Subgroups of failure after surgery for pelvic organ prolapse and associations with quality of life outcomes: a longitudinal cluster analysis

J. Eric JELOVSEK, MD, MMEd, MSDS[1], **Marie G. GANTZ, PhD**[2], **Emily S. LUKACZ, MD**[3], **Halina M. ZYCZYNSKI, MD**[4], **Amaanti SRIDHAR, MS**[2], **Caroline KERY, MS**[5], **Rob CHEW, MS**[5], **Heidi S. HARVIE, MD, MSCE**[6], **Gena DUNIVAN, MD**[7], **Joseph SCHAFFER, MD**[8], **Vivian SUNG, MD**[9], **R. VARNER, MD**[10], **Donna MAZLOOMDOOST, MD**[11], **Matthew D. BARBER, MD, MHS**[11] **NICHD Pelvic Floor Disorders Network**

[1]Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina

[2]Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, North Carolina

[3]Department of Obstetrics and Gynecology, University of California San Diego, San Diego, California

[4]Department of Obstetrics and Gynecology, University of Pittsburgh/Magee-Womens Research Institute, Pittsburgh, Pennsylvania

[5]Division for Statistical and Data Sciences, RTI International, Research Triangle Park, North Carolina

[6]Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, Pennsylvania

[7]Department of Obstetrics and Gynecology, University of New Mexico, Albuquerque, New Mexico

[8]Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, Texas

[9]Department of Obstetrics and Gynecology, Brown University/Women & Infants Hospital of Rhode Island, Providence, Rhode Island

Corresponding Author: J. Eric Jelovsek, MD, MMEd, MSDS, DUMC 3084, Durham, NC 27710, eric.jelovsek@duke.edu phone: 919-681-3864.

[10]Department of Obstetrics and Gynecology, University of Alabama at Birmingham, Birmingham, Alabama

[11]NICHD

Abstract

Background: Treatment outcomes after pelvic organ prolapse (POP) surgery are often presented as dichotomous 'success or failure' based upon anatomic and symptom criteria. However, clinical experience suggests some women with outcome 'failures' are asymptomatic and perceive their surgery to be successful, while others have anatomic resolution, but continue to report symptoms. Characterizing failure types could be a useful step to refine definitions of success, understand mechanisms of failure, and identify individuals who may benefit from specific therapies.

Objectives: To identify clusters of women with similar failure patterns over time and assess associations between clusters and the Pelvic Organ Prolapse Distress Inventory (POPDI), Short-Form Six-Dimension health index (SF-6D), patient global impression of improvement (PGI-I), patient satisfaction items questionnaire (PSIQ), and quality-adjusted life years (QALYs).

Study Design: Outcomes were evaluated for up to 5 years in a cohort of participants (N=709) with stage 2 or greater POP who underwent surgical POP repair and had sufficient follow-up in one of 4 multi-center surgical trials conducted by the NICHD Pelvic Floor Disorders Network. Surgical success was defined as a composite measure requiring anatomic success (Pelvic Organ Prolapse Quantification system points Ba, Bp, and C ≤ 0), subjective success (absence of bothersome vaginal bulge symptoms), and absence of retreatment for POP. Participants who experienced surgical failure and attended ≥ 4 visits from baseline through 60 months after surgery were longitudinally clustered accounting for similar trajectories in Ba, Bp, and C, and degree of vaginal bulge bother; missing data were imputed. Participants with surgical success were grouped into a separate cluster.

Results: Surgical failure was reported in 39% (276/709) of the women included in the analysis. Failures clustered into 4 mutually exclusive subgroups: A) asymptomatic intermittent anterior wall failures, B) symptomatic intermittent anterior wall failures, C) asymptomatic intermittent anterior and posterior wall failures and D) symptomatic all-compartment failures. Each cluster had different bulge symptoms, anatomy, and retreatment associations with quality of life outcomes. Asymptomatic intermittent anterior wall failures (N=150) were most similar to surgical successes with Ba values that averaged around −1 cm but fluctuated between anatomic success (Ba ≤ 0) and failure (Ba > 0) over time. Symptomatic intermittent anterior wall failures (N=82) were anatomically similar to asymptomatic intermittent anterior failures but women in this cluster persistently reported bothersome bulge symptoms and the lowest QOL, SF-6D scores, and perceived success. Women with asymptomatic intermittent anterior and posterior wall failures (N=28) had the most severe preoperative POP but the lowest symptomatic failure rate and retreatment rate. Participants with symptomatic all-compartment failures (N=16) had symptomatic and anatomic failure early after surgery and the highest retreatment of any cluster.

Conclusions: Four clusters of POP surgical failure were identified in participants up to 5 years after POP surgery: asymptomatic intermittent anterior wall failures, symptomatic intermittent anterior wall failures, asymptomatic intermittent anterior and posterior wall failures,

and symptomatic all-compartment failures. These groups provide granularity about the nature of surgical failures after POP surgery. Future work is planned for predicting these distinct outcomes using patient characteristics that can be used for counseling individual women.

Condensation:

Longitudinal clustering of pelvic organ prolapse surgery outcomes identified four distinct failure types in participants up to five years after surgery.

Keywords

pelvic organ prolapse; surgical outcomes; success definition; failure definition; failure subtypes; machine learning; clustering; quality of life; quality adjusted life years

Introduction

Significant variation exists in how outcomes of pelvic organ prolapse (POP) surgery are reported.[1] Treatment outcomes after POP surgery are often presented as dichotomous 'success or failure' based on anatomic criteria or, more recently, both anatomic and symptom criteria combined.[1, 2] However, clinical experience suggests some women with 'failures' are asymptomatic and perceive their surgery to be successful, while others, whose anatomy is 'successfully' improved, continue to report symptoms. These paradoxical findings were highlighted by investigators who explored 18 different surgical success definitions in women two years after abdominal sacrocolpopexy.[1] Despite concluding that bulge symptoms should be included in any definitions of success after POP surgery due to their strong association with the patients' assessment of overall improvement and treatment success,[1] the researchers recognized that 17% of participants who reported vaginal bulge symptoms demonstrated good anatomic support (stage 0 or 1).[1] They also noted that the clinical relevance of asymptomatic stage 2 prolapse was unclear since it was unknown whether these patients have symptomatic progression over time.[1]

The objective of this study was to better understand the heterogeneity of outcomes across time for women surgically treated for POP and determine whether clinically relevant failure types could be identified. Specifically, we aimed to identify whether there were common patterns, or clusters, using each participant's trajectory of anatomic findings and bulge symptoms through time. To assess whether the clustering was meaningful from the patient's perspective, clusters were compared with condition-specific symptom and generic health-related quality of life instruments, patient global impression of improvement patient satisfaction items questionnaire, and quality-adjusted life years.

Material and Methods

This was a retrospective study of women enrolled in one of four prospective surgical trials conducted across 17 centers in the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Pelvic Floor Disorders Network (PFDN). All participants planned to undergo reconstructive surgery for stage 2-4 pelvic organ prolapse and were assessed for recurrent prolapse using standardized validated measures. The study design

and results of the primary and extended studies have been published.[3-8] The Colpopexy and Urinary Reduction Efforts (CARE) trial enrolled stress-continent women undergoing abdominal sacrocolpopexy surgery for prolapse between March 2002 and February 2005.[3] The intervention evaluated the effectiveness of prophylactic Burch cystourethropexy continence surgery versus no Burch in reducing de novo stress urinary incontinence. CARE participants who completed two-year follow-up were offered enrollment in the extended CARE study with up to 7 additional years of follow-up.[4] The Outcomes Following Vaginal Prolapse Repair and Midurethral Sling trial (OPUS) trial enrolled stress continent women undergoing vaginal prolapse surgery between May 2007 and October 2009.[5] The trial evaluated the effectiveness of prophylactic retropubic midurethral sling versus sham in reducing de novo stress urinary incontinence, and participants were followed for one year postoperatively. The OPUS trial also included women who declined randomization, but who participated in a patient-preference cohort. The Operations and Pelvic Muscle Training in the Management of Apical Support Loss (OPTIMAL) trial compared 2-year outcomes in women undergoing vaginal apical prolapse repair with midurethral sling for stress urinary incontinence.[6, 9] Participants enrolled between January 2008 and March 2011 and were randomized in a 2×2 factorial design to 2 intervention arms: (1) perioperative behavioral therapy with pelvic floor muscle training versus usual care and (2) surgical intervention (either uterosacral ligament suspension or sacrospinous ligament suspension). Participants who completed 2 year follow-up were invited to enroll in the 3-year extended trial for a total of 5 years of follow-up.[7] The Study of Uterine Prolapse Procedures Randomized Trial (SUPeR) trial compared the efficacy and adverse events of vaginal hysterectomy with uterosacral ligament suspension versus vaginal mesh hysteropexy.[8] The study enrolled women between April 2013 and February 2015. The primary endpoint was 3 years after the last randomization, with the earliest enrolled participants followed for up to 5 years. All studies received institutional review board approval at each site and all participants provided written informed consent.

For this analysis, the outcome of surgical success was defined as a composite measure requiring: 1) anatomic success (Pelvic Organ Prolapse Quantification system[10] (POPQ) Ba, Bp, and C 0); 2) subjective success (absence of bothersome vaginal bulge symptoms); and 3) absence of retreatment for POP.[1] Surgical failure was defined as the occurrence of an anatomic failure, subjective failure, or retreatment for pelvic organ prolapse with surgery or pessary. Anatomic failure was defined as POP of one or more compartments beyond the hymen (i.e., POPQ point Ba, Bp, or C > 0). Subjective failure was defined as an affirmative response to either of the Pelvic Floor Distress Inventory (PFDI) questions "Do you usually have a sensation of bulging or protrusion from the vaginal area?" or "Do you usually have a bulge or something falling out that you can see or feel in the vaginal area?" with a degree of bother more than "Not at all." [11, 12] Relevant outcomes were evaluated at 3 or 6, 12, 24, 36, 48, and 60 months after surgery. The 3-month visit for participants in OPUS and CARE studies were combined with the 6-month visit for subjects in OPTIMAL and SUPeR studies.

Secondary outcomes included pelvic organ prolapse distress inventory (POPDI) subscale scores from the Pelvic Floor Distress Inventory (PFDI)-46 or PFDI-20 short form, Patient Global Impression of Improvement (PGII) or Patient Satisfaction Items Questionnaire (PSIQ) score, and Short-Form Six-Dimension health index (SF-6D) scores.[11-15] The PGII

was collected at follow-up visits in OPUS and OPTIMAL. This question asked, "What best describes how your bladder function is now, compared to how it was before you had prolapse surgery?" In the SUPeR trial, the question was worded as follows, "Check the number that best describes how your post-operative condition is now, compared with how it was before you had the surgery." The PSIQ was collected at follow-up visits in CARE. This question asked, "Compared to how you were doing before your recent pelvic floor operation, would you say that now you are:" The response items were similar, so a combined overall success measure (PGII/PSIQ) was derived with 5 levels ranging from l='Much Better or Very Much Better' to 5='Much Worse or Very Much Worse' by collapsing the tail end values on the PGII. Quality-adjusted life years (QALYs) were calculated using an area under the curve approach following the trapezoidal rule from each subject's SF-6D Index score reported at baseline, 3 or 6 months, 12 months, and 24 months. If the SF-6D Index score was missing at 24 months, it was assumed to be the same as at 12 months.

Participants were included in analysis if they received reconstructive surgery and either attended at least 4 visits at baseline through 60 months after surgery or were retreated for POP at any time after surgery. Patients receiving colpocleisis were excluded. Since 6 follow up time points were analyzed, participants who were not retreated and attended fewer than 3 postoperative visits were excluded to avoid imputing more than half of the outcomes. Participants who had surgical failures at any follow-up time point were longitudinally clustered based on joint trajectories of POPQ points Ba, Bp, C, and bulge symptoms at baseline, 3 or 6, 12, 24, 36, 48, and 60 months. Missing data were imputed using the copyMean algorithm for anatomic and subjective failures, and values at the last visit prior to retreatment were carried forward for retreatment failures.[16]

Clustering is the process of finding distinct groups of observations based on their covariate similarity.[17] Clustering of participants was performed using an implementation of the k-means algorithm designed to work specifically on trajectories in longitudinal data from the Kml3d package in R. K-means is a clustering algorithm that assumes there are "k" mutually exclusive clusters present in a dataset. The algorithm first chooses "k" points as the initial cluster centers, called centroids, through random assignment. An iterative process of assigning observations to clusters based on the closest centroid and updating centroids by calculating the mean of all of the points assigned to the cluster then occurs until the centroids are stable. In the longitudinal setting, instead of each subject being represented by a vector of covariates, each is represented as a matrix of multiple outcome trajectories. Distances between matrices are computed by independently finding (1) the distances between outcome trajectories ("line-distances") and (2) the distances between visits, for the outcome cross-sections ("column-distances"). The two distance vectors are combined by taking the standard p-norm to capture dependency between both measurements at different visits and between different outcomes. Additional details on choosing an optimal number of clusters and for imputation with longitudinal data are reported in the literature.[16, 18] Participants whose outcomes met success criteria at all time points were grouped into a 'persistent successes' cluster.

POPDI subscale scores, SF-6D scores, PGII/PSIQ, and QALYs were compared between clusters using Fisher's Exact Test for categorical measures or general linear models

for continuous outcomes. Outcomes assessed at more than one follow up time point were compared between clusters using general linear mixed models with independent variables for cluster, visit, interaction between cluster and visit, and modeling within-subject correlation using an unstructured covariance structure. All tests were conducted at a significance level of 0.05.

Results

The flow of participants for this study originating from each of the primary studies is demonstrated in Figure 1. A total of 1337 eligible participants were enrolled into their respective studies and 1297 of these participants underwent reconstructive surgery without colpocleisis. An additional 588 participants were excluded who did not undergo retreatment for POP and had insufficient follow-up of the POPQ exam or completion of the PFDI. The majority (418/588) of these excluded participants were from the OPUS trial since the study only followed participants for one year; therefore, OPUS participants were only included for this study if they were retreated. The analysis cohort consisted of 709 total participants including 433 (61%) whose outcomes met the definition of surgical success across all follow-up time points and 276 (39%) whose outcomes met criteria for surgical failure at least once during follow-up. The numbers of participants at each follow-up visit for each study are shown in Figure 1.

Clustering of the 276 women with surgical failures and 433 with successes resulted in five mutually exclusive groups labelled descriptively as follows:

- Cluster A: Asymptomatic Intermittent Anterior Wall Failures (N=150)

- Cluster B: Symptomatic Intermittent Anterior Wall Failures (N=82)

- Cluster C: Asymptomatic Intermittent Anterior and Posterior Wall Failures (N=28)

- Cluster D: Symptomatic All-Compartment Anatomic Failures (N=16)

- Cluster E: Persistent Successes (N=433)

Figure 2 displays the mean clustered pattern of the anatomic measures and bulge symptoms over time (individual patterns are illustrated in Appendix Figure 1). Table 1 shows the percentage of participant evaluations (summed over all time points) that met each individual definition of failure (anatomic, subjective, and retreatment criteria) by cluster. Percentages based on completed (non-missing) visits and those based on imputed data were consistent, with the biggest variations due to carrying forward of retreatment failures to subsequent visits (Table 1). Cluster A failures were most similar to persistent successes (Cluster E) and were predominately asymptomatic, intermittent anterior wall failures with Ba values that averaged around −1 but fluctuated between success (Ba ≤ 0) and failure (Ba > 0) over time (Figure 2, upper left panel, Appendix Figure 1A). Cluster B failures were similar to Cluster A anatomically (Figure 2, top panels and bottom left panel, Appendix Figure 1A-C) but, on average, reported bothersome bulge symptoms (Figure 2, bottom right panel, Appendix Figure 1D). Cluster C failures consisted mostly of asymptomatic, intermittent anterior and posterior wall failures with Ba and Bp values that fluctuated between success and failure for individual women (Figure 2, top panels, Appendix Figure 1A-B). This subgroup had the

most severe preoperative POP and low symptomatic failure rate and retreatment rate. Cluster D failures had frequent symptomatic and anatomic failure in all vaginal compartments with onset early after surgery and the highest retreatment of any cluster (Figure 2, Appendix Figure 1A-D, Table 1). Cluster A and Cluster C asymptomatic failures comprised 178/276, 64.5% of total failures in the cohort.

Adjusted analyses of mean symptom bother, quality of life outcomes and QALYs by cluster assignment are shown in Table 2 and individual scores are illustrated in Appendix Figure 2. Adjusted mean scores were significantly different between clusters for all outcomes. Despite the Symptomatic Intermittent Anterior Wall Failure subgroup (Cluster B) having mean postoperative POPQ measures above the hymen, women in the cluster reported the highest degree of prolapse symptom distress (highest POPDI Scores) compared to all others (Table 2, Appendix Figure 2A). Women in Cluster B also reported the worst generalized health-related QOL, and PGII/PSIQ scores (Table 2, Appendix Figures 2B-2I). In contrast, participants in the Asymptomatic Intermittent Anterior Wall Failure subgroup (Cluster A) and Asymptomatic Intermittent Anterior and Posterior Wall Failure subgroup (Cluster C) reported prolapse symptom distress scores that were most similar to the Persistent Success group (Cluster E), though the average scores in Clusters A and C increased (worsened) over time (Table 2, Appendix Figure 2A). Women in the Symptomatic All-Compartment Anatomic Failures subgroup (Cluster D) reported higher prolapse symptom distress scores, and worse PGII/PSIQ scores, than all other groups besides Cluster B (Table 2, Appendix Figures 2A and 2I).

Structured Comment

Principal Findings

In this retrospective study of 709 participants of 4 PFDN surgical trials we used longitudinal clustering to identify four mutually exclusive clinical failure profiles as well as a group whose outcomes persistently met all criteria for success. The failure profiles in order of predominance were: asymptomatic intermittent anterior wall failure, symptomatic anterior wall failure, asymptomatic intermittent anterior and posterior wall failure, and symptomatic all-compartment failure. Each failure profile had distinct associations with condition-specific patient-reported outcomes, SF-6D scores, global impression of improvement or satisfaction, and QALYs, further corroborating the failure profiles as real and meaningful from the patient perspective. These categorizations of failure provide nuanced, clinically relevant descriptions of surgical outcomes that were lost with application of the dichotomous composite definition of success or failure.

Results

Historically, surgery for POP has largely been judged by anatomic outcomes. With increased availability of validated patient-reported outcome measures for women with pelvic floor disorders, the impact of surgery on symptom resolution has been incorporated into outcome assessments and improved our understanding of the patient experience after POP surgery. In 2009, investigators from the PFDN recommended that any definition of success after POP surgery should include the absence of bulge symptoms in addition to anatomic

criteria and the absence of retreatment.[1] Since then, many high quality trials have used a composite definition of surgical success for prolapse procedures that defined success as absence of anatomic or subjective prolapse or retreatment. This rigorous standard applied to randomized controlled trials resulted in failure rates that confounded clinicians and investigators who were challenged to reconcile the discord between high failure rates (26-38%) but overall high patient reported satisfaction (89-90%) endorsed by low retreatment rates (5-12%).[8, 9]

Clinical Implications

Our findings that the majority of women with outcomes qualifying as surgical failures were asymptomatic (178/276, 64.5% in Cluster A and Cluster C) may be reassuring to patients contemplating treatment choices that are currently reported to have high surgical failure rates. This confirms previous work which showed that treatment success from the patient's perspective, defined as absence or improvement in prolapse symptoms, correlates only weakly with anatomic success. (1)

Identifying specific subcategorizations of failure may allow for a better understanding of factors which can influence surgical outcomes. We found that patients with asymptomatic intermittent anterior and posterior wall failures (Cluster C) had the most severe preoperative POP but low symptomatic failure rate and retreatment rate. This corresponds with the authors' clinical impression that those patients with the worst anatomic prolapse often have excellent treatment success from the patient's perspective even though they might have anatomic surgical failure by strict criteria. This information is important in counselling patients about success and failure and when setting appropriate expectations of surgical outcomes. Similarly, those in Cluster D demonstrated anatomic and symptomatic failures early after surgery suggesting that either there was a technical failure with the surgery or that this group may be at particular high risk of failure, impacting preoperative counseling and choice of surgical approach.

Research Implications

Employing the proposed subgroups of failure in future studies provides opportunity to identify clinical characteristics prognostic of each failure experience. This descriptive information of expected outcomes could guide surgeon's advice and patient's decisions surrounding treatment choices. For example, identifying factors which distinguish patients at high risk for Cluster A (asymptomatic intermittent anterior failure) versus Cluster B (symptomatic intermittent anterior failure) may help identify those who might benefit from pelvic floor therapy combined with surgery or a different surgical procedure. Similarly, identifying those at high risk of persistent symptoms in spite of anatomic restoration (Cluster B) will aid with surgical counseling and may discourage patients from proceeding with surgical correction unlikely to improve their symptoms. For example, vaginal bulge, pressure, and heaviness symptoms are common in women with myofascial pain conditions and pelvic floor dysfunction and identifying these conditions in women who are planning surgery for pelvic organ prolapse could help providers counsel them about expected changes in those symptoms after surgery.

Likewise, women at high risk for the most severe failure, Cluster D, may need to be counseled that they are at highest risk for reoperation and they may consider selecting an augmented repair such as primary sacrocolpopexy over native tissue vaginal reconstruction. Clusters may also be differentially represented after a surgical approach. For example, a woman may have a high probability of being in Cluster A (asymptomatic intermittent anterior wall prolapse beyond the hymen) after vaginal approaches or a high probability of being in Cluster B (symptomatic anterior wall prolapse) after sacrocolpopexy, although this was not determined from our analysis. Further investigation of the clinical and surgical characteristics associated with each failure cluster will improve our understanding of treatment response by approach. It may also help us better understand the mechanisms of failure, whether symptomatic or anatomic.

Strengths and Limitations

A strength of this study is the use of data from four well designed, high quality multi-center clinical trials that used multiple validated outcome measures. Similarly, our use of cluster analysis to identify mutually exclusive failure groups in this patient population is novel for POP. The fact that each of these clusters have unique patterns of health care outcomes like health-related quality of life and QALYs strengthens our findings. One of the limitations of our studies is that we included and combined trials for this analysis with differing treatments and lengths of follow-up. For example, OPUS trial participants were only followed for one year; therefore, only OPUS participants who were retreated (N=9) were included in this analysis. Also, outcome measures, such as the PGII/PSIQ, used in these trials had significant but not complete overlap. Similarly, some commonly used POP surgeries, particularly laparoscopic or robotic sacrocolpopexy, were not represented in this dataset. In spite of these limitations, our findings provide significant new insight into patient's experience after POP Surgery.

Conclusions

In summary, our analyses have identified 5 distinct types of surgical outcomes up to 5 years after prolapse surgery that distinguish between wholistic and persistent success, asymptomatic intermittent anterior wall failures, symptomatic intermittent anterior wall failures, asymptomatic intermittent dual compartment failures, and all compartment failures. These groups provide information about how POP surgeries fail based on our definitions and may be useful in counseling women and refining definitions of success after POP surgery. Future work is planned for predicting the distinct surgical outcomes using patient characteristics known before surgery that can be used for counseling individual women.
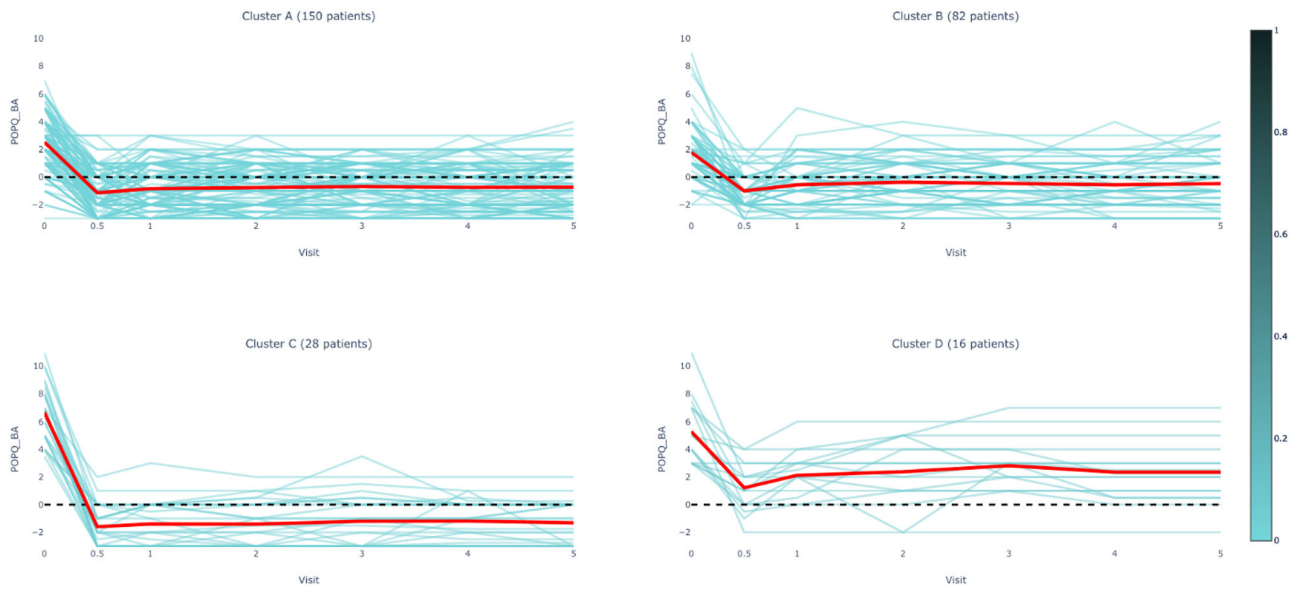
Financial Support for the research:

Appendix

Figure 1A.
POPQ Ba Measurements by Cluster Assignment and Visit
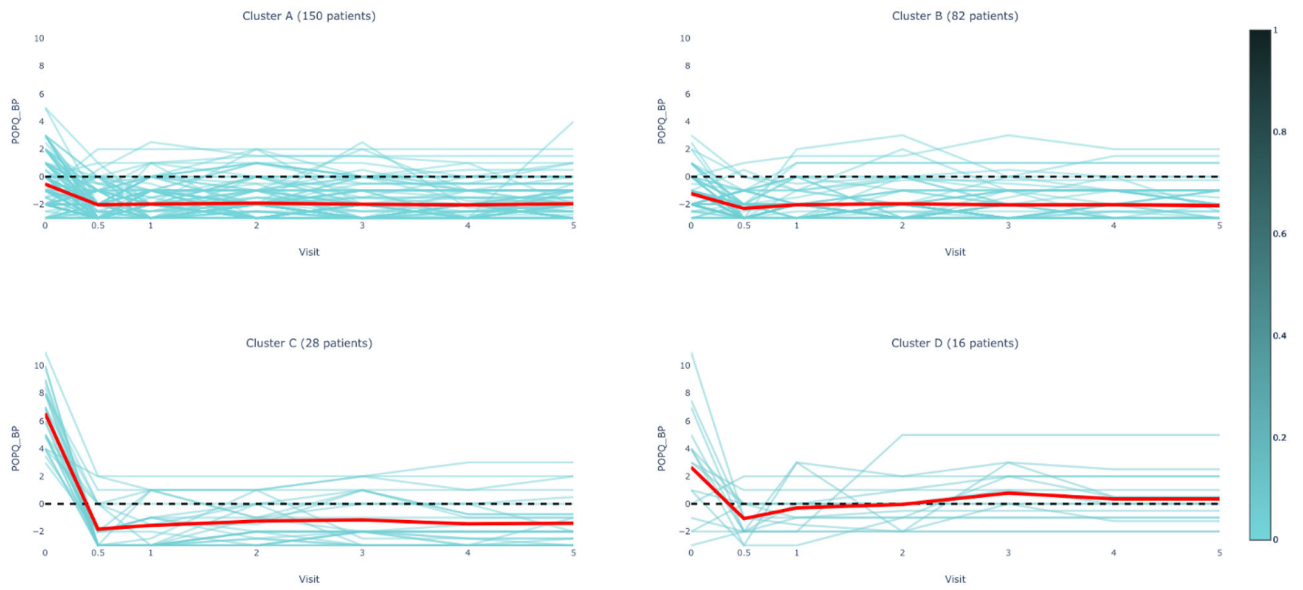
Figure 1B.
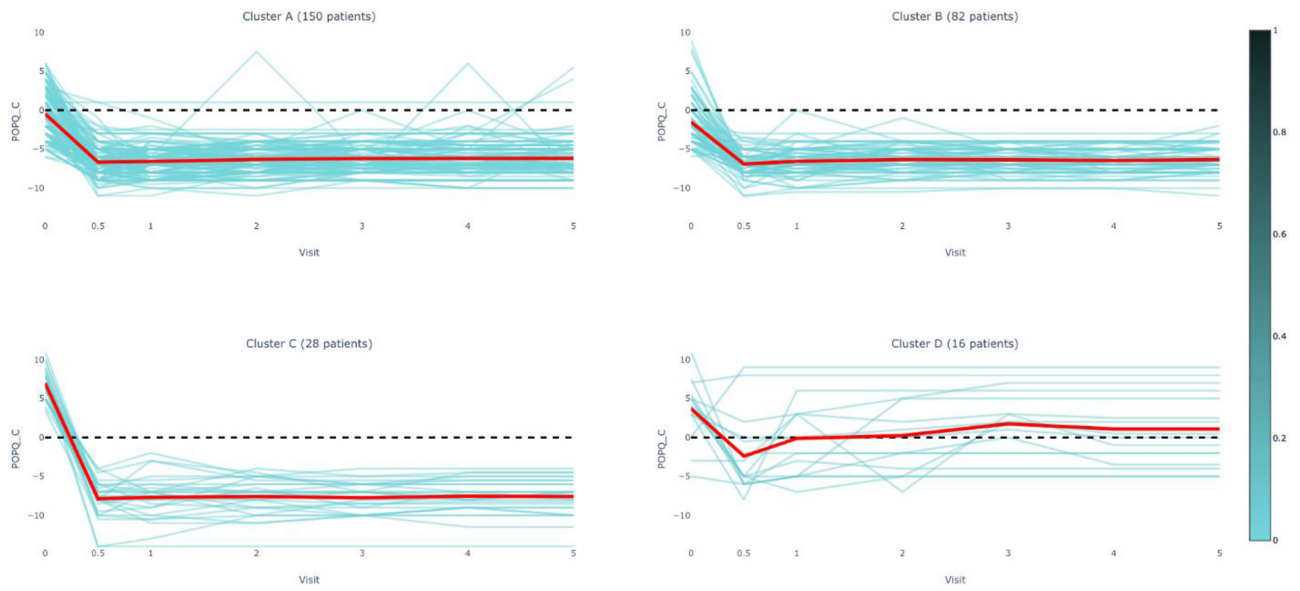POPQ Bp Measurements by Cluster Assignment and Visit

Figure 1C.
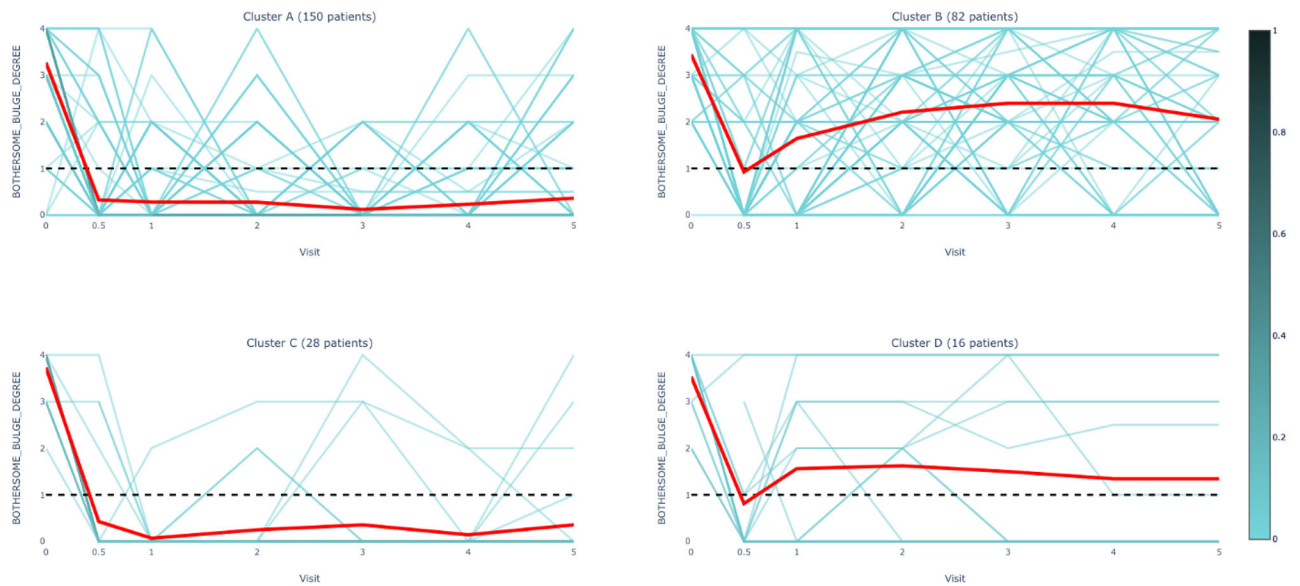POPQ C Measurements by Cluster Assignment and Visit

Figure 1D.
Degree of Bothersome Bulge Symptoms by Cluster Assignment and Visit
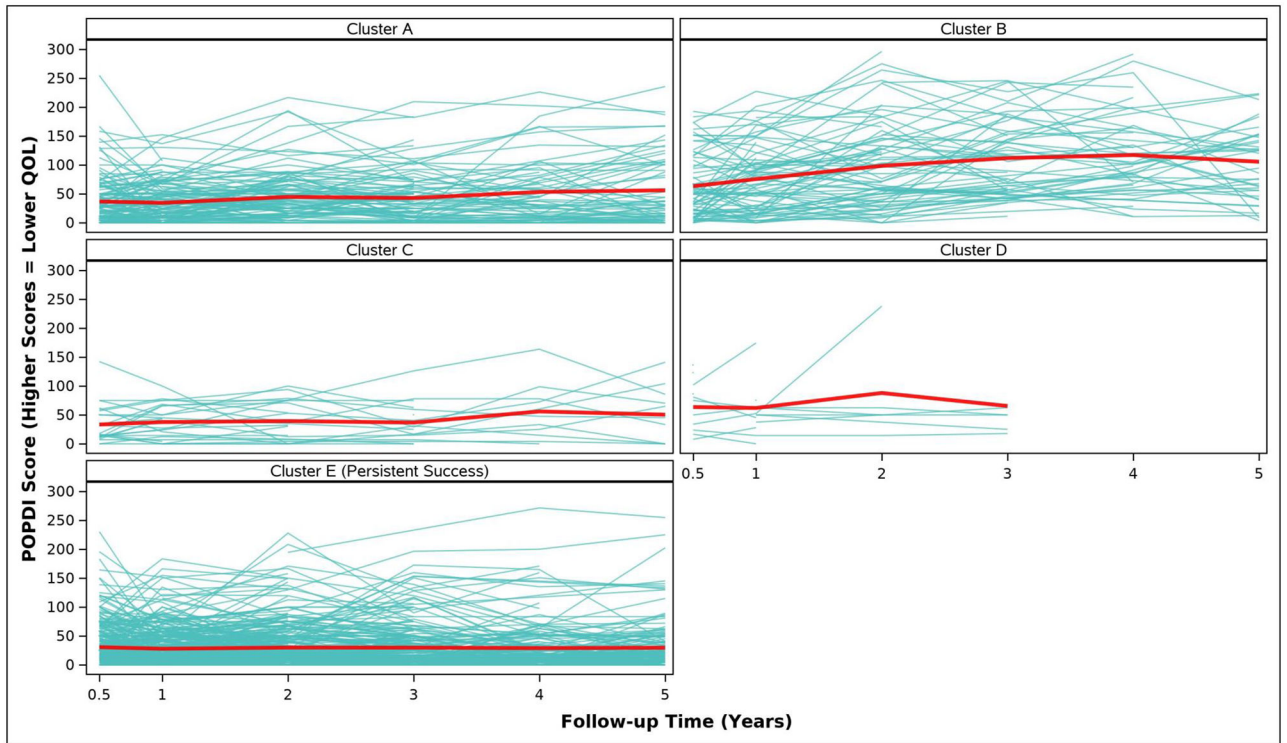
Figure 2A.
POPDI Score by Cluster Assignment and Visit

Figure 2B.

SF-6D Physical Function Score by Cluster Assignment and Visit

Figure 2C.
SF-6D Role Limitation Score by Cluster Assignment and Visit

Figure 2D.
SF-6D Social Function Score by Cluster Assignment and Visit

Figure 2E.

SF-6D Pain Score by Cluster Assignment and Visit

Figure 2F.
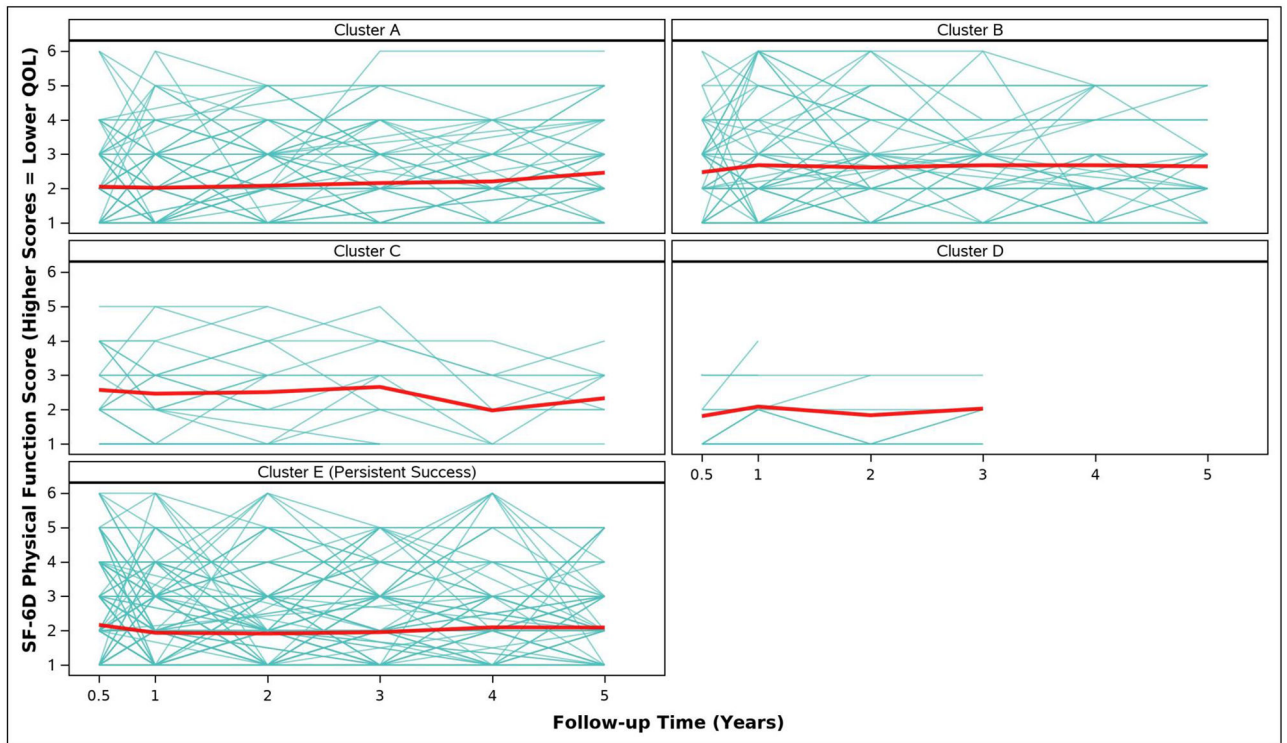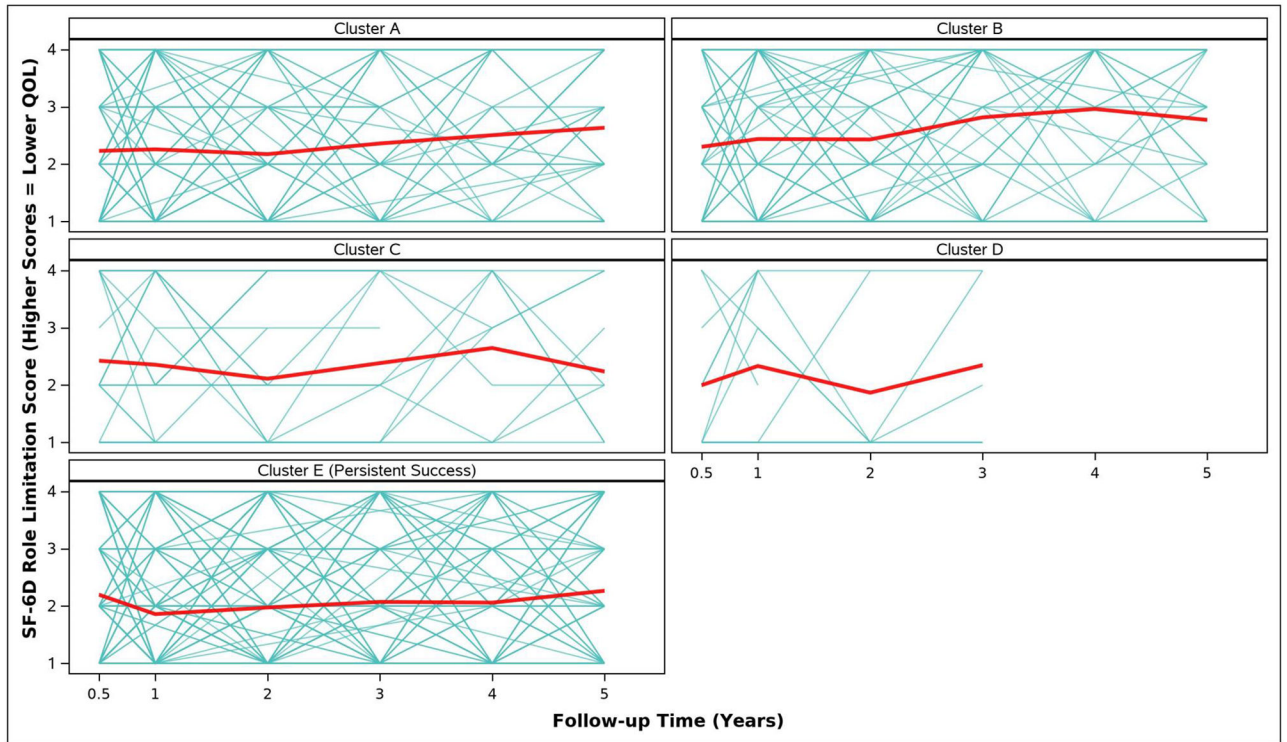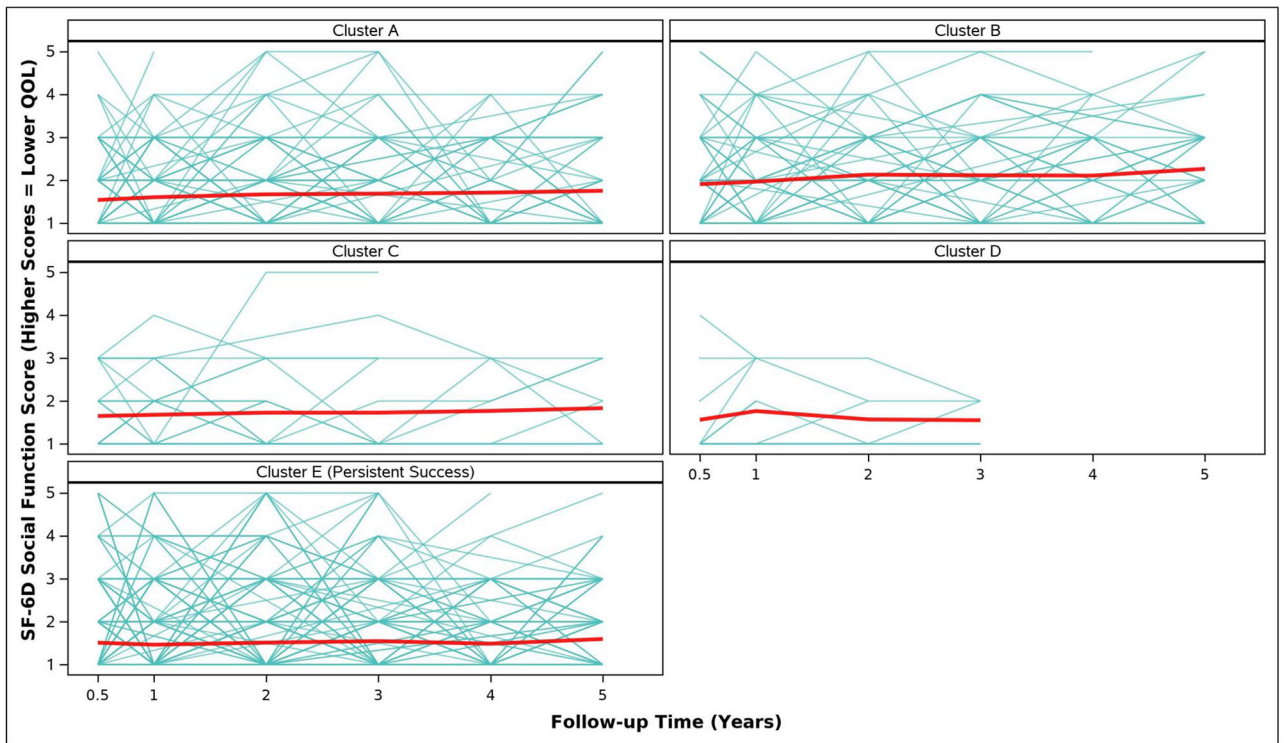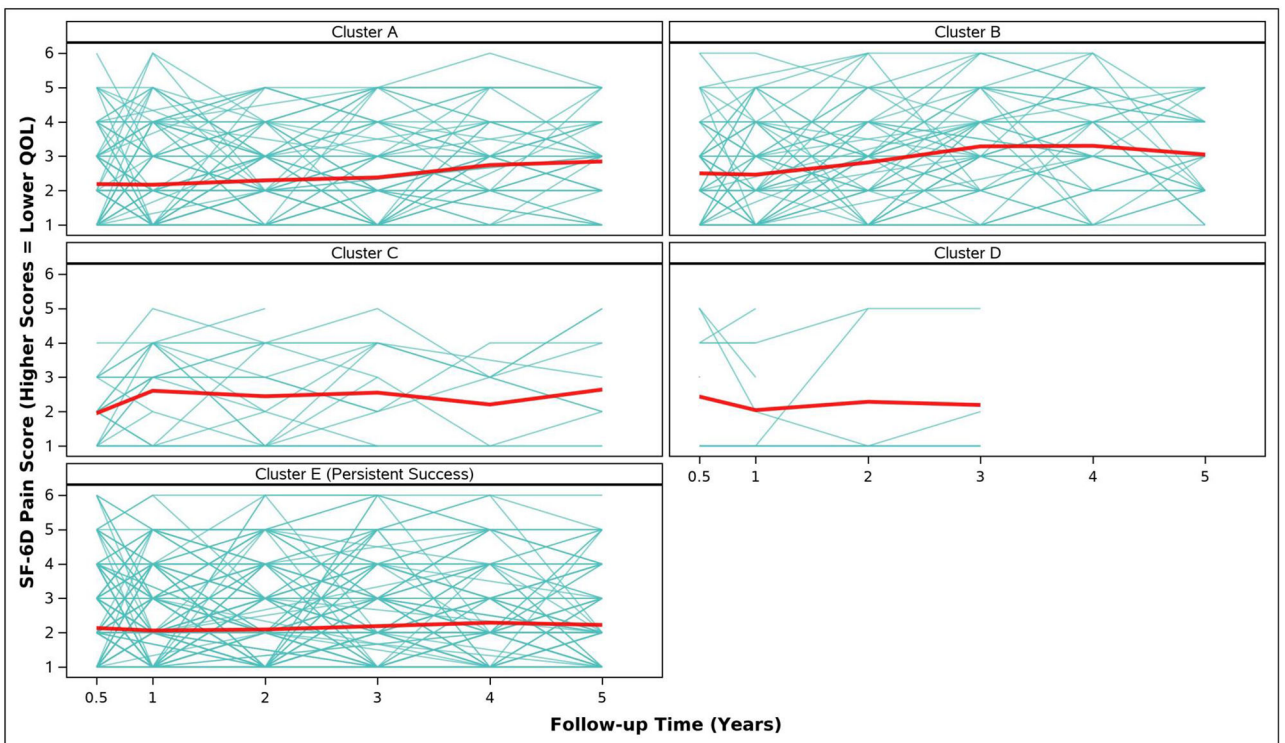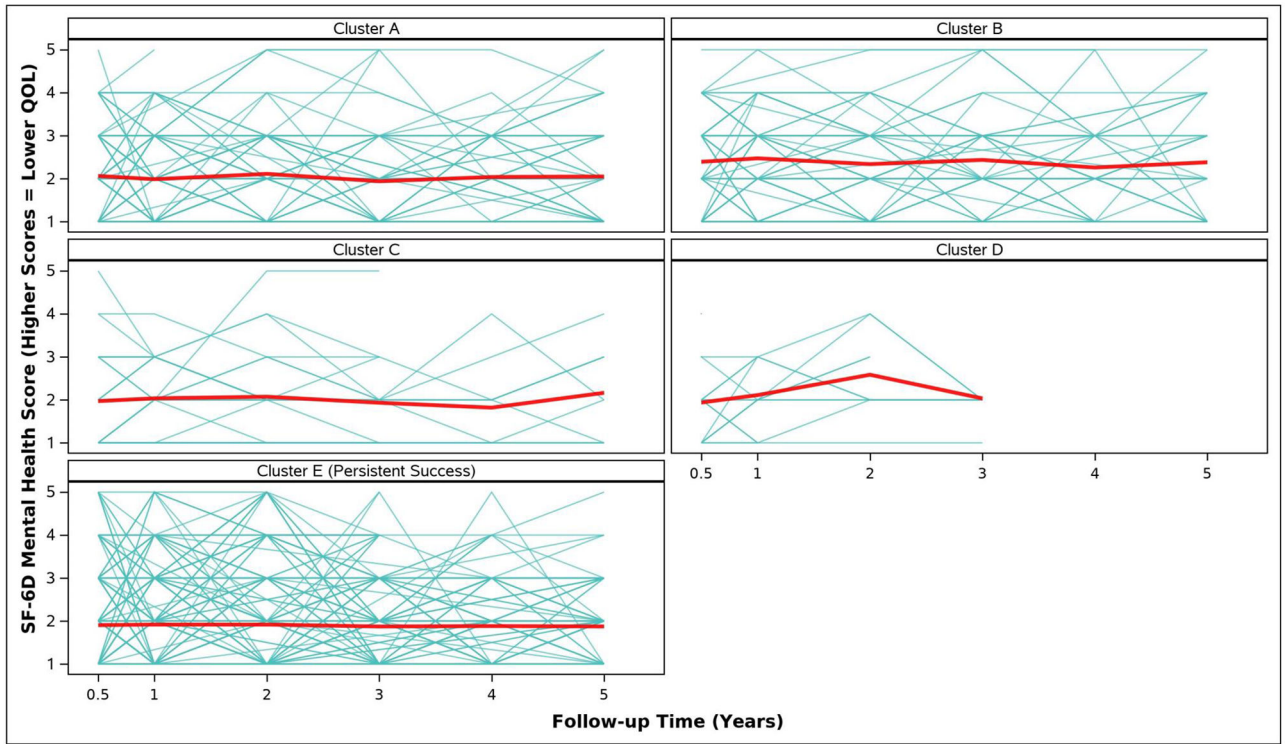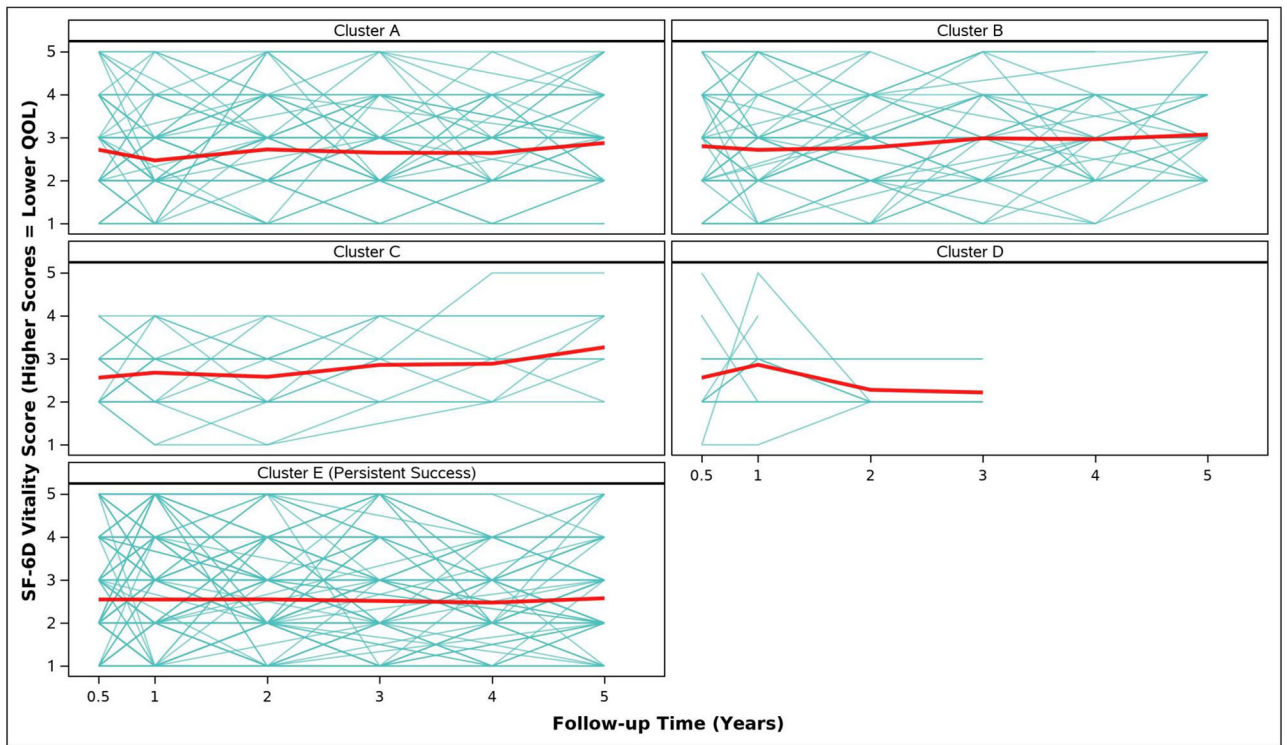SF-6D Mental Health Score by Cluster Assignment and Visit

Figure 2G.
SF-6D Vitality Score by Cluster Assignment and Visit
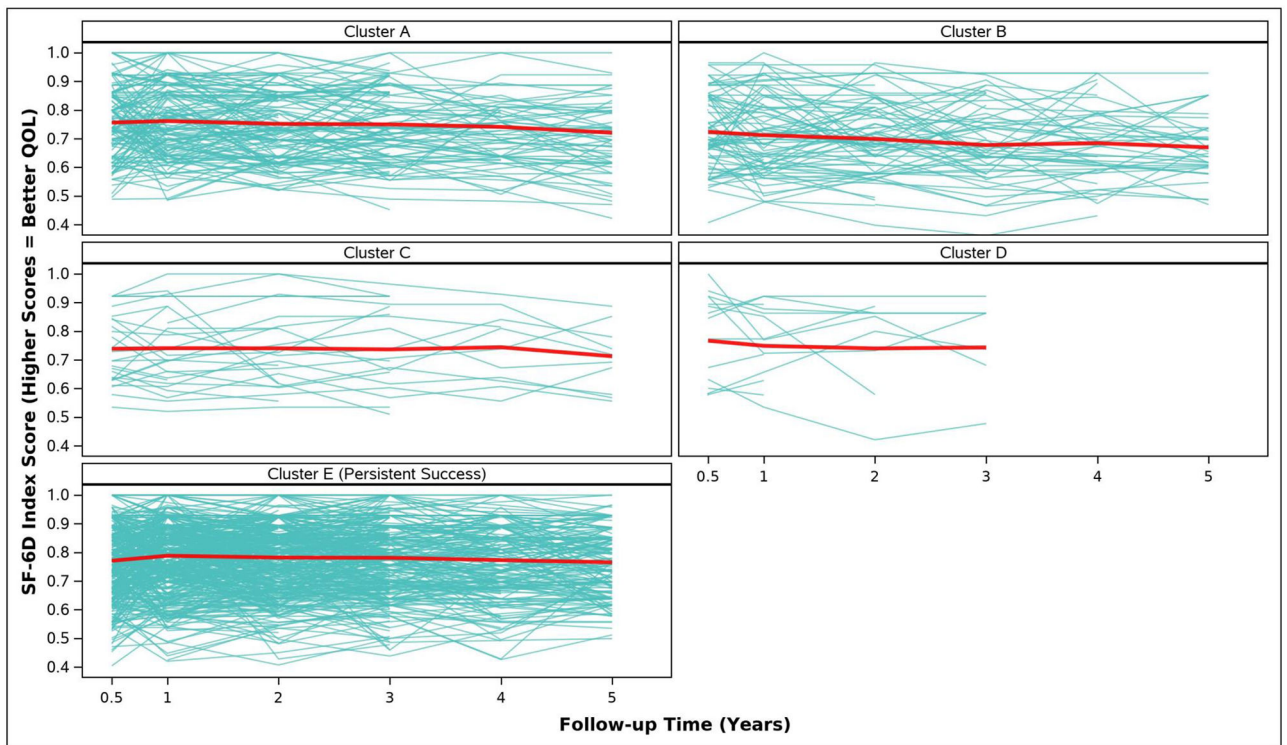

Figure 2H.

SF-6D Index Score by Cluster Assignment and Visit
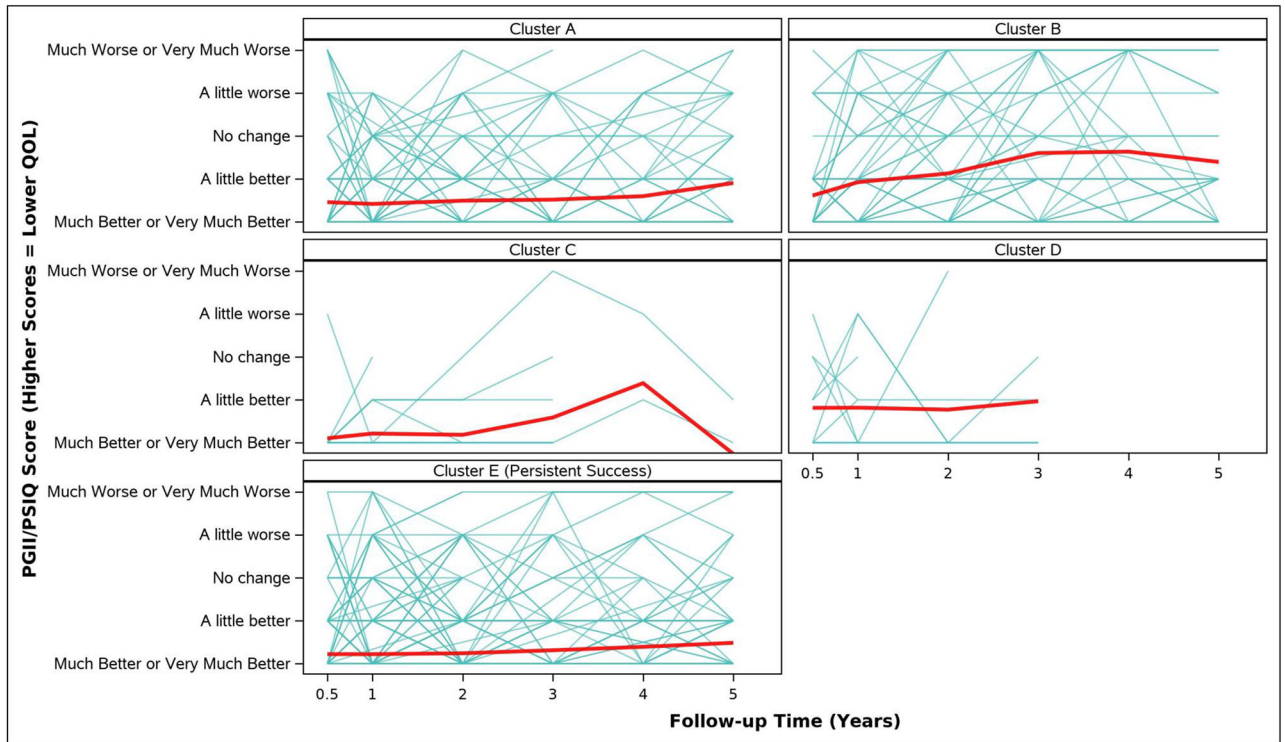
Figure 2I.
PGII/PSIQ Score by Cluster Assignment and Visit

References

1. Barber MD, Brubaker L, Nygaard I, et al. Defining success after surgery for pelvic organ prolapse. Obstetrics and gynecology 2009;114:600–09. [PubMed: 19701041]

2. Barber MD, Maher C. Epidemiology and outcome assessment of pelvic organ prolapse. Int Urogynecol J 2013;24:1783–90. [PubMed: 24142054]

3. Brubaker L, Cundiff GW, Fine P, et al. Abdominal sacrocolpopexy with Burch colposuspension to reduce urinary stress incontinence. N Engl J Med 2006;354:1557–66. [PubMed: 16611949]

4. Nygaard I, Brubaker L, Zyczynski HM, et al. Long-term outcomes following abdominal sacrocolpopexy for pelvic organ prolapse. JAMA : the journal of the American Medical Association 2013;309:2016–24. [PubMed: 23677313]

5. Wei JT, Nygaard I, Richter HE, et al. A midurethral sling to reduce incontinence after vaginal prolapse repair. N Engl J Med 2012;366:2358–67. [PubMed: 22716974]

6. Barber MD, Brubaker L, Menefee S, et al. Operations and pelvic muscle training in the management of apical support loss (OPTIMAL) trial: design and methods. Contemporary clinical trials 2009;30:178–89. [PubMed: 19130903]

7. Jelovsek JE, Barber MD, Brubaker L, et al. Effect of Uterosacral Ligament Suspension vs Sacrospinous Ligament Fixation With or Without Perioperative Behavioral Therapy for Pelvic Organ Vaginal Prolapse on Surgical Outcomes and Prolapse Symptoms at 5 Years in the OPTIMAL Randomized Clinical Trial. JAMA 2018;319:1554–65. [PubMed: 29677302]

8. Nager CW, Visco AG, Richter HE, et al. Effect of Vaginal Mesh Hysteropexy vs Vaginal Hysterectomy With Uterosacral Ligament Suspension on Treatment Failure in Women With Uterovaginal Prolapse: A Randomized Clinical Trial. JAMA 2019;322:1054–65. [PubMed: 31529008]

9. Barber MD, Brubaker L, Burgio KL, et al. Comparison of 2 transvaginal surgical approaches and perioperative behavioral therapy for apical vaginal prolapse: the OPTIMAL randomized trial. JAMA : the journal of the American Medical Association 2014;311:1023–34. [PubMed: 24618964]

10. Bump RC, Mattiasson A, Bo K, et al. The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. Am J Obstet Gynecol 1996;175:10–7. [PubMed: 8694033]

11. Barber MD, Kuchibhatla MN, Pieper CF, Bump RC. Psychometric evaluation of 2 comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. American Journal of Obstetrics & Gynecology 2001;185:1388–95. [PubMed: 11744914]

12. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). Am J Obstet Gynecol 2005;193:103–13. [PubMed: 16021067]

13. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42:851–9. [PubMed: 15319610]

14. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92. [PubMed: 11939242]

15. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. American Journal of Obstetrics & Gynecology 2003 7;189(1):98–101 2003. [PubMed: 12861145]

16. Genolini C, Lacombe A, Ecochard R, Subtil F. CopyMean: A new method to predict monotone missing values in longitudinal studies. Comput Methods Programs Biomed 2016;132:29–44. [PubMed: 27282225]

17. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning : with applications in R. New York: Springer; Number of pages.

18. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R Packages to Cluster Longitudinal Data. J Stat Softw 2015;65:1–34.

AJOG at a Glance:

Why was the study conducted?

Since clinical experience suggests some women after pelvic organ prolapse surgery have asymptomatic anatomic failures and perceive their surgery to be successful, while others have anatomic success but continue to report bothersome bulge symptoms, we clustered the longitudinal patterns of anatomy and bulge symptoms in women with prolapse surgery failures up to five years later to characterize these patterns.

What are the key findings?

Four mutually exclusive subgroups of pelvic organ prolapse outcome failure were identified, and each cluster had different and important associations with symptom bother, quality of life, and impression of overall improvement.

What does this study add to what is already known?

Identifying outcome failure types could be useful in refining definitions of success after pelvic organ prolapse surgery, in understanding mechanisms of failure, and in identifying individuals who may benefit from specific therapeutic approaches.
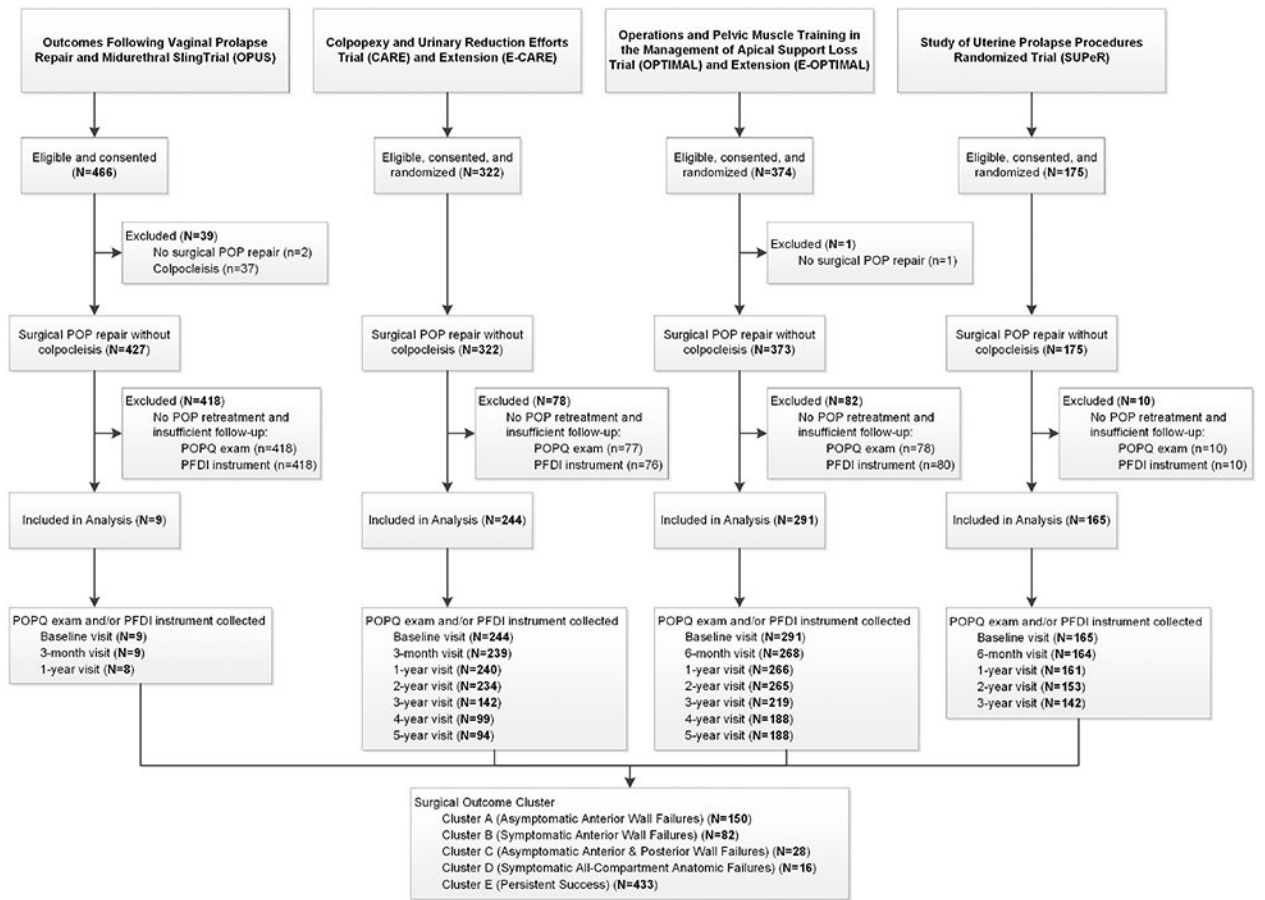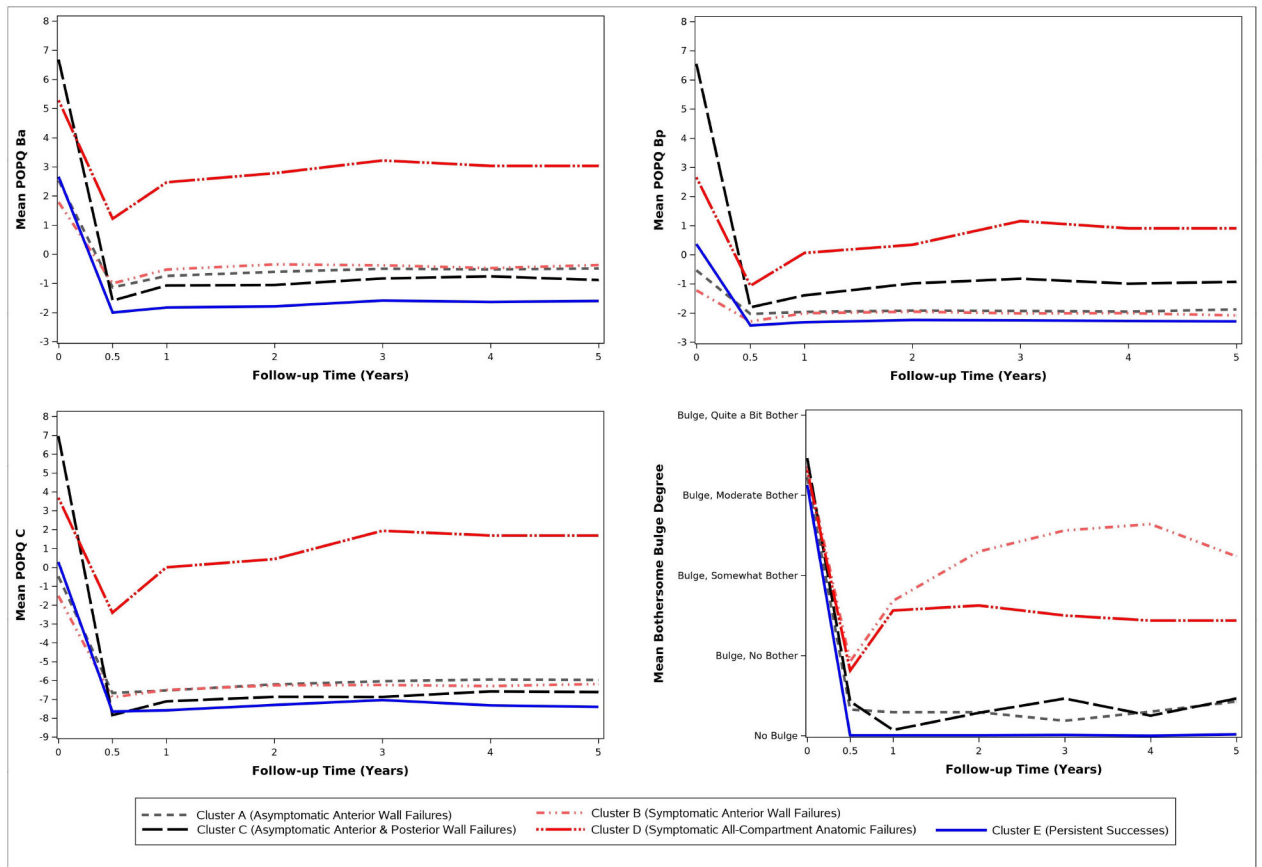
Figure 1.
Participant Flow

Figure 2.
Mean POPQ Ba, Bp, C Measurements and Bothersome Bulge Symptoms by Cluster
Assignment and Visit

Table 1.

Surgical Failure Types by Cluster Assignment [a]

Failure Types [b]	Cluster A - Asymptomatic Intermittent Anterior Wall Failures (N=150)	Cluster B - Symptomatic Intermittent Anterior Wall Failures (N=82)	Cluster C - Asymptomatic Intermittent Anterior and Posterior Wall Failures (N=28)	Cluster D - Symptomatic All-Compartment Anatomic Failures (N=16)	P-value [c]
Imputed Anterior Failure, n/N (%)	248/900 (27.6)	114/492 (23.2)	39/168 (23.2)	84/96 (87.5)	**<.0001**
Imputed Posterior Failure, n/N (%)	47/900 (5.2)	20/492 (4.1)	37/168 (22.0)	38/96 (39.6)	**<.0001**
Imputed Apical Failure, n/N (%)	11/900 (1.2)	0/492 (0.0)	0/168 (0.0)	49/96 (51.0)	**<.0001**
Imputed Anatomic Failure, n/N (%)	294/900 (32.7)	132/492 (26.8)	65/168 (38.7)	85/96 (88.5)	**<.0001**
Imputed Subjective Failure, n/N (%)	96/900 (10.7)	342/492 (69.5)	19/168 (11.3)	41/96 (42.7)	**<.0001**
Imputed Retreatment Failure, n/N (%)	135/900 (15.0)	52/492 (10.6)	14/168 (8.3)	50/96 (52.1)	**<.0001**
Imputed Surgical Failure, n/N (%)	426/900 (47.3)	373/492 (75.8)	78/168 (46.4)	86/96 (89.6)	**<.0001**
Anterior Failure, n/N (%)	140/627 (22.3)	77/338 (22.8)	16/114 (14.0)	34/44 (77.3)	**<.0001**
Posterior Failure, n/N (%)	28/627 (4.5)	16/338 (4.7)	19/114 (16.7)	12/44 (27.3)	**<.0001**
Apical Failure, n/N (%)	6/627 (1.0)	0/338 (0.0)	0/114 (0.0)	17/44 (38.6)	**<.0001**
Anatomic Failure, n/N (%)	167/627 (26.6)	91/338 (26.9)	34/114 (29.8)	35/44 (79.5)	**<.0001**
Subjective Failure, n/N (%)	78/632 (12.3)	218/350 (62.3)	16/117 (13.7)	18/45 (40.0)	**<.0001**
Retreatment Failure, n/N (%)	33/649 (5.1)	15/358 (4.2)	3/118 (2.5)	11/47 (23.4)	**<.0001**
Surgical Failure, n/N (%)	255/649 (39.3)	249/358 (69.6)	47/118 (39.8)	39/47 (83.0)	**<.0001**

[a] Cluster assignments A through D are the result of cluster analysis based on POPQ Ba, POPQ Bp, POPQ C, and bulge degree of bother reported on PFDI-46 items 4 or 5/PFDI-20 item 3 including all subjects with at least one surgical failure post-surgery, and prolapse retreatment or at least 4 visits with POPQ and bulge degree of bother collected. Among all subjects included in the cluster analysis, missing POPQ measurements and bulge degree of bother were imputed by the mean of the collected values.

[b] Anterior prolapse failure is defined as the prolapse of the anterior wall beyond the hymen (i.e. POPQ point Ba > 0). Posterior prolapse failure is defined as the prolapse of the posterior wall beyond the hymen (i.e. POPQ point Bp > 0). Apical prolapse failure is defined as the prolapse of the apical wall beyond the hymen (i.e. POPQ point C > 0). Anatomic failure is defined as a prolapse beyond the hymen (i.e. POPQ point Ba, Bp, or C > 0). Subjective failure is defined as the presence of a bothersome bulge (i.e. a positive response to PFDI-46 item 4 or 5, or PFDI-20 item 3 with a degree of bother more than "Not at all"). Retreatment failure is defined as re-operation or pessary for pelvic organ prolapse. Surgical failure is defined as the occurrence of an anatomic failure, subjective failure, or retreatment failure.

[c] P-values were obtained using Fisher's Exact Test for categorical measures. All tests were conducted at a significance level of 0.05.

Table 2.

Adjusted Analyses of Symptom Bother and Quality of Life Outcomes by Cluster Assignment [a, b]

Outcome Measures	Cluster A - Asymptomatic Intermittent Anterior Wall Failures (N=150)	Cluster B - Symptomatic Intermittent Anterior Wall Failures (N=82)	Cluster C - Asymptomatic Intermittent Anterior and Posterior Wall Failures (N=28)	Cluster D - Symptomatic All-Compartment Anatomic Failures (N=16)	Cluster E - Persistent Success (N=433)	P-value [b]
POPDI Score, Adjusted Mean (SE)						
3/6 Months [c]	36.9 (3.4)	63.3 (4.5)	33.5 (7.9)	63.6 (11.4)	30.6 (2.1)	**<.0001**
1 Year	34.5 (3.3)	75.6 (4.3)	37.7 (7.6)	62.1 (10.5)	27.9 (1.9)	**<.0001**
2 Years	44.8 (4.0)	98.4 (5.2)	39.4 (9.6)	88.0 (15.9)	30.0 (2.3)	**<.0001**
3 Years	42.9 (4.1)	111.8 (5.4)	36.9 (9.8)	65.5 (16.8)	29.8 (2.4)	**<.0001**
4 Years	53.3 (5.4)	117.2 (6.5)	56.1 (12.5)		29.0 (3.1)	**<.0001**
5 Years	56.3 (5.6)	105.6 (7.2)	50.4 (13.9)		29.6 (3.4)	**<.0001**
SF-6D Physical Function Score, Adjusted Mean (SE)						
3/6 Months [c]	2.1 (0.1)	2.5 (0.1)	2.6 (0.2)	1.8 (0.3)	2.2 (0.1)	**0.0260**
1 Year	2.0 (0.1)	2.7 (0.1)	2.5 (0.2)	2.1 (0.3)	1.9 (0.1)	**<.0001**
2 Years	2.1 (0.1)	2.6 (0.1)	2.5 (0.2)	1.8 (0.3)	1.9 (0.1)	**<.0001**
3 Years	2.2 (0.1)	2.7 (0.1)	2.7 (0.2)	2.0 (0.4)	2.0 (0.1)	**<.0001**
4 Years	2.2 (0.1)	2.7 (0.2)	2.0 (0.3)		2.1 (0.1)	**0.0177**
5 Years	2.5 (0.1)	2.6 (0.2)	2.3 (0.3)		2.1 (0.1)	**0.0056**
SF-6D Role Limitation Score, Adjusted Mean (SE)						
3/6 Months [c]	2.2 (0.1)	2.3 (0.1)	2.4 (0.2)	2.0 (0.3)	2.2 (0.1)	0.7727
1 Year	2.3 (0.1)	2.4 (0.1)	2.4 (0.2)	2.3 (0.3)	1.9 (0.1)	**<.0001**
2 Years	2.2 (0.1)	2.4 (0.1)	2.1 (0.2)	1.9 (0.4)	2.0 (0.1)	**0.0332**
3 Years	2.4 (0.1)	2.8 (0.2)	2.4 (0.3)	2.3 (0.4)	2.1 (0.1)	**0.0005**
4 Years	2.5 (0.2)	3.0 (0.2)	2.6 (0.4)		2.1 (0.1)	**<.0001**
5 Years	2.6 (0.1)	2.8 (0.2)	2.2 (0.4)		2.3 (0.1)	**0.0305**
SF-6D Social Function Score, Adjusted Mean (SE)						

Outcome Measures	Cluster A - Asymptomatic Intermittent Anterior Wall Failures (N=150)	Cluster B - Symptomatic Intermittent Anterior Wall Failures (N=82)	Cluster C - Asymptomatic Intermittent Anterior and Posterior Wall Failures (N=28)	Cluster D - Symptomatic All-Compartment Anatomic Failures (N=16)	Cluster E - Persistent Success (N=433)	P-value [b]
3/6 Months [c]	1.5 (0.1)	1.9 (0.1)	1.7 (0.2)	1.6 (0.2)	1.5 (0.0)	**0.0098**
1 Year	1.6 (0.1)	2.0 (0.1)	1.7 (0.2)	1.8 (0.2)	1.5 (0.0)	**0.0002**
2 Years	1.7 (0.1)	2.1 (0.1)	1.7 (0.2)	1.6 (0.3)	1.5 (0.0)	**<.0001**
3 Years	1.7 (0.1)	2.1 (0.1)	1.7 (0.2)	1.6 (0.4)	1.5 (0.1)	**0.0023**
4 Years	1.7 (0.1)	2.1 (0.1)	1.8 (0.3)		1.5 (0.1)	**0.0004**
5 Years	1.8 (0.1)	2.3 (0.2)	1.8 (0.3)		1.6 (0.1)	**0.0012**
SF-6D Pain Score, Adjusted Mean (SE)						
3/6 Months [c]	2.2 (0.1)	2.5 (0.1)	2.0 (0.2)	2.4 (0.3)	2.1 (0.1)	0.1185
1 Year	2.2 (0.1)	2.5 (0.1)	2.6 (0.2)	2.0 (0.3)	2.1 (0.1)	**0.0316**
2 Years	2.3 (0.1)	2.8 (0.1)	2.4 (0.3)	2.3 (0.4)	2.1 (0.1)	**0.0002**
3 Years	2.4 (0.1)	3.3 (0.2)	2.6 (0.3)	2.2 (0.4)	2.2 (0.1)	**<.0001**
4 Years	2.7 (0.1)	3.3 (0.2)	2.2 (0.3)		2.3 (0.1)	**<.0001**
5 Years	2.8 (0.1)	3.0 (0.2)	2.6 (0.3)		2.2 (0.1)	**<.0001**
SF-6D Mental Health Score, Adjusted Mean (SE)						
3/6 Months [c]	2.1 (0.1)	2.4 (0.1)	2.0 (0.2)	1.9 (0.2)	1.9 (0.0)	**0.0024**
1 Year	2.0 (0.1)	2.5 (0.1)	2.0 (0.2)	2.1 (0.3)	1.9 (0.0)	**0.0004**
2 Years	2.1 (0.1)	2.3 (0.1)	2.1 (0.2)	2.6 (0.3)	1.9 (0.1)	**0.0034**
3 Years	1.9 (0.1)	2.4 (0.1)	1.9 (0.2)	2.0 (0.3)	1.9 (0.1)	**0.0015**
4 Years	2.0 (0.1)	2.3 (0.1)	1.8 (0.3)		1.9 (0.1)	0.0714
5 Years	2.1 (0.1)	2.4 (0.1)	2.2 (0.3)		1.9 (0.1)	**0.0137**
SF-6D Vitality Score, Adjusted Mean (SE)						
3/6 Months [c]	2.7 (0.1)	2.8 (0.1)	2.6 (0.2)	2.6 (0.2)	2.5 (0.0)	0.1331
1 Year	2.5 (0.1)	2.7 (0.1)	2.7 (0.2)	2.9 (0.2)	2.5 (0.0)	0.2637
2 Years	2.7 (0.1)	2.8 (0.1)	2.6 (0.2)	2.3 (0.3)	2.6 (0.0)	0.0924
3 Years	2.7 (0.1)	3.0 (0.1)	2.9 (0.2)	2.2 (0.3)	2.5 (0.0)	**0.0018**

Outcome Measures	Cluster A - Asymptomatic Intermittent Anterior Wall Failures (N=150)	Cluster B - Symptomatic Intermittent Anterior Wall Failures (N=82)	Cluster C - Asymptomatic Intermittent Anterior and Posterior Wall Failures (N=28)	Cluster D - Symptomatic All-Compartment Anatomic Failures (N=16)	Cluster E - Persistent Success (N=433)	P-value [b]
4 Years	2.6 (0.1)	3.0 (0.1)	2.9 (0.2)		2.5 (0.1)	**0.0020**
5 Years	2.9 (0.1)	3.1 (0.1)	3.3 (0.3)		2.6 (0.1)	**0.0004**
SF-6D Index Score, Adjusted Mean (SE)						
3/6 Months [c]	0.8 (0.0)	0.7 (0.0)	0.7 (0.0)	0.8 (0.0)	0.8 (0.0)	**0.0321**
1 Year	0.8 (0.0)	0.7 (0.0)	0.7 (0.0)	0.7 (0.0)	0.8 (0.0)	**<.0001**
2 Years	0.8 (0.0)	0.7 (0.0)	0.7 (0.0)	0.7 (0.0)	0.8 (0.0)	**<.0001**
3 Years	0.8 (0.0)	0.7 (0.0)	0.7 (0.0)	0.7 (0.0)	0.8 (0.0)	**<.0001**
4 Years	0.7 (0.0)	0.7 (0.0)	0.7 (0.0)		0.8 (0.0)	**<.0001**
5 Years	0.7 (0.0)	0.7 (0.0)	0.7 (0.0)		0.8 (0.0)	**<.0001**
PGII/PSIQ Score [d], Adjusted Mean (SE)						
3/6 Months [c]	1.5 (0.1)	1.6 (0.1)	1.1 (0.2)	1.8 (0.2)	1.2 (0.0)	**<.0001**
1 Year	1.4 (0.1)	1.9 (0.1)	1.2 (0.2)	1.8 (0.2)	1.2 (0.0)	**<.0001**
2 Years	1.5 (0.1)	2.1 (0.1)	1.2 (0.2)	1.8 (0.3)	1.2 (0.0)	**<.0001**
3 Years	1.5 (0.1)	2.6 (0.1)	1.6 (0.3)	2.0 (0.3)	1.3 (0.1)	**<.0001**
4 Years	1.6 (0.1)	2.6 (0.2)	2.4 (0.6)		1.4 (0.1)	**<.0001**
5 Years	1.9 (0.1)	2.4 (0.2)	0.7 (0.7)		1.5 (0.1)	**<.0001**
1-Year QALY [e], Adjusted Mean (SE)	0.7 (0.0)	0.7 (0.0)	0.7 (0.0)	0.8 (0.0)	0.8 (0.0)	**0.0020**
2-Year QALY [e], Adjusted Mean (SE)	0.8 (0.0)	0.7 (0.0)	0.7 (0.0)	0.8 (0.0)	0.8 (0.0)	**<.0001**

POPDI = pelvic organ prolapse distress inventory (POPDI) subscale scores from the Pelvic Floor Distress Inventory (PFDI)-46 or PFDI-20 short form. Higher scores indicate worse symptom distress.

PGII = Patient Global Impression of Improvement

PSIQ = Patient Satisfaction Items Questionnaire

SF-6D = Short-Form Six-Dimension health index

[a] Cluster assignments A through D are the result of cluster analysis based on POPQ Ba, POPQ Bp, POPQ C, and bulge degree of bother reported on PFDI-46 items 4 or 5, or PFDI-20 item 3 including all subjects with at least one surgical failure post-surgery, and prolapse retreatment or at least 4 visits with POPQ and bulge degree of bother collected. Among all subjects included in the cluster analysis,

missing POPQ measurements and bulge degree of bother were imputed by the mean of the collected values. Cluster E assignment was given to all subjects with persistent successes having at least 4 visits with POPQ and bulge degree of bother collected.

[b] Adjusted means, standard errors, and p-values for repeated measures outcomes were obtained from general linear mixed models adjusting for cluster, post-surgical visit, interaction between cluster and visit, and modeling within-subject correlation using an unstructured covariance structure. Adjusted means, standard errors, and p-values for all other outcomes were obtained from general linear models adjusting for cluster. All tests were conducted at a significance level of 0.05.

[c] 3-month visit for subjects in OPUS and CARE studies, 6-month visit for subjects in OPTIMAL and SUPER studies

[d] Patient Global Impression Inventory (PGII) was collected at post-surgical visits in OPUS, OPTIMAL/E-OPTIMAL, and SUPeR. Patient Satisfaction Items Questionnaire (PSIQ) was collect at post-surgical visits in CARE/E-CARE. A combined quality of life measure (PGII/PSIQ) was constructed with 5 levels ranging from 1=Much Better or Very Much Better' to 5='Much Worse or Very Much Worse' by collapsing the tail end values on the PGII and including PSIQ item 3.

[e] Quality adjusted life years (QALYs) are calculated using an area under the curve approach following the trapezoidal rule from each subject's SF-6D Index score reported at baseline, 3/6 months, 12 months, and 24 months. Each subject must have at least an SF-6D Index score reported at baseline and 12 months in order for a 1-year QALY to be calculated. Each subject must have at least an SF-6D Index score reported at baseline and either at 12 or 24 months in order for a 2-year QALY to be calculated. If the SF-6D Index score is missing at 24 months, it was assumed to be the same as at 12 months.