.951. What we found valuable, beyond the statistics, was the clinical utility of routine assessment. Some patients with perfectly acceptable surgical results scored high on psychosocial distress measures. Without structured questioning, these individuals wouldn't have been identified until they presented with complaints. Early identification opens the door to appropriate support.

4. Why This Matters Clinically

4.1. *Refining Patient Selection*

Preoperative FACE-Q scores establish baseline satisfaction and—just as importantly—reveal which specific concerns are driving the consultation [19]. A patient with low nasolabial fold scores but high overall facial satisfaction might be better served by targeted fillers than by a full surgical approach. Postoperative scores then quantify what changed and flag cases where intervention may be needed despite technically sound surgery.

Psychological function scores deserve particular attention. Very low baseline scores predict poor outcomes regardless of surgical quality—these patients may need psychological support before proceeding [19,20]. In our experience, moderate dissatisfaction with specific features, paired with realistic expectations, tends to respond well to targeted treatment. But recognizing these patterns requires having the data in the first place.

4.2. *Benchmarking and Quality*

Reoperation rates miss patients who are unhappy but don't come back. Standardized PROM scores capture satisfaction whether or not someone pursues revision. This allows for meaningful comparison across surgeons, techniques, and settings—the kind of data that can actually drive quality improvement rather than just box-ticking.

4.3. *Documentation and Medicolegal Considerations*

When disputes arise, documented baseline expectations and postoperative satisfaction provide objective evidence that pure clinical notes often lack. Several insurers have begun offering premium

reductions for practices with systematic PROM collection. The medicolegal argument, in isolation, might not justify implementation—but combined with clinical utility, it adds up.

4.4. *Where Healthcare Is Heading*

Value-based care is coming, whether we embrace it or not [4]. Demonstrating patient benefit through validated instruments will increasingly matter for reimbursement, regulation, and informed consumer choice. Practices that start collecting PROMs now will have historical data when they need it. Those that wait will be building infrastructure under pressure.

5. New Technologies and What They Might Offer

5.1. *Adaptive Testing*

Standard PROMs ask everyone the same questions regardless of relevance. Computerized Adaptive Testing takes a different approach: based on each response, the algorithm selects the next most informative item [8]. A rhinoplasty patient might answer mostly nose-specific questions while a facelift patient focuses on age-related items. The questionnaire adapts to the individual.

We ran simulation studies on 1000 FACE-Q administrations to see what CAT could do [8]. Question burden dropped by up to 71%. Correlation with full-form scores was 0.98—essentially no information loss. Patients preferred the shorter format. From a practical standpoint, cutting a five-minute questionnaire to under two minutes removes one of the main objections clinicians raise about PROM use.

5.2. *Predictive Models*

Machine learning trained on PROM databases opens up interesting possibilities [21]. Could we estimate satisfaction likelihood from preoperative data? Predict which patients might regret their decision? Flag cases where psychological screening should be prioritized? The work is still preliminary, but early results suggest these predictions might eventually be clinically useful. The key word is eventually—we're not there yet, and overpromising would be a mistake.

5.3. *Combining Subjective and Objective Data*

The logical next step is integrating PROMs with objective measures: 3D imaging for volumetric analysis, AI-based symmetry assessment, skin quality metrics, dynamic facial analysis [22]. Neither subjective nor objective data alone tells the complete story. Together, they could provide comprehensive outcome assessment that captures both what changed physically and how patients experience those changes.

6. Making It Work in Practice

6.1. *The Usual Objections*

We hear the same concerns repeatedly. Time: electronic administration takes three to five minutes, and pre-visit completion via patient portal takes no clinic time at all. Cost: the FACE-Q is available through Q-Portfolio at no charge for clinical use. You can start with paper if electronic systems aren't ready. Interpretation: published norms exist, and the learning curve is manageable. Patient buy-in: response rates average around 83% when PROMs are framed as standard quality monitoring rather than optional research.

6.2. *A Realistic Rollout*

Based on what we've seen work: spend the first two months selecting modules, training staff, and building basic protocols. Don't overcomplicate it. Months three and four, pilot with new consultations only—paper forms are fine at this stage. Review the data weekly as a team and adjust

the process. Expand to all aesthetic patients and transition to electronic collection around months five and six. From month seven onward, focus on automation, dashboards, and using the data for quality improvement. The whole process takes about half a year to feel routine.

7. Limitations and Honest Criticisms

Some colleagues argue that reducing aesthetic surgery to numbers misses the point—that this is fundamentally artistic work. There's something to that concern, but PROMs don't replace artistic judgment. They ensure that judgment aligns with what patients actually want. Ignoring patient preferences isn't artistry; it's arrogance.

The risk of score misuse is real. Insurers might deny coverage based on predicted satisfaction. Hospitals might penalize surgeons with lower scores without adjusting for case mix. Appropriate safeguards—risk adjustment, confidence intervals, focus on improvement trends—matter. But the solution is thoughtful implementation, not avoidance of measurement altogether.

This review has its own limitations. Narrative synthesis invites selection bias. We couldn't perform quantitative meta-analysis. Much of the evidence comes from specialty centres; generalizability to community settings isn't assured. These caveats don't undermine the basic argument for PROMs, but they do suggest where more research would help.

8. Where This Is Going

Real-time smartphone monitoring is probably coming—daily micro-assessments tracking recovery, detecting complications early, guiding postoperative care. International registries, facilitated by the FACE-Q's availability in more than 30 languages, could enable cross-cultural comparison and rare event detection. Integration with electronic health records and value-based payment will accelerate adoption whether individual practices embrace it or not.

The practices building PROM infrastructure now will be ready when this arrives. Those that haven't started will find themselves catching up.

9. Conclusions

The evidence for PROM implementation in facial aesthetics is strong. The FACE-Q provides reliable, sensitive measurement of outcomes that matter to patients. Implementation is feasible—we've seen it work. Emerging technologies promise efficiency gains that address the main practical objections.

Our patients trust us with their appearance, their sense of identity, their confidence. Demonstrating that we actually improve their lives—with evidence, not just reassurance—seems the least we can do.

Author Contributions: Conceptualization, M.J.O.; Methodology, M.J.O.; Investigation, M.J.O.; Writing—Original Draft Preparation, M.J.O.; Writing—Review and Editing, M.J.O. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: Thanks to colleagues at Clinique Rebelle for their support and to the Harvard PROVE Center team for ongoing collaboration.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Ching, S.; Thoma, A.; McCabe, R.E.; Antony, M.M. Measuring Outcomes in Aesthetic Surgery: A Comprehensive Review of the Literature. *Plast. Reconstr. Surg.* **2003**, *111*, 469–480.
2. Dobbs, T.D.; Hughes, S.; Mowbray, N.; Hutchings, H.A.; Whitaker, I.S. How to Decide Which Patient-Reported Outcome Measure to Use? A Practical Guide for Plastic Surgeons. *J. Plast. Reconstr. Aesthet. Surg.* **2018**, *71*, 957–966.
3. Porter, M.E.; Teisberg, E.O. *Redefining Health Care: Creating Value-Based Competition on Results*; Harvard Business School Press: Boston, MA, USA, 2006.
4. Porter, M.E. What Is Value in Health Care? *N. Engl. J. Med.* **2010**, *363*, 2477–2481.
5. Klassen, A.F.; Cano, S.J.; Scott, A.; Snell, L.; Pusic, A.L. Measuring Patient-Reported Outcomes in Facial Aesthetic Patients: Development of the FACE-Q. *Facial Plast. Surg.* **2010**, *26*, 303–309.
6. Klassen, A.F.; Cano, S.J.; Scott, A.M.; Pusic, A.L. Measuring Outcomes That Matter to Face-Lift Patients: Development and Validation of FACE-Q Appearance Appraisal Scales and Adverse Effects Checklist for the Lower Face and Neck. *Plast. Reconstr. Surg.* **2014**, *133*, 21–30.
7. Cano, S.J.; Klassen, A.; Pusic, A.L. The Science behind Quality-of-Life Measurement: A Primer for Plastic Surgeons. *Plast. Reconstr. Surg.* **2009**, *123*, 98e–106e.
8. Ottenhof, M.J.; Geerards, D.; Harrison, C.; Klassen, A.F.; Hoogbergen, M.M.; van der Hulst, R.R.W.J.; Pusic, A.L. Applying Computerized Adaptive Testing to the FACE-Q Skin Cancer Module: Individualizing Patient-Reported Outcome Measures in Facial Surgery. *Plast. Reconstr. Surg.* **2021**, *148*, 1137–1145.
9. Dixit, R.R.; Balasubramanian, S.; Dattatray Dhaygude, A.; Devarajan, H.R.; Jhapte, R.; Chunawala, H. Predictive Modelling of Patient Outcomes Using Machine Learning: A Comparative Analysis of Algorithmic Approaches in Healthcare. In Proceedings of the 2024 International Conference on Artificial Intelligence and Emerging Technologies, Global AI Summit, Virtual, 4 September 2024; pp. 418–423.
10. Kosowski, T.R.; McCarthy, C.; Reavey, P.L.; Scott, A.M.; Wilkins, E.G.; Cano, S.J.; Klassen, A.F.; Rubin, J.P.; Pusic, A.L. A Systematic Review of Patient-Reported Outcome Measures after Facial Cosmetic Surgery and/or Nonsurgical Facial Rejuvenation. *Plast. Reconstr. Surg.* **2009**, *123*, 1819–1827.
11. Honrado, C.P.; Larrabee, W.F., Jr. Update in Three-Dimensional Imaging in Facial Plastic Surgery. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2004**, *12*, 327–331.
12. Ware, J.E., Jr.; Sherbourne, C.D. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection. *Med. Care* **1992**, *30*, 473–483.
13. Pusic, A.L.; Klassen, A.F.; Scott, A.M.; Cano, S.J. Development and Psychometric Evaluation of the FACE-Q Satisfaction with Appearance Scale: A New Patient-Reported Outcome Instrument for Facial Aesthetics Patients. *Clin. Plast. Surg.* **2013**, *40*, 249–260.
14. Andrich, D. Rating Scales and Rasch Measurement. *Expert Rev. Pharmacoecon. Outcomes Res.* **2011**, *11*, 571–585.
15. Bond, T.G.; Fox, C.M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 3rd ed.; Routledge: New York, NY, USA, 2015.
16. Ottenhof, M.J.; Veldhuizen, I.J.; Hensbergen, L.J.V.; Blankensteijn, L.L.; Bramer, W.; Lei, B.V.; Pusic, A.L.; van der Hulst, R.R.W.J.; Cano, S.J. The Use of the FACE-Q Aesthetic: A Narrative Review. *Aesthetic Plast. Surg.* **2022**, *46*, 2769–2780.
17. Citron, I.; Townley, W. Assessing Outcomes from Rhinoplasty Using Clinical and Patient Reported Measures (FACE-Q™). *J. Plast. Reconstr. Aesthet. Surg.* **2023**, *84*, 182–186.
18. Ottenhof, M.J.; Meulendijks, M.Z.; Lardinois, A.; Deibel, D.; van der Hulst, R.; van der Pot, W.; Hoogbergen, M.M. Nasal Tip Defects: Satisfaction with Rintala Flap for Reconstruction—A Report of 38 Cases. *Eur. J. Plast. Surg.* **2022**, *45*, 741–745.
19. Gallo, L.; Churchill, I.; Kim, P.; Rae, C.; Voineskos, S.H.; Thoma, A.; Cano, S.J.; Pusic, A.L. Patient Factors That Impact FACE-Q Aesthetics Outcomes: An Exploratory Cross-Sectional Regression Analysis. *Aesthet. Surg. J.* **2025**, *45*, 543–551.
20. Honrado, C.P.; Lee, S.; Bloomquist, D.S.; Larrabee, W.F., Jr. Quantitative Assessment of Nasal Changes after Maxillomandibular Surgery Using a 3-Dimensional Digital Imaging System. *Arch. Facial Plast. Surg.* **2006**, *8*, 26–35.

21. Dixit, R.R.; Balasubramanian, S.; Dattatray Dhaygude, A.; Devarajan, H.R.; Jhapte, R.; Chunawala, H. Predictive Modelling of Patient Outcomes Using Machine Learning: A Comparative Analysis of Algorithmic Approaches in Healthcare. In Proceedings of the 2024 International Conference on Artificial Intelligence and Emerging Technologies, Global AI Summit, Virtual, 4 September 2024; pp. 418–423.
22. Popat, H.; Richmond, S.; Playle, R.; Marshall, D.; Rosin, P.L.; Cosker, D. Three-Dimensional Motion Analysis—An Exploratory Study. Part 2: Reproducibility of Facial Movement. *Orthod. Craniofac. Res.* **2008**, *11*, 224–228.
23. 23. Clinique Rebelle. Clinique Rebelle—Plastic Surgery for Mothers. www.cliniquerebelle.com. Accessed 2026.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.