

European Journal of Pediatric Surgery

Evaluating Inter- and Intraobserver Agreement on Pectus Carinatum Severity and Treatment Outcomes: A Comparison of Subjective and Objective Assessment Methods

Hendrik van Braak, Sjoerd de Beer, Sander Zwaveling, Matthijs W Oomen, L.W. Ernest van Heurn, Justin R de Jong.

Affiliations below.

DOI: 10.1055/a-2466-6407

Please cite this article as: van Braak H, de Beer S, Zwaveling S et al. Evaluating Inter- and Intraobserver Agreement on Pectus Carinatum Severity and Treatment Outcomes: A Comparison of Subjective and Objective Assessment Methods. European Journal of Pediatric Surgery 2024. doi: 10.1055/a-2466-6407

Conflict of Interest: The authors declare that they have no conflict of interest.

Abstract:

Background

Visual examination is crucial for assessing pectus carinatum (PC) severity and treatment results. This cross-sectional study evaluates the inter- and intraobserver agreement of PC deformities before and after treatment.

Methods

Observers examined medical photographs of patients before and after treatment. Primary outcome was inter- and intraobserver agreement on esthetic results after treatment. Secondary outcomes included inter- and intraobserver agreement on severity and symmetry before treatment, differences in esthetic results after Ravitch surgery and dynamic compression bracing (DCS-bracing), and the impact of scars, age and treatment duration on esthetic results. Results

Medical photographs of 201 patients (aged 4-18) were evaluated by five surgeons and five peers. Surgeons and peers demonstrated inadequate ($\kappa < .61$) interobserver agreement on esthetic results ($\kappa = .26$, $\kappa = .22$), severity of PC ($\kappa = .43$, $\kappa = .38$) and symmetry ($\kappa = .37$, surgeons only). Agreement between surgeons and peers on esthetic results ($\kappa = .37$) and severity before treatment ($\kappa = .54$) was similarly inadequate. Surgeons and peers demonstrated inadequate intraobserver agreement on esthetic results ($\kappa = .49$, $\kappa = .34$), severity of PC ($\kappa = .54$, $\kappa = .48$) and symmetry ($\kappa = .60$, surgeons only). Deformities treated with Ravitch surgery were perceived as more severe but yielded better results. Peers, unlike surgeons, viewed scars as negatively impacting results. No relationship was found between results after treatment and treatment duration ($p = .682$, $p = .062$) or age ($p = .205$, $p = .527$).

Conclusions

Subjective assessment of PC severity and esthetic results is inconsistent. 3D-scanning could help standardize treatment completion and aid patients and surgeons in determining treatment completion. The psychosocial effects of scars should be addressed when discussing treatment options.

Corresponding Author:

Hendrik van Braak, Amsterdam UMC Locatie AMC, Department of Pediatric Surgery, Meibergdreef 9, 1105 AZ Amsterdam, Netherlands, h.vanbraak@amsterdamumc.nl

Affiliations:

Hendrik van Braak, Amsterdam UMC Locatie AMC, Department of Pediatric Surgery, Amsterdam, Netherlands
Sjoerd de Beer, Amsterdam UMC Locatie AMC, Department of Pediatric Surgery, Amsterdam, Netherlands

Sander Zwaveling, Amsterdam UMC Locatie AMC, Department of Pediatric Surgery, Amsterdam, Netherlands
[...]
Justin R de Jong, Amsterdam UMC Locatie AMC, Department of Pediatric Surgery, Amsterdam, Netherlands



This article is protected by copyright. All rights reserved.

Accepted Manuscript

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.



Introduction

Pectus carinatum (PC) is a common pediatric condition characterized by overgrowth of costal cartilage which causes protrusion of the sternum and adjacent cartilage. In our institution, patients with PC were previously treated with surgical options like the Ravitch procedure and Abramson procedures. However, non-invasive treatment using Dynamic Compression System (DCS) bracing has become the preferred treatment¹⁻⁴. Although treatment outcomes are similar⁵, a significant clinical challenge persists. Surgical outcomes are typically definitive, providing clear correction, but determining endpoints for DCS-bracing therapy remains challenging.

The subjective nature of PC's esthetic impact makes determining treatment success challenging. While objective tools like 3D-scanning and the Haller index provide measurable data on the reduction of the chest deformity^{6,7}, they do not fully capture the patient's own view of success. At our institution, patients expectations vary widely; some are satisfied with modest improvement, while others aim for an ideal chest appearance. As a result, even with objective measurements, outcomes ultimately hinge on the patients' own view of success and satisfaction, in mutual agreement with the physician^{5, 8, 9}. Differences in treatment success expectations between patients and physicians can influence treatment decisions, like whether or not to start the retainer phase, and treatment duration, raising questions about consistency in evaluating treatment results. We aimed to assess whether these inconsistencies exist.

The aim of this study was to evaluate the inter- and intraobserver agreement, between surgeons and peers, primarily on esthetic results after treatment and secondary on severity and symmetry before treatment. Additionally we assessed differences in average ratings between surgeons and peers and explored the influence of scars, age and treatment duration on esthetic results.

Materials and methods

Study design

At time of study conceptualization, we already had a sufficient collection of medical photographs available for analysis. This allowed us to conduct a retrospective cross-sectional study wherein medical photographs of pediatric patients with pectus carinatum were visually evaluated by reviewers. A METC official waiver of ethical approval was granted from the METC of the Amsterdam Medical Center. The need for informed consent was waived.

Patients and reviewers

All patients who successfully finished treatment for pectus carinatum with DCS-bracing or Ravitch surgery in the Amsterdam Pectus Center between 2009 and 2019, and had before and after treatment photographs taken, were included. Treatment with DCS-bracing was regarded successful if patients achieved a satisfactory result, in consultation with the clinician. The results of Ravitch surgery were consistently regarded as successful, as the procedure allows for a precise and complete correction of the deformity. Figure 1 shows the diagnostic and therapeutic flowchart used in treatment. Eight medical photographs (four before and four after treatment (Figure 2 and Figure 3) were taken from different angles as part of our standard protocol. Photographs were intended to be taken from every patient, not just the extremes. Photographs were reviewed by five pediatric surgeons and five peers of patients. The limited sample size reflects the exploratory nature of this study in assessing observer agreement on esthetic outcomes.

Peers

We refer to individuals who share similar age-related characteristics with the patients as peers. Peers were selected instead of patients to provide an external perspective on esthetic outcomes, one that is less influenced by personal treatment experiences or individual biases. These peers were not directly involved in the patients' care nor were they family members and were recruited by posting an open invitation at a local high school and university, asking students to participate in the research. The peers, aged 14, 14, 16, 16 and 23 years, included three males and

two females. Their age matches the age of patients, to offer a perspective more aligned with the patients' age related social and esthetic standards, which can be particularly sensitive in adolescent and young adult populations. The number of peers was matched with the number of surgeons to ensure a balanced evaluation, minimizing potential biases from unequal sample sizes.

Data collection and measurements

Data were obtained retrospectively from patient records. Observers were asked to visually examine medical photographs of PC patients and rate the severity before treatment on a three point scale (one = mild, two = moderate, three = severe) and judge the results after treatment (one = minimal improvement, two = moderate improvement, three = good improvement). Both surgeons and peers were asked to assess esthetic results of patients who underwent surgery with the scar present, and subsequently, visualize the photographs without the scar, providing an extra rating. The scars were not digitally removed, patients were just asked to imagine the photograph without the scar. Furthermore surgeons were asked to determine if the deformity before treatment was symmetrical or not. This additional question was included exclusively for surgeons, as this factor can influence treatment selection and approach. This question was not posed to peers because they lack the clinical context to understand how symmetry affects treatment decisions, rendering the question redundant. Observers were blinded to all clinical information and completed the assessment twice, with at least a two weeks interval between both assessments.

Outcomes

Primary outcomes were interobserver and intraobserver agreement on esthetic results after treatment, secondary outcomes were inter- and intraobserver agreement on severity and symmetry before treatment, differences in average ratings between the Ravitch and DCS-

bracing group and the influence of scars, age and treatment duration on esthetic results after treatment.

Statistical analysis

Descriptive measurements were utilized to characterize the study population. To assess interobserver agreement on esthetic results, severity and symmetry, we applied Cohen's kappa statistics to all pairs of observers using ratings from the initial assessment round. Cohen's kappa was selected because it accounts for chance agreement, providing a more robust measure of agreement than simple percent agreement. The weighted arithmetic mean of all observer pairs, determined by the number of evaluated subjects, was adopted as a measure of overall agreement¹⁰. Additionally, the interobserver agreement per observer was based on the weighted arithmetic mean of all observer pairs that included the observer in question. We computed two-sided 95% confidence intervals (CI). A similar approach was taken for intraobserver and scar-versus-no-scar agreement, but this time, we involved repeated assessments by the same surgeon. Average agreement was based on the mean of all observers. Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) was employed to assess agreement when there was substantial disparity in the prevalence of response options¹¹. PABAK was specifically chosen as it adjusts for imbalances in response distribution, allowing a clearer understanding of agreement where one response may be more common than others. The interpretation of kappa coefficients was as follows: coefficients ranging from .00 to .20, .21 to .40, .41 to .60, .61 to .80, and .81 to 1.00 corresponded respectively to slight, fair, moderate, substantial, and almost perfect agreement. These ranges provide context for the clinical significance of observed agreement, with higher values indicating more reliable consistency among observers, a kappa value of .61 or above is typically considered adequate agreement, indicating a substantial level of consistency beyond chance¹². We used Spearman's rank correlation to evaluate the statistical relationship between average ratings on severity before treatment and improvement after treatment. Descriptive

statistics were used to calculate differences in average ratings. Descriptive statistics were used to analyze the relationship between treatment duration and age on results after treatment.

Statistical significance was determined at a probability value of less than 0.05. Data were analyzed using IBM SPSS Statistics 28.0.

Results

Between September 2009 and December 2019 a total of 736 patients underwent either Ravitch surgery or DCS-bracing for PC in the Amsterdam Pectus Center. Among these, 68.6% (n=505/736) completed treatment successfully. Medical photographs were taken before and after treatment for 201 patients (aged 4-18) out of this group (39.8%, n=201/505). While protocol dictates photographing all pectus carinatum patients who finalize treatment, not all attended their medical photography appointments. Both pre- and posttreatment medical photographs are needed to be eligible for evaluation. The majority of included patients were male (96.5%, n=194), with an median age of 15.0 (IQR 13.5-16.0) years. DCS-bracing was the primary treatment for most patients (69.2%, n=139), while the remaining patients underwent Ravitch surgery (30.8%, n=62). Six (9.7%) patients who underwent Ravitch surgery had previously failed DCS-bracing treatment. Median age in the Ravitch group was 15.0 (IQR 14.0-16.0) years compared to 14.0 (IQR 13.0-15.0) years in the DCS-bracing group ($p < .001$).

Interobserver agreement on severity before treatment, symmetry and results after treatment

[Table 1] shows the results of the interobserver agreement. Overall, agreement was inadequate across all assessments, as all kappa values were below .61, indicating less than substantial consistency.

Fair agreement was reached among both surgeons and peers ($\kappa = .26(.21-.31)$ and $\kappa = .22(.19-.26)$ resp.) on results after treatment. The overall agreement of severity before treatment among surgeons showed a moderate agreement ($\kappa = .43(.37-.49)$). This was slightly higher than the fair agreement observed among peers ($\kappa = .38(.34-.41)$). Surgeons exhibited fair agreement on

symmetry ($\kappa=.37(.29-.45)$). There was a moderate agreement between surgeons and peers on severity before treatment ($\kappa=.54(.48-.60)$) and a fair agreement on results after treatment ($\kappa=.37(.30-.45)$).

There was a significant relationship between average rating on severity before treatment and improvement after treatment (surgeons and peers $r=.39$, $p<.001$, $r=.43$, $p<.001$ resp.), with more severe deformities resulting in more noticeable improvements.

Intraobserver agreement on severity before treatment, symmetry and results after treatment, [Table 2] shows the results of the intraobserver agreement. Overall, agreement was inadequate across all assessments, as all kappa values were below 0.61, indicating less than substantial consistency.

Surgeons exhibited more consistent results on severity before treatment ($\kappa=.54(.47-.61)$) and results after treatment ($\kappa=.49(.40-.58)$) compared to peers ($\kappa=.48(.38-.59)$ and $\kappa=.34(.28-.40)$ resp.). Surgeons showed an moderate average intraobserver agreement ($\kappa=.60(.46-.74)$) on symmetry.

Difference in average ratings between the Ravitch and DCS-bracing group

[Table 3] presents the differences in average ratings between the two treatment groups. Both surgeons and peers consistently rated deformities in patients treated with Ravitch surgery as more severe compared to those treated with DCS-bracing. The analysis of the results after treatment reveals a similar trend, with superior results in the Ravitch group compared to the DCS-bracing group. On average, peers rated deformities before treatment as more severe than surgeons (2.11 versus 1.91, $p<.001$). Peers on the other hand were less positive on the results compared to surgeons (2.23 versus 2.40, $p<.001$).

Influence of scar's, age and treatment duration on esthetic results

All surgeons exhibited an adequate and almost perfect agreement ($\kappa=.89(.85-.94)$) on results with scars compared to results ‘imagining no scar’, as shown in [table 4]. This in contrast with patients’ peers ($\kappa=.39(.19-.60)$), exhibiting inadequate agreement.

There was a significant relationship between age and severity of PC before treatment (surgeons and peers, $p=.010$, $p=.029$ resp.), but not between age and improvement after treatment (surgeons and peers, $p=.205$, $p=.527$ resp.).

There was no statistically significant relationship between ratings after treatment and duration of treatment (surgeons and peers, $p=.682$, $p=.062$ resp.) in de DCS-bracing group.

Discussion

No adequate agreement was reached on esthetic results after treatment, severity of PC and symmetry. Surgeons exhibited better, but still inadequate intraobserver agreement compared to peers. Deformities of patients who underwent Ravitch surgery were regarded as more severe, but with better esthetic results afterwards compared to patients who underwent DCS-bracing. Peers rated chest deformities more negative than surgeons, assigning higher severity ratings before treatment and indicating less improvement afterward. Age was related to severity of PC before treatment. Both age and treatment duration were not related to final outcomes. More severe deformities before treatment obviously resulted in more noticeable improvements, as opposed to milder deformities.

Inter- and intraobserver agreement on results after treatment

While several studies report that treatment outcomes for pectus carinatum are determined based on patient satisfaction and mutually agreed-upon observed correction of the deformity¹³⁻¹⁵, our findings reveal that there was inadequate agreement on esthetic outcomes post-treatment, with low intraobserver agreement indicating limited consistency among raters. Inconsistent assessment of esthetic outcomes can lead to unclear treatment endpoints, misaligned patient

expectations and challenges in standardizing care. This underscores the need for objective evaluation methods in clinical practice.

A further complication is that developed pectoral muscles and a higher BMI can effectively mask PC. Many patients in our study with ‘good improvement’ developed big pectoral muscles, questioning whether PC actually improved, or was merely concealed by the pectoral muscles. If the results are primarily attributed to changes in muscle mass or BMI, there is a risk of recurrence of esthetic complaints when these factors change again.

Currently, patients are encouraged to build muscle as part of their treatment but are also cautioned that PC could reappear if they lose weight or muscle mass. Therefore, managing expectations before starting treatment is essential. Similar to observations in patients with pectus excavatum (PE) treated with vacuum bell therapy¹⁶, we find that as patients grow older and reach adulthood, they often become less concerned with their PC and are more likely to accept it if mild recurrence occurs over time.

Thus, an effective approach for obtaining a more objective assessment of PC outcomes should be one that is unbiased, but also capable of adjusting for variables in muscle mass and BMI to ensure accurate assessment of treatment success. 3D scanning is one such a technique^{6, 7, 17, 18, 19, 20, 21}, and has been shown to be able to account for BMI²². This method can evaluate the initial severity of PC and, importantly, monitor progress throughout treatment. Over time, it may also serve to establish objective treatment endpoints, helping to create a more standardized approach to patient care.

However, in practice, some patients may feel satisfied earlier than others, regardless of set endpoints. Some patients may wish to end therapy before reaching the set endpoints, while others may prefer to continue beyond them. Despite having determined treatment endpoints, the cosmetic nature of PC means there are no risks associated with shorter or longer treatment durations, making these varying preferences a complication in the treatment process..

Additionally, while 3D scanning is user-friendly, it demands extra time from physicians, whose time is already limited.

Inter- and intraobserver agreement on severity and symmetry before treatment

Determining which treatment is suitable for patients is usually decided based on pressure of initial correction (PIC), symmetry and flaring. When PC is too rigid or when the deformity is too asymmetrical or accompanied by flaring, a surgical correction is technically necessary⁵. Therefore the fair to moderate inter- and intraobserver agreement on PC severity is therefore less of a problem compared to the inconsistency regarding outcomes, since subjective evaluation of severity of PC does ultimately not influence the choice of treatment. Symmetry, on the contrary, does influence the choice of treatment, suggesting that the inadequate agreement on symmetry could have implications.

However, symmetry was not defined prior to the study because in practice also no consistent definition of symmetry was used. This may have led to the different interpretations of symmetry between surgeons. A clear, consistent definition of asymmetry is necessary since it can potentially influence clinical decision making. An asymmetry analysis as described in the literature could be utilized for this purpose¹⁷.

We found no relationship between severity of PC and treatment duration, likely due to inadequate inter- and intraobserver agreement on severity.

Ravitch versus DCS-bracing

Deformities in surgically treated patients were consistently rated as more severe than those in patients treated with DCS-therapy, by both surgeons and peers. Naturally, more severe deformities yield more noticeable improvements, compared to the milder deformities in the DCS-bracing group. This likely explains the better perceived results in the Ravitch group relative to those treated with DCS-bracing.

Interestingly, when examining the same data, peers tend to rate both the initial deformities and post-treatment results more negatively than surgeons. This may be due to surgeons' greater familiarity with these deformities, leading to lower severity ratings based on their clinical experience. Additionally, peers may be more sensitive to esthetic factors like scarring, which influences their assessments of post-treatment results more critically than those of surgeons.

Influence of scars on esthetic results after treatment

Our results show a significant difference in how surgeons and patients' peers perceive the impact of scars on results after treatment. Surgeons may regard scars as a minor issue due to their focus on achieving structural correction and clinical outcomes, viewing scars as a common, unavoidable aspect of effective surgical intervention. This highlights the importance of considering individual perspectives when counseling patients on treatment options like the Ravitch procedure, and especially the minimally invasive Abramson procedure, as scars can significantly impact patients and are avoided with DCS-bracing.

Age and severity of treatment and influence on esthetic results

The relationship between age and severity of deformities before treatment may reflect the age difference between the Ravitch and DCS-bracing groups, with the Ravitch group generally being older. Patients in the Ravitch group also had more severe deformities, suggesting that severity might confound the observed association between age and deformity. While earlier bracing for younger patients could be considered to prevent deformity progression, we view the risk of relapse as a greater concern than progression and therefore prefer to initiate treatment at an older age to minimize this risk⁵. Additionally, age was not found to be related to better esthetic outcomes, suggesting there is no advantage to starting treatment earlier in terms of achieving improved results.

Influence of treatment duration on esthetic results

No relationship was found between treatment duration and results after treatment. This result might be biased because all our patients were already regarded as successfully treated.

Limitations

The fact that only 39.8% of patients had photographs available for analysis raises a concern regarding potential selection bias. While the reduced sample size may influence the results, there's no clear evidence that patients without photographs differ significantly in outcomes from those included in the analysis, since all patients were treated successfully and were satisfied with the treatment.

A limitation of our study is the potential bias introduced by the difference in experience between the peer and surgeon groups in assessing PC. However, this contrast was intentional, as it mirrors real-world scenarios where patients are inexperienced in evaluation of PC in contrast to surgeons. Understanding how these groups differ in their assessments provides valuable insights into patient experiences and expectations post-treatment. Additionally, the small number of peers reflects the exploratory nature of this experimental study, aimed at gaining preliminary insights rather than achieving broad generalizability.

Despite the impact of flaring on clinical decision making, a question regarding flaring was not included, because surgeons indicated that assessing flaring was not always feasible in our photographs. This would make analysis of flaring in our study unreliable. Although flaring is sometimes considered when choosing between surgery and bracing, its presence or absence does not impact treatment outcomes, which is the primary focus of this study.

In our study we asked observers to visualize the photograph without scar, which may have introduced some bias by placing the burden of imagination on observers and potentially impacting the accuracy of their assessments. A more objective method, such as digitally removing scars from the photographs, could have helped to standardize evaluations and minimize bias by allowing observers to focus solely on the correction of the deformity itself.

There may be a difference in visualization skills between peers and surgeons. However, given the clear distinction observed in their assessments, we believe our method proved effective overall.

Conclusion

Subjective assessments of pectus carinatum (PC) severity and treatment outcomes showed fair to moderate agreement, with peers rating deformities more severely and being more critical of scars. Ravitch surgery led to greater improvements for severe cases, but age and treatment duration had no significant impact on results. The need for objective tools like 3D scanning is clear to improve consistency and treatment evaluation. Future efforts should focus on refining these tools and addressing the psychosocial effects of scars in PC treatment.

Acknowledgments

No funds, grants, or other support were received during the preparation of this manuscript.

Conflicts of interest

None declared

References

- 1 Welch KJ, Vos A. Surgical correction of pectus carinatum (pigeon breast). *J Pediatr Surg.* 1973;8:659-67
- 2 Martinez-Ferro M, Bellia Munzon G, Fraire C, et al. Non surgical treatment of pectus carinatum with the FMF(®) Dynamic Compressor System. *J Vis Surg.* 2016;2:57
- 3 Martinez-Ferro M, Fraire C, Bernard S. Dynamic compression system for the correction of pectus carinatum. *Semin Pediatr Surg.* 2008;17:194-200
- 4 Abramson H, Aragon X, Blanco JB, Ciano A, Abramson L. Minimally invasive repair of pectus carinatum and how to deal with complications. *J Vis Surg.* 2016;2:64

- 5 Van Braak H, de Beer SA, Zwaveling S, Oomen MWN, De Jong JR. Ravitch Surgery or Dynamic Compression Bracing for Pectus Carinatum: A Retrospective Cohort Study. *Ann Thorac Surg.* 2024;117:144-152
- 6 Song SH, Kim CH, Moon DH, Lee S. Usefulness of 3-Dimensional Body Surface Scanning in the Evaluation of Patients with Pectus Carinatum. *Korean J Thorac Cardiovasc Surg.* 2020;53:301-5
- 7 Belgacem A, Tricard J, Dutoit A, et al. Efficiency of non-operative management for pectus deformities in children using an X-ray-free protocol. *Interdiscip Cardiovasc Thorac Surg.* 2023;36
- 8 de Beer SA, Blom YE, Lopez M, de Jong JR. Measured dynamic compression for pectus carinatum: A systematic review. *Semin Pediatr Surg.* 2018;27(3):175-82
- 9 Emil S, Sévigny M, Montpetit K, et al. Success and duration of dynamic bracing for pectus carinatum: A four-year prospective study. *J Pediatr Surg.* 2017;52:124-9
- 10 Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychol Bull.* 1971;76:365-77
- 11 Byrt T, Bishop J, Carlin J. Bias, Prevalence and Kappa. *J Clin Epidemiol.* 1993;33:159-74
- 12 Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 1977;33:159-74
- 13 Dekonenko C, Dorman RM, Pierce A, et al. Outcomes Following Dynamic Compression Bracing for Pectus Carinatum. *J Laparoendosc Adv Surg Tech A.* 2019;29:1223-7
- 14 Trinidad S, McKenna E, Jenkins T, et al. Outcomes From a Nurse Practitioner Led Dynamic Compression System Bracing Program for Pectus Carinatum. *J Pediatr Surg.* 2024;59:950-5

- 15 Kelly RE, Obermeyer RJ, Goretsky MJ, et al. Ten-year experience with staged management of pectus carinatum: Results and lessons learned. *J Pediatr Surg.* 2021;56:1835-40
- 16 van Braak H, de Beer SA, al Ghouch Y, et al. 15 years of vacuum bell therapy for pectus excavatum: long-term outcomes and influencing factors. *J Pediatr Surg.* 2024
- 17 Lain A, Garcia L, Gine C, Tiffet O, Lopez M. New Methods for Imaging Evaluation of Chest Wall Deformities. *Front Pediatr.* 2017;5:257
- 18 Daemen JHT, Loonen TGJ, Lozekoot PWJ, et al. Optical imaging versus CT and plain radiography to quantify pectus severity: a systematic review and meta-analysis. *J Thorac Dis.* 2020;12:1475-87
- 19 Wong KE, Gorton GE 3rd, Tashjian DB, Tirabassi MV, Moriarty KP. Evaluation of the treatment of pectus carinatum withh compressive orthotic bracing using three dimensional body scans. *J Pediatr Surg.* 2014;49:924-7
- 20 Daemen JHT, de Loos ER, Geraedts TCM, et al. Visual diagnosis of pectus excavatum: An inter-observer and intra-observer agreement analysis. *J Pediatr Surg.* 2022;57:526-31
- 21 Port E, Hebal F, Hunter CJ, Malas B, Reynolds M. Measuring the impact of brace intervention on pediatric pectus carinatum using white light scanning. *J Pediatr Surg.* 2018;53:2491-4
- 22 Taylor JS, Szafer D, Madhavan S, et al. Algorithm for Calculating Haller Index for Pectus Excavatum Using 3-Dimensional Optical Scanning. *J Am Coll Surg.* 2018;227:S192
- 23 de Beer S, Volcklandt S, de Jong J, Oomen M, Zwaveling S, van Heurn E. Dynamic Compression Therapy for Pectus Carinatum in Children and Adolescents: Factors for Success. *J Pediatr Surg.* 2023;58:1440-5

Table 1

Interobserver agreement between surgeons and peers

Table 2

Intraobserver agreement of peers and surgeons

Table 3

Average rating of severity before treatment and results after treatment

Table 4

Scars versus no scars

Figure 1

Diagnostic and therapeutic flowchart (*CT = Computed Tomography, MRI = Magnetic Resonance Imaging, PIC = Pressure of Initial Correction, psi = pounds per square inch*)

Figure 2

Example of four photographs of a patient before treatment

Figure 3

Example of four photographs of a patient after treatment

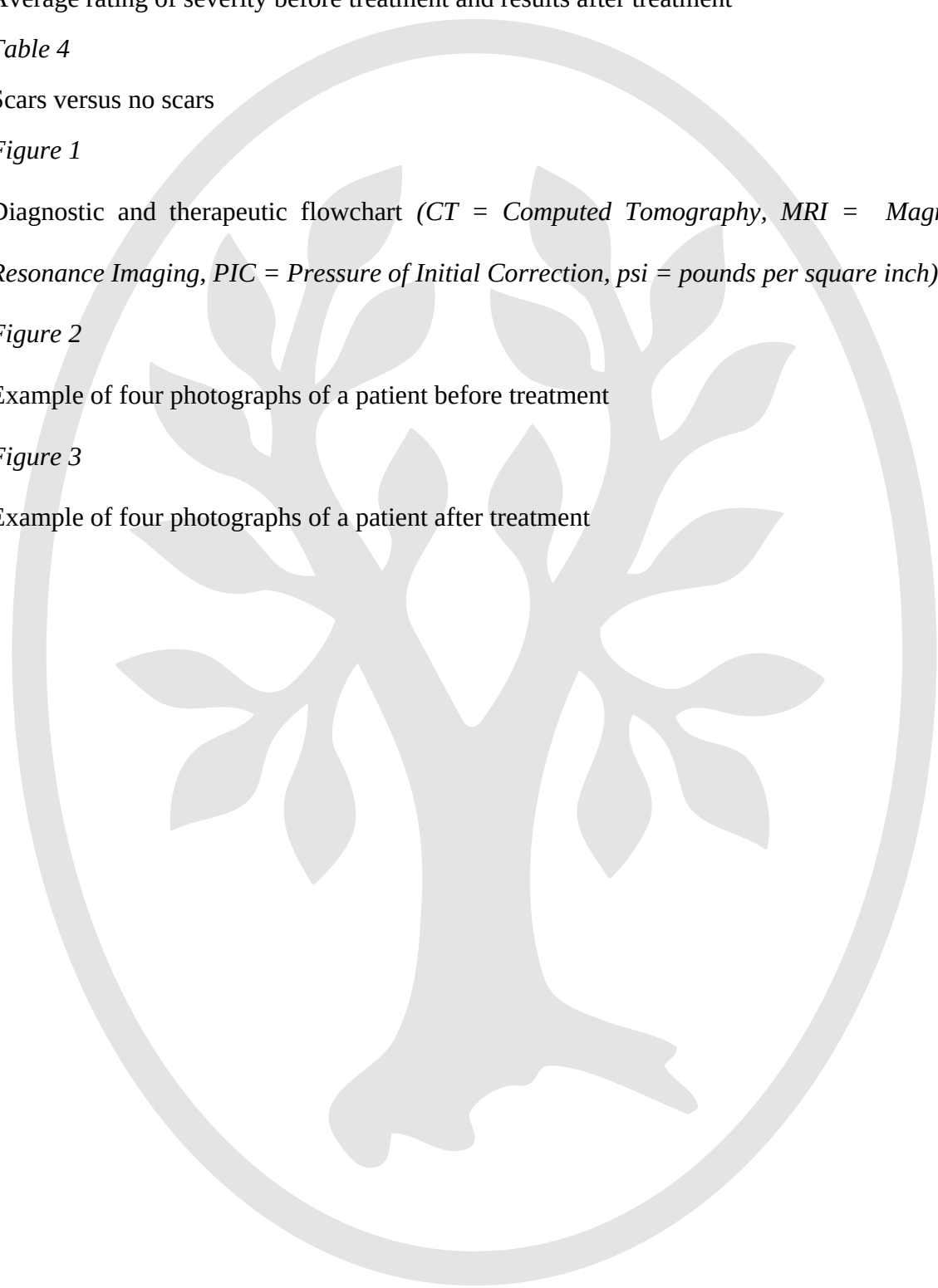


Table 1.

	Severity before treatment		Results after treatment ^a
	Severity ^a	Symmetry ^a	
SURGEONS			
Overall agreement	.43 (.37-.49)	.37 (.29-.45)	.26 (.21-.31)
Surgeon A versus B-E	.35 (.27-.42)	.45 (.38-.53)	.24 (.16-.32)
Surgeon B versus A, C-E	.45 (.39-.52)	.43 (.36-.51)	.36 (.28-.44)
Surgeon C versus AB, DE	.49 (.43-.56)	.31 (.23-.40)	.36 (.28-.44)
Surgeon D versus A-C, E	.42 (.35-.49)	.29 (.19-.39)	.29 (.20-.37)
Surgeon E versus A-D	.47 (.39-.54)	.38 (.30-.47)	.20 (.14-.27)
PEERS			
Overall agreement	.38 (.34-.41)		.22 (.19-.26)
Peer A versus B-E	.35 (.27-.43)		.23 (.15-.31)
Peer B versus A, C-E	.47 (.41-.53)		.22 (.14-.30)
Peer C versus AB, DE	.39 (.32-.46)		.23 (.16-.31)
Peer D versus A-C, E	.38 (.31-.46)		.30 (.23-.37)
Peer E versus A-D	.44 (.37-.51)		.23 (.15-.30)
Surgeons versus peers ^b			
	.54 (.48-.60)		.37 (.30-.45)
^a κ (95% CI)			
^b Arithmetic mean of surgeons versus arithmetic mean of peers			
^c PABAK (95% CI)			

Table 2.

	Severity before treatment		Results after treatment ^a
	Severity ^a	Symmetry ^a	
SURGEONS			
Average agreement	.54 (.47-.61)	.60 (.46-.74)	.49 (.40-.58)
Surgeon A versus A	.40 (.28-.51)	.75 (.66-.84)	.50 (.37-.63)
Surgeon B versus B	.55 (.45-.64)	.63 (.52-.73)	.56 (.46-.67)
Surgeon C versus C	.59 (.41-.77)	.71 (.52-.89)	.61 (.41-.81)
Surgeon D versus D	.59 (.49-.68)	.34 (.21-.47)	.34 (.24-.45)
Surgeon E versus E	.55 (.45-.65)	.58 (.47-.69)	.44 (.34-.54)
PEERS			
Average agreement	.48 (.38-.59)		.34 (.28-.40)
Peer A versus A	.34 (.23-.46)		.22 (.11-.34)
Peer B versus B	.61 (.53-.70)		.36 (.24-.47)
Peer C versus C	.50 (.40-.59)		.40 (.29-.50)
Peer D versus D	.38 (.28-.48)		.34 (.23-.44)
Peer E versus E	.60 (.51-.70)		.37 (.26-.37)
^a κ (95% CI)			

Table 3.

	Severity before treatment				Results after treatment			
	<i>Overall</i> <i>l</i>	<i>DCS-</i> <i>bracing</i>	<i>Ravitch</i> <i>h</i>	<i>p-value</i>	<i>Overall</i> <i>l</i>	<i>DCS-bracing</i>	<i>Ravitch</i>	<i>p-value</i>
Surgeon A	1.67	1.60	1.84	.016	2.71	2.65	2.94	.002
Surgeon B	1.83	1.63	2.26	<.001	2.50	2.37	2.79	<.001
Surgeon C	1.97	1.76	2.44	<.001	2.33	2.12	2.79	<.001
Surgeon D	2.00	1.76	2.56	<.001	2.51	2.41	2.74	.004
Surgeon E	2.08	1.91	2.45	<.001	1.92	1.78	2.21	<.001
Peer A	2.33	2.22	2.60	<.001	2.34	2.30	2.44	.004
Peer B	2.21	2.01	2.66	<.001	2.33	2.38	2.21	.006
Peer C	1.84	1.65	2.26	<.001	2.05	2.01	2.15	.002
Peer D	1.96	1.83	2.24	.002	2.33	2.27	2.48	.020
Peer E	2.20	2.03	2.58	<.001	2.09	2.03	2.23	<.001
<i>DCS-bracing = Dynamic Compression System Bracing</i>								

Table 4.

	Scars versus no scars
SURGEONS	
Average agreement	.89 (.85-.94) ^b
Surgeon A	.90 (.81-.99) ^a
Surgeon B	.97 (.90-1.0) ^a
Surgeon C	.90 (.78-1.0) ^a
Surgeon D	.84 (.70-.98) ^a
Surgeon E	.86 (.75-.97) ^a
PEERS	
Average agreement	.39 (.19-.60) ^b
Peer A	.67 (.50-.83) ^b
Peer B	.27 (.14-.40) ^b
Peer C	.62 (.43-.80) ^b
Peer D	.20 (.06-.34) ^b
Peer E	.20 (.09-.31) ^b
^a PABAK (95% CI)	
^b κ (95% CI)	

Evaluating Inter- and Intraobserver Agreement on Pectus Carinatum Severity and Treatment Outcomes: A Comparison of Subjective and Objective Assessment Methods

Hendrik van Braak MD^{a,b}, Sjoerd A. de Beer MD^a, Sander Zwaveling MD, PhD^a, Matthijs W. N. Oomen MD, PhD^a, L.W. Ernest van Heurn MD, PhD^{a,b}, Justin R. de Jong MD, PhD^a

^a Department of Pediatric Surgery, Amsterdam University Medical Center, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands.

^b Department of Surgery, Leiden University Medical Center, Albinusdreef 2, 2333ZA Leiden, The Netherlands

Corresponding author: Hendrik van Braak^a, h.vanbraak@amsterdamumc.nl, Department of Pediatric Surgery, Emma Kinderziekenhuis Amsterdam University Medical Center, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands. Room H7-214

