

ARTICLE

Benchmarking with Patient Outcomes — Unmasked Surgeon Comparisons within a Learning Community for Continuous Improvement

Gregory Katz, PhD, PharmD, MBA, Cécile Rousseau, MSc, Béatrice Cochener-Lamard, PhD, MD, Jean-Pierre Thierry, MD, Jens Deerberg-Wittram, MD, Zofia Das-Gupta, PhD, Dominique Monnet, MD, PhD, Zirui Song, MD, PhD

Vol. 7 No. 5 | May 2026

DOI: 10.1056/CAT.25.0467

Patient-reported outcome measures (PROMs) are increasingly applied in clinical practice to improve quality of care and patient outcomes. To date, however, little is known about their use for peer comparisons of performance between clinicians, which the authors refer to as benchmarking. In this study of cataract surgeons, benchmarking was implemented through a stepwise approach: (1) Before benchmarking, each surgeon could not access their performance data or compare it with that of their peers; (2) once benchmarking began, all surgeons' identities and individual results became visible to their peers. This benchmarking design was intended to create a learning community free from judgment or stigmatization, drawing each surgeon into a collaborative dynamic of continuous improvement. The authors hypothesized that surgeons within such a benchmarking community would be encouraged to modify their surgical indications, improve their patients' average outcomes, and reduce the volume of procedures without meaningful patient benefit. Between 2021 and 2025, the authors analyzed a benchmarking community of cataract surgeons practicing at four institutions in France. The benchmarking model included voluntary participation, patient involvement, improvement cycles, and respect for patient–surgeon autonomy in clinical decision-making. Dashboards allowed surgeons to compare their PROMs, clinician-reported outcome measures (CROMs), and case-mix-adjusted results. Access

to the benchmarking dashboard was supplemented through noncoercive, nonpunitive quarterly meetings with peers to discuss individual and community results, enabling improvement cycles. The authors analyzed data from over 2600 patients with complete PROM and CROM data and found that five times more surgeons modified the surgical indications they used during benchmarking than before benchmarking — 10 of 24 (42%) versus 2 of 24 (8%). The authors also found that only surgeons who actively consulted the benchmarking dashboard modified their surgical indications. Among those who modified their surgical indications during benchmarking, the average postsurgery health gain score of their patients increased by 22% (from 28.7 before to 35.1 after modifying the indications), and the rate of patients without meaningful benefit decreased by 34% (from 19% to 13%). The differences were similar when compared with performance before benchmarking, which yielded a 21% increase in average health gain scores and a 35% decrease in the proportion of patients below the meaningful outcome threshold. These findings support the potential of unmasked peer benchmarking to improve patient outcomes, reduce unnecessary procedures, and advance value-based surgical care.

Introduction

Increasingly, health care providers are integrating patient-reported outcome measures (PROMs) into routine practice to help improve care pathways and patient outcomes, but little is known about their use for peer comparisons of performance, which we refer to as benchmarking.¹ Some studies about open benchmarking — where peers' identities and results are unmasked — suggest that transparency can motivate quality improvement.^{2,3} The effectiveness of this may depend on a noncoercive, nonpunitive organizational culture, where surgeons are not judged but inspired to improve patient outcomes.^{1,4,5} In such an environment, surgeons could ideally share data on patient outcomes without negative consequences and focus on collaborative improvement through Plan-Do-Study-Act cycles, enabling them to make rapid adjustments.^{6,7}

Based on these studies, we hypothesized that surgeons within an unmasked peer benchmarking community would be more likely to modify their surgical indications and to avoid inappropriate surgeries than those not in a benchmarking setting. To this end, we began implementing open benchmarking with PROMs in four ophthalmic units in France that perform cataract surgery, one of the most frequently performed surgeries worldwide.⁸ From 2021 to 2025, we collected data in a registry with a standardized data architecture that combined PROMs, clinician-reported outcome measurements (CROMs), and case-mix variables aligned with the recommendations of the International Consortium for Health Outcomes Measurement (ICHOM)⁹ and the European Society of Cataract and Refractive Surgeons.¹⁰

Methods

Study Design

We collected data from two university hospitals — the Brest University Hospital Center (*Centre Hospitalier Universitaire de Brest*) (Brest) and the Nantes University Hospital Center (*Centre Hospitalier Universitaire de Nantes*) (Nantes) — and two private clinics — Chénieux Ophthalmology (*Ophthalmologie Chénieux*) (Limoges) and Sourdille-Atlantique Ophthalmologic Institute (*Institut Ophthalmologique Sourdille-Atlantique*) (Saint-Herblain) — in France between May 31, 2021, and May 30, 2025. Benchmarking was implemented through a stepwise approach in which surgeons initially had no access to their own performance data (PROMs, CROMs, case-mix variables) or to their peers' data. This was intended to allow for the collection of baseline data. After the implementation of benchmarking, surgeons accessed their performance data alongside that of their peers. We implemented the [ICHOM standard set of patient-centered outcome measures for cataracts](#),⁹ which includes a core set of outcome measures covering visual acuity, refractive outcomes, and surgical complications (CROMs); patient-reported limitations in daily activities (PROMs); and key case-mix variables, enabling risk-adjusted comparisons of cataract care across providers. We also applied an international methodology for implementing PROMs and value-based health care.¹¹

The methodology was validated by the scientific council of the French National Cataract Registry, which comprises representatives from French professional ophthalmology societies and the French National Union of Accredited Patient Associations (Appendix, Supplementary Methods). This scientific council served as the study's governing body, working with the French Ministry of Health and the public payer throughout the study. The registry and benchmarking were deployed over four consecutive 12-month phases: Year 1 — May 31, 2021, through May 30, 2022; Year 2 — May 31, 2022, through May 30, 2023; Year 3 — May 31, 2023, through May 30, 2024; and Year 4 — May 31, 2024, through May 30, 2025.

Before benchmarking, surgeons could not access their performance data or compare it with that of their peers. Once benchmarking began, all surgeons' identities and individual results became visible to them and their peers. In Year 1, presurgery PROM and CROM data were collected from surgeons at three institutions (Nantes University Hospital Center, Sourdille-Atlantique Ophthalmologic Institute, and Chénieux Ophthalmology), but no feedback was provided and no benchmarking occurred. In Year 2, the surgeons at these three institutions were randomly assigned at a 1:1 ratio to either immediate benchmarking via an online dashboard or delayed benchmarking starting in Year 3. In addition, in Year 3 Chénieux Ophthalmology was replaced by Brest University Hospital Center, which started benchmarking in Year 4.

Peer comparison was encouraged in two ways: (1) Starting in Year 1, a financial incentive of €30 was introduced for each complete case a surgeon shared with the benchmarking community to compensate for the reputational risk; (2) starting in Year 3, public recognition by health authorities was instituted, with participating surgeons' names published on the public payer and regional health authority portals; however, their individual performance results remained accessible to only the benchmarking participants.

Collection of PROMs Data

We invited adult patients (≥ 18 years of age) to complete a digital version of a validated PROM questionnaire specific to cataract surgery (Catquest-9SF) that measures patients' visual limitations in daily activities. The Catquest-9SF instrument was specifically calibrated for digital utilization in a French-speaking population, with its psychometric properties validated through Rasch analysis and published in 2021.¹² Before the consultations, the patients were informed about completing the Catquest-9SF questionnaire by staff and posters in the waiting room. We excluded patients with intumescent lenses and closed angles, ocular pathologies precluding postoperative visual improvement, or who were under legal guardianship. Patients provided their consent and completed the questionnaire via text message, email, or phone with a third-party provider. Patients were invited to complete the questionnaire 65 days post surgery. If needed, a first reminder was sent after 3 days, and a second after 7 days, via text message, email, or phone call, using the contact information the patient provided. Patients were not provided with financial incentives for their participation.

Health Gain Score

The Catquest-9SF total score is calculated by summing the responses to all nine questionnaire items, each of which is scored from 0 (no visual difficulty) to 4 (cannot perform owing to vision).¹² "Not applicable" responses are excluded from the calculation. The total score ranges from 0 to 36, where 0 indicates no perceived visual disability and 36 represents the highest level of self-reported visual difficulty.

For each patient, the health gain score was defined as the difference between their CatQuest-9SF scores before and after surgery, each rescaled to a 0–100 range through linear transformation based on Rasch-derived item calibration, yielding a range from –100 to +100. A positive health gain score indicated visual improvement after surgery, whereas a negative result indicated visual deterioration.

Definition of the Threshold for a Meaningful Outcome in Health Gain Score

When the study began in 2021, no published minimum important difference (MID) for cataract surgery had been established. In the absence of such data, the scientific council defined a priori a threshold of 10 points as the minimum health gain required to define a meaningful patient benefit, supported by expert consensus.

In 2022, Grimfors et al. published an anchor-based MID for the Catquest-9SF.¹³ Converting their logit-scale estimates to our 0–100 metric yielded MID values of approximately 4.6 for patients with better preoperative visual function and 16.6 for those with greater preoperative impairment. The threshold of 10 points lies between these two estimates and is consistent with the known baseline dependency of the MID documented by Grimfors et al., providing a posteriori support to our a priori threshold.¹⁴

Case-Mix-Adjusted Benchmarking

Because the four participating institutions had different patient populations, we tested a series of benchmarking algorithms using different case-mix variables, including CROM data,

visual comorbidities, age, and sex. Based on these variables, the scientific council adopted a stratification algorithm that defined three patient groups (severe, intermediate, and nonsevere). The benchmarking platform enabled surgeons from various institutions to compare patient health gain scores in each patient group.

Surveys of Surgeons

Every quarter for the last 3 years of the study, we invited participating surgeons to complete surveys on changes in their communication with patients and colleagues, as well as changes in their surgical decisions, techniques, lens choices, patient relationships, and overall practice patterns (Appendix, Supplementary Methods). Survey data were collected using Forms (Microsoft, Redmond, Washington, United States) between June 2022 and June 2025. All surgeons reviewed and authenticated their survey results with an electronic signature.

Benchmarking Dashboards

The benchmarking platform enabled surgeons to compare their results with peers, teams, and the community using a series of indicators, including PROM and CROM scores and patient health gain scores. The platform displays surgeons' case-mix-adjusted results for the selected period (Fig. 1). The benchmarking dashboards were proposed, developed, and refined through a number of iterative cycles with surgeon delegates from each participating institution, who contributed to their scientific content and ergonomic design. Each final dashboard was approved by a formal vote of the scientific council.

The names and results of all surgeons are visible to the dashboard community. The dashboard also includes a diagram showing all patient health gain scores as a function of their presurgery Catquest-9SF score, with thresholds for different levels of visual limitation in daily activities (Fig. 2).

The visit metrics for the benchmarking dashboard were measured using Clarity (Microsoft, Redmond, Washington, United States) from December 2023 to June 2025.

Feedback Cycles

Every quarter, surgeons in the benchmarking program were invited to a video conference with peers from participating institutions. Each session brought together 7 to 12 surgeons, with an overall participation rate of 17 out of 24 surgeons (70.8%) over the 4-year study period. During each session, surgeons discussed their individual results against community benchmarks, which were stratified by patient profiles. The sessions also served as a forum for refining dashboard ergonomics and case-mix adjustments. Beyond data review, these peer exchanges fostered a collegial community dynamic — allowing surgeons to engage in open discussion of their performance without fear of judgment or stigmatization, reinforcing a shared commitment to continuous improvement.

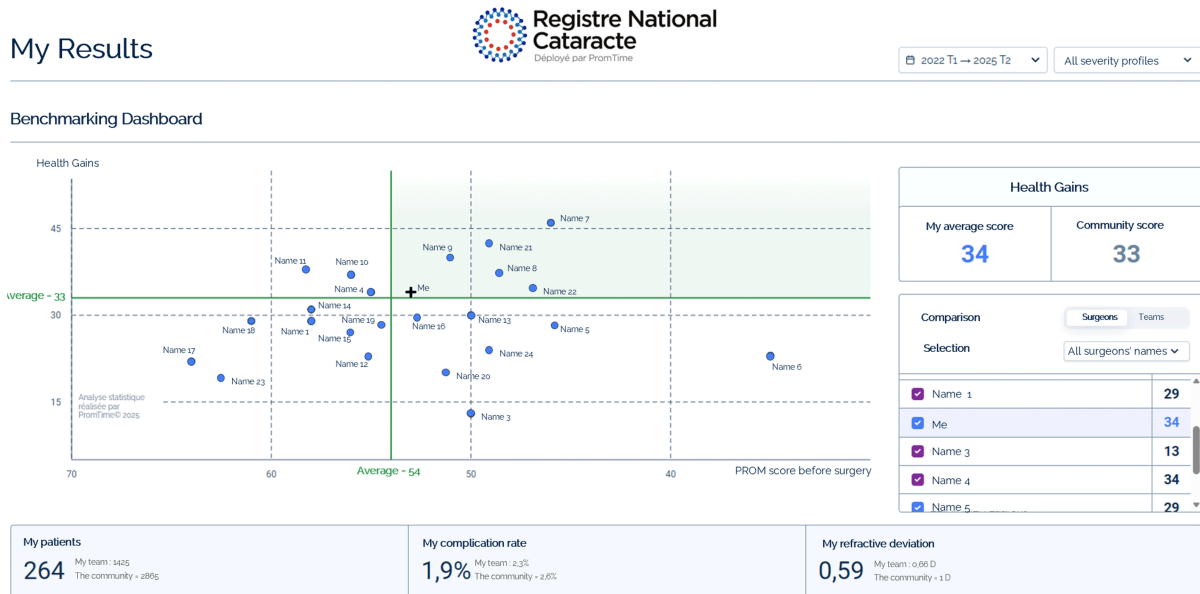
Statistical Analysis

The primary analysis compared surgeon-level outcomes before and after the introduction of unmasked peer benchmarking. We defined the analyzed population as surgeons who had completed at least one survey before and one survey during benchmarking and had complete

FIGURE 1

Benchmarking Dashboard: Main Screen.

The benchmarking dashboard displays each surgeon’s average health gain score (y-axis) for their patients relative to their patients’ average preoperative PROM score (x-axis). Each dot represents an individual surgeon, identified by name, positioned against community averages (the green lines). The logged-in surgeon (“Me”) is highlighted for direct peer comparison. Summary metrics — patient volume, complication rate, and refractive deviation — are shown against team and community benchmarks. The dashboard can be updated dynamically using two filters: a time period selector (from 2022 T1 onward) and a severity-based profile filter (severe, intermediate, and nonsevere), allowing case-mix-adjusted comparisons across surgeons and teams. PROM denotes patient-reported outcome measure.



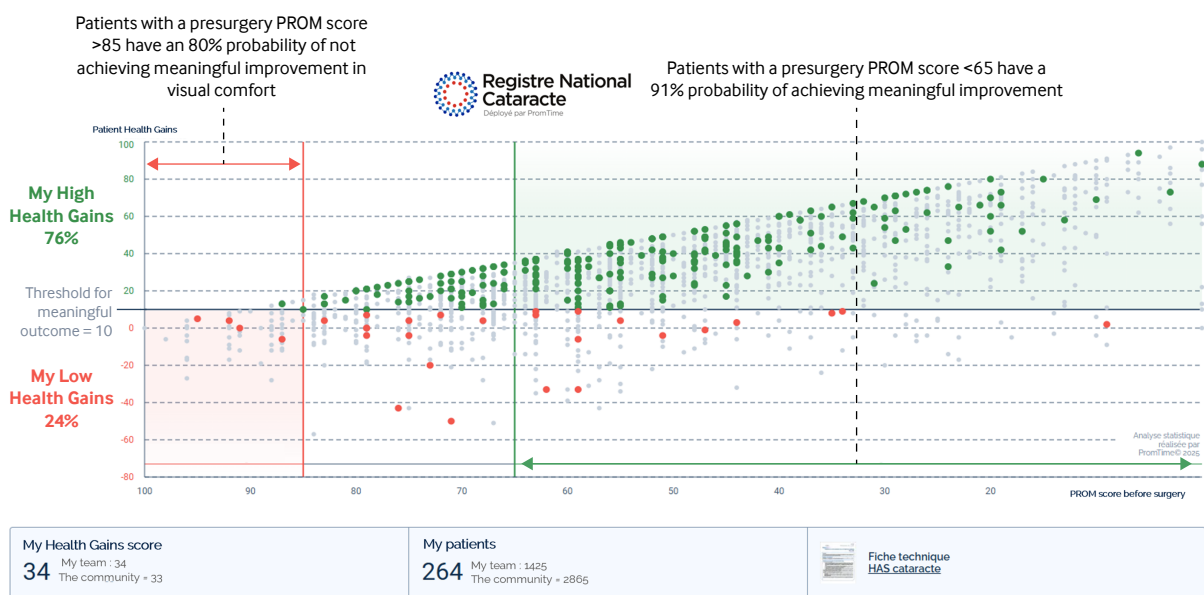
Source: PromTime dashboard for the French National Cataract Registry.
NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

PROM, CROM, and case-mix data for at least five patients. In addition, we analyzed the patient data of only the patients with complete pre- and postsurgery PROM and CROM data. We first compared the rate of surgeons who reported modifying their surgical indications before and during benchmarking (a pre- vs. postintervention design). For these surgeons, we also compared mean health gain scores before and after they modified their surgical indications. The association between benchmarking and the survey results was assessed using an exact McNemar test. Changes in health gain scores and modifications of surgical indications in relation to the benchmarking intervention were estimated using a linear regression model. The likelihood of a health gain score below the threshold for a meaningful outcome (<10) was calculated using a logistic regression model, with the model estimates converted to odds ratios. To improve the inference, the regression models used cluster-robust standard errors¹⁵ corrected by Bayesian bootstrapping with 10,000 draws. The probability of achieving a health gain above the threshold for a meaningful improvement (>10) was

FIGURE 2

Benchmarking Dashboard: Distribution of Patient Health Gain Score According to Presurgery PROM Score.

In this diagram, each dot represents an individual surgical patient (n=2865; some overlap), whose postoperative health gain (y-axis) is plotted against their preoperative PROM score (x-axis, inverted order). The logged-in surgeon's patients are highlighted: green dots indicate health gains above the threshold for a meaningful outcome (=10) and red dots indicate health gains below this threshold. Gray dots represent the entire patient community. The red and green vertical lines delineate preoperative profiles: Patients with a preoperative PROM score greater than 85 have an 80% probability of not achieving a meaningful improvement in visual comfort, while patients with a preoperative PROM score below 65 have a 91% probability of achieving meaningful improvement. The horizontal line at the threshold for a meaningful outcome (=10) marks the minimum health gain associated with a change in the patient's visual limitations in daily activities. PROM denotes patient-reported outcome measure.



Source: PromTime dashboard for the French National Cataract Registry.
NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

calculated for low and high levels of visual limitations in daily activities (presurgery Catquest-9SF scores >85 and <65, respectively). Statistical analyses were conducted using R version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria). A P value below 0.05 was considered statistically significant.

Regulatory and Ethical Framework

The study protocol and ethical framework were approved through three decrees published in the Official Journal of the French Republic (*Journal Officiel de la République Française*) (NOR codes SSAH2019853A, SPRH2328415A, and TSSZ2514457A). The study was overseen by a scientific council comprising representatives from French professional ophthalmology societies and the

French National Union of Accredited Patient Associations (Appendix, Supplementary Methods). Data were collected and managed in compliance with Article L. 1111-8 of the French Public Health Code and Regulation (EU) 2016/679.

Results

Sample

Of the 52 surgeons enrolled over the 4-year study period, 24 were included in the final analysis, as they met the criteria of having complete survey, PROM, and CROM data. These 24 surgeons collectively accounted for 2621 patients with complete PROM, CROM, and case-mix variable data (Table 1; Table S1). Among the 28 surgeons with incomplete data, the reasons for noncompletion were largely structural: Seven surgeons left their institution or ceased surgical activity before or during the benchmark phase; four had an insufficient caseload during the study period; eight had administrative omissions in their survey responses; and nine withdrew for reasons unrelated to the study objectives. No systematic differences in age, years of practice, or surgical volume at enrollment were identified between completers and noncompleters. While motivation-related self-selection cannot be fully excluded, the predominantly structural nature of the reasons for noncompletion plausibly limits the risk of motivation-related bias in the findings.

Regarding institutional participation, Nantes University Hospital Center and Sourdille-Atlantique Ophthalmologic Institute contributed data in all 4 years. Chénieux Ophthalmology stopped

Table 1. Breakdown of Study Participants.*

Institution	Surgeons	Year 1	Year 2	Year 3	Year 4	Overall
Brest University Hospital Center	Participated	Institution not yet participating		7	7	7
	With complete data			3	3	3
	With missing data			4	4	4
Chénieux Ophthalmology	Participated	16	16	Institution left the study		16
	With complete data	5	5			5
	With missing data	11	11			11
Nantes University Hospital Center	Participated	14	14	14	14	16†
	With complete data	7	7	7	7	7
	With missing data	7	7	7	7	9
Sourdille-Atlantique Ophthalmologic Institute	Participated	12	12	13	13	13
	With complete data	9	8	9	9	9
	With missing data	3	4	4	4	4
Total	Participated	42	42	34	34	52
	With complete data	21	20	19	19	24
	With missing data	21	22	15	15	28

*This table provides a breakdown of the study participants across the four organizations that provided cataract surgery to patients over the 4-year study. The study began with three institutions (Chénieux Ophthalmology, Nantes University Hospital Center, and Sourdille-Atlantique Ophthalmologic Institute). In Year 3, Chénieux Ophthalmology was replaced by Brest University Hospital Center, which started benchmarking in Year 4. The analyzed population was defined as surgeons who had completed at least one survey before and one survey during benchmarking and had complete PROM, CROM, and case-mix data for at least five patients. Of the 52 surgeons who participated in the study, 24 had these data and were included in the analyzed population. †Two surgeons without complete data left Nantes University Hospital Center (and, therefore, the study) during Year 2 and were replaced by two new surgeons. Therefore, a total of 16 surgeons from this institution participated in this study at some point.

participating in 2023 for institutional reasons and was replaced by Brest University Hospital Center in the same year, ensuring the continuity of the multicenter design.

Survey Completion Rates

Over the last 3 years of the study, surgeons were invited to complete semiannual surveys reporting modifications to their practice since the introduction of benchmarking. A total of 168 surveys were administered to the 24 surgeons included in the analysis, each of which was systematically reviewed and electronically signed by the responding surgeon to confirm the data's accuracy.

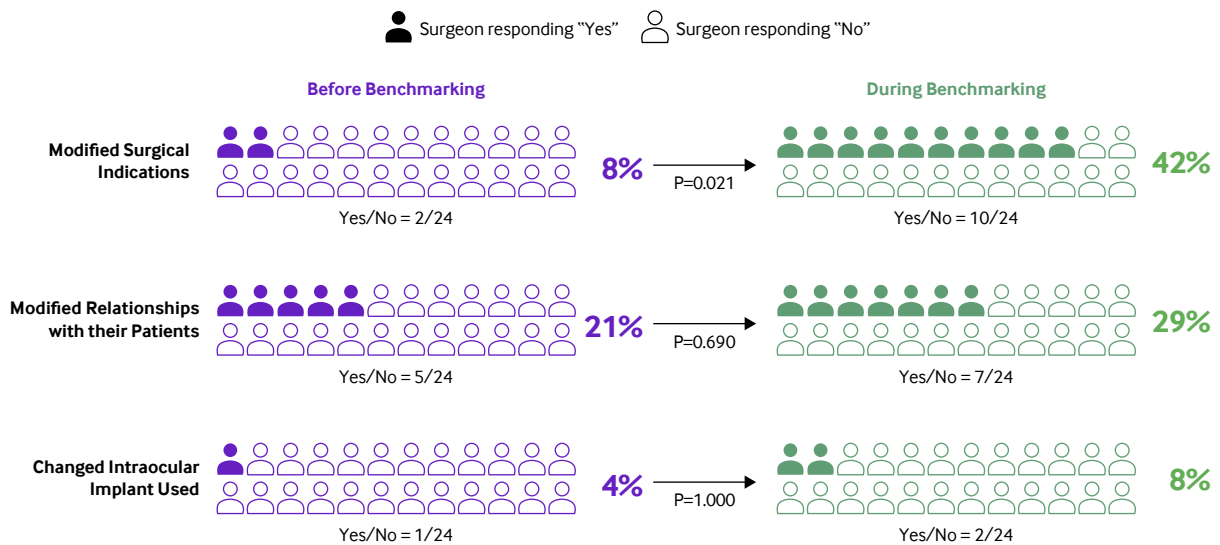
Survey Results

Five times more surgeons reported modifying their surgical indications during benchmarking than before (41.7% [10 out of 24] vs. 8.3% [2 out of 24]; $P=0.021$) (Fig. 3). In contrast, no statistically significant differences were observed in the proportion of surgeons reporting a modification in their patient–surgeon relationships (29.2% [7 out of 24] vs. 20.8% [5 out of 24]; $P=0.69$) or change in intraocular lens selection (8.3% [2 out of 24] vs. 4.2% [1 out of 24]; $P=1.0$).

FIGURE 3

Results from Surveys of Surgeons.

This figure shows the proportion of surgeons reporting modifications in their practice before (purple) and during (green) the benchmarking phase across three domains: the modification of surgical indications, the modification of patient–surgeon relationships, and a change in intraocular lens selection. The only statistically significant difference between the phases was observed for the modification of surgical indications. The reasons for the changes in the three categories were not determined by the survey. The survey questions and response choices are provided in the Appendix, Supplementary Methods section.

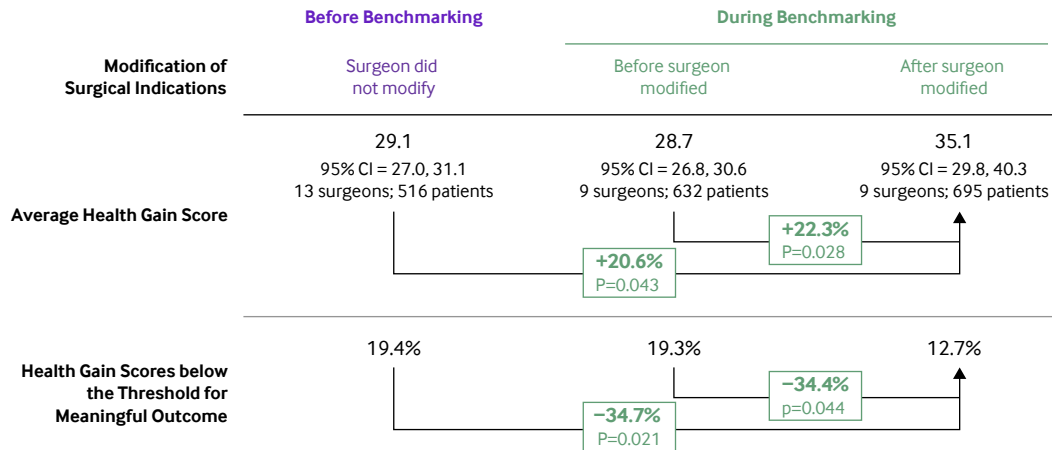


NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

FIGURE 4

Patient Health Gain Scores.

Average health gain scores and the proportion of health gains below the threshold for a meaningful outcome were compared before and after surgeons modified their surgical indications using linear regression models. Both metrics improved following modifications in surgical indications: Average health gain scores increased by 22.3% (P=0.028), and the proportion of health gain scores below the threshold for a meaningful outcome decreased by 34.4% (P=0.044). The differences were similar when compared with those before benchmarking, where there was a 20.6% increase in average health gain scores (P=0.043) and a 34.7% decrease in the proportion of health gain scores below the threshold for a meaningful outcome (P=0.021). CI denotes confidence interval.



NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

Patients' Health Gain Scores during Benchmarking

Patients' mean health gain scores increased by 22.3% (P=0.028), from 28.7 among 632 patients to 35.1 among 695 patients (Fig. 4 and Tables S2–S6). The results were similar when compared with those from before benchmarking — a 20.6% increase (P=0.043) from 29.1 among 516 patients to 35.1 among 695 patients.

Patient Health Gain Scores below the Threshold for Meaningful Outcome

The rate of patient health gain scores below the threshold for a meaningful outcome (<10) decreased by 34.4% (P=0.044) from 19.3% before to 12.7% (Fig. 4 and Tables S5 and S6). The results were similar when compared with those before benchmarking, where there was a 34.7% decrease in the proportion of patients below the meaningful outcome threshold (P=0.021).

“During benchmarking, surgeons modified their surgical indications five times more frequently than before benchmarking (41.7% vs. 8.3%; P=0.021).”

Benchmarking Dashboard and Changes in Surgical Indications

Of the 18 surgeons for whom dashboard data were available, 13 (72.2%) consulted the benchmarking dashboard at least once, of whom 9 (69.2%) subsequently modified their surgical indications. None of the five surgeons who did not consult the dashboard reported modifying their surgical indications.

Distribution of Patient Health Gain Scores According to Presurgery PROM Score

Among the 2621 patients included in the analysis, 92% achieved a health gain score above the threshold for a meaningful outcome (>10) when their preoperative Catquest-9SF score was below 65. By contrast, 80% of patients achieved a health gain score below the threshold for a meaningful outcome (<10) when their preoperative Catquest-9SF score was above 85.

Discussion

These results support the integration of unmasked peer benchmarking into routine surgical practice and encourage further evaluation of incentive models that prioritize engagement in outcome improvement as a central lever for advancing value-based health care.

Key Findings

Over 4 years, we implemented a surgeon-identified peer benchmarking model across four French institutions, enabling cataract surgeons to compare their PROMs, CROMs, and case-mix-adjusted health gain scores within a learning community. During benchmarking, surgeons modified their surgical indications five times more frequently than before benchmarking (41.7% vs. 8.3%; $P=0.021$). Patients operated on by surgeons who revised their indications showed 22.3% greater average health gains ($P=0.028$) and a 34.4% reduction in procedures without a meaningful patient benefit (12.7% vs. 19.3%; $P=0.044$). The differences were similar when comparing to before benchmarking, where there was a 20.6% increase in average health gain scores ($P=0.043$) and a 34.7% decrease in the proportion of patients below the meaningful outcome threshold ($P=0.021$). These findings are consistent with individual studies showing improvements in patient-reported outcomes when clinicians receive PROMs-based feedback, although clear benefits had not previously been demonstrated at the health system level.¹⁶

“*Patients operated on by surgeons who revised their indications showed 22.3% greater average health gains ($P=0.028$) and a 34.4% reduction in procedures without a meaningful patient benefit ($P=0.044$).*”

Contextualizing the Results

The proportion of patients with health gain scores below the threshold for a meaningful outcome among surgeons who had not been benchmarked (19.3%) is consistent with registry data from

Sweden¹⁷ and Wales,¹⁸ which also used the Catquest-9SF questionnaire and reported that 18% to 20% of cataract surgeries do not improve patients' visual limitations from the patient's perspective. In our study, this proportion fell to 12.7% among the patients of surgeons who modified their indications during benchmarking, supporting the hypothesis that PROMs-based benchmarking within a safe learning community may help avoid unnecessary procedures, reduce health care spending,¹⁹ and improve patient outcomes.

Toward PROM-Guided Surgical Indications

There is a subset of patients who are unlikely to achieve a meaningful health gain — not because of surgical failure, but because they enter the operating room with an insufficient symptom burden. Our retrospective data show that patients with presurgery PROM scores above 85 carry an 80% probability of having a nonmeaningful outcome; conversely, those with scores below 65 have a 91% probability of crossing the benefit threshold. These empirically derived thresholds potentially reflect a structural ceiling effect inherent to the indication rather than the procedure itself. We further observed that, during preoperative consultations, many surgeons actually asked patients three Catquest-9SF items to inform their decision-making: Patients reporting minimal visual disability were more likely to have their surgery postponed, while those with greater self-reported impairment were prioritized. These observations reflect a shift toward integrating presurgery PROM scores into decision-making about surgical indications to reduce the number of low-value procedures conducted.

Extending the Martini Klinik Peer Benchmarking Model

Our benchmarking model was inspired by the unmasked peer-comparison model pioneered at the Martini Klinik in Hamburg — a landmark for open-identity benchmarking.²⁰ By systematically comparing patient outcomes following prostate cancer treatment within a closed community of surgeons, the Martini Klinik demonstrated that voluntary exposure to comparative performance data drives continuous improvement in outcomes without coercive mandates. Our study sought to determine whether this core mechanism — the motivational power of unmasked peer benchmarking within a continuous-improvement community — is generalizable across public and private institutions, surgical cultures, and conditions. Our results suggest that this benchmarking model is likely transferable from prostate cancer in Germany to cataracts in France, and potentially generalizable across other health care systems.

“ *Unmasked peer benchmarking encourages a model of horizontal self-regulation, where professionals collectively assume responsibility for patient outcomes without top-down oversight.* ”

Mitigation of Naming and Shaming Risk

From the outset, the benchmarking community was designed to have a noncoercive, nonpunitive collegial culture — one in which data comparisons would encourage a collective dynamic rather than individual judgments. To mitigate the risk of naming and shaming, open communication and

shared learning were explicitly embedded in the community's culture. Participating surgeons were not direct competitors — they belonged to teams with long-standing collaborative relationships. Participation in structured quarterly dialogue sessions was voluntary, supporting iterative cycles of continuous improvement. We selected an unmasked peer benchmarking approach to encourage peer accountability, hypothesizing that it would provide stronger motivation for improvement than masked benchmarking.

Peer Accountability and Self-Regulation

Unmasked peer benchmarking encourages a model of horizontal self-regulation, where professionals collectively assume responsibility for patient outcomes without top-down oversight. This organizational model operates on two complementary levels. First, it realigns caregivers around patient recovery and quality of life rather than process or volume metrics. Second, it breaks down professional silos by creating structured opportunities for peer dialogue, transforming benchmarking from a solitary performance review into part of a cycle of collegial improvement. Together, these two dimensions suggest that unmasked peer benchmarking may be a sustainable model for professional self-governance, capable of aligning individual motivation and collective aspiration within a single value-based framework grounded in peer accountability.

“*When funding was discontinued at the end of the study, all participating surgeons continued to benchmark without compensation, suggesting that peer comparison and professional recognition may represent more compelling drivers of engagement than monetary rewards alone.*”

Limitations

Our study faced implementation barriers commonly observed when integrating PROMs and benchmarking into routine care.^{19,21,22} The entry of one institution and the exit of another during the study period may have introduced unmeasured confounding. Unobserved differences in patients, skills, or preferences across surgeons may have also introduced a source of bias. Participation was voluntary raising the potential for recruitment bias; a possible mitigation strategy would be to randomly assign each surgeon to a new data collection week each month, reducing burden while improving the sample's representativeness. Financial incentives may have influenced surgeons' willingness to participate; however, when funding was discontinued at the end of the study, all participating surgeons continued to benchmark without compensation, suggesting that peer comparison and professional recognition may represent more compelling drivers of engagement than monetary rewards alone.

Toward a Learning Community Payment Model

Among the payment models most commonly associated with value-based health care, bundled payments consolidate the reimbursement of all services delivered across a defined care episode into a single prospectively determined payment. By aligning financial incentives around the entire

care pathway, this model encourages coordination and reduces unnecessary health care utilization. However, bundled payments remain largely outcome-agnostic: Providers are rewarded for cost containment regardless of whether patient-reported outcomes improve. A complementary model — which we propose calling the “learning community payment” model — conditions reimbursement not on the outcomes achieved but on participation in a continuous improvement community in which surgeons’ performance data are shared transparently among their peers. In this model, surgeons who consent to unmasked peer benchmarking receive larger reimbursements, while those who opt out receive proportionally smaller payments. This model preserves professional autonomy: No surgeon is compelled to participate, and no financial penalty is tied to the level of performance attained. By decoupling financial incentives from outcomes and anchoring them instead to unmasked peer benchmarking, the learning community payment model creates a structurally meaningful incentive for a self-regulated professional environment engaged in advancing value-based health care.

Gregory Katz, PhD, PharmD, MBA

Professor and Chair of Value-Based Health Care, Université Paris Cité School of Medicine, Paris, France

President, PromTime, Paris, France

Cécile Rousseau, MSc

Chief Operating Officer, PromTime, Paris, France

Béatrice Cochener-Lamard, PhD, MD

President, French National Professional Council of Ophthalmology, Paris, France

Chair, Ophthalmology Department, Brest University Hospital Center, Brest, France

Dean, Faculty of Medicine and Medical Sciences, Université de Bretagne Occidentale, Brest, France

Jean-Pierre Thierry, MD

President, French College of Patient Outcomes, Andresy, France

Medical Advisor, French National Union of Accredited Patient Associations, France Assos Santé, Paris, France

Jens Deerberg-Wittram, MD

Professor for Healthcare Management, Technische Universität Berlin, Berlin, Germany

Senior Fellow, Charité Center for Patient-Centered Outcomes Research, Berlin, Germany

Honorary Professor, Technical University of Berlin, Berlin, Germany

Zofia Das-Gupta, PhD

Senior Director, International Consortium for Health Outcomes Measurement, London, United Kingdom

Dominique Monnet, MD, PhD

Professor, Department of Ophthalmology, Cochin Hospital, Assistance-Publique Hôpitaux de Paris, Université Paris Cité, Paris, France

President, French Society of Intraocular Lenses and Refractive Surgery (SAFIR), Paris, France

Zirui Song, MD, PhD

Associate Professor of Health Care Policy and Medicine, Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

Practicing Physician, Massachusetts General Hospital, Boston, MA, USA

Zirui Song can be contacted at song@hcp.med.harvard.edu.

Disclosures: Author disclosures are available at catalyst.nejm.org.

Appendix

Acknowledgments

We thank the surgeons who participated in this study: Dr. Loic Bazin, Dr. Julien Begey, Dr. Anas-Alexis Benyoussef, Dr. Marianne Bourhi, Dr. Julien Corre, Dr. Martin Coquelet, Dr. Jeanne Le Dinahet, Dr. Aude Meal, Dr. Pauline Rolland, and Dr. Yanis Saib from Brest University Hospital Center; Dr. Florian Bodenes, Dr. Manon Clément, Dr. Charlène Cornée, Dr. Jean-Baptiste Ducloyer, Dr. Hélène Masse, Dr. Driss Mazhar, Dr. Marie Le Grignou, Dr. Olivier Lebreton, Dr. Guillaume Lebreton, Dr. Quentin Lefebvre, Dr. Isabelle Orignac, Dr. Maxime Routier, Dr. Marion Servant, Dr. Bertrand Vabres, Dr. Robin Vasseur, and Prof. Michel Weber from Nantes University Hospital Center; and Dr. Cyrille Albouy, Dr. Gaëlle Boulanger, Dr. Bruno Cisneros, Dr. Dimitri Coeuru, Dr. Vincent Grué, Dr. Jean-Michel Halbardier, Dr. Laurent Leininger, Dr. François Lignereux, Dr. Guillaume Peigné, and Dr. Lionel Stork from Sourdille-Atlantique Ophthalmologic Institute. Assistance with medical writing was provided by Phillip Leventhal (PPD Clinical Research services, Thermo Fisher Scientific), and statistical analysis was provided by Marine Gosset and Nicolas Krucien (PPD Evidera Patient-Centered Research). The Paris Cité University Foundation paid the costs of medical writing assistance and statistical analysis.

Funding

This study was funded by public and private grants from Paris Cité University Medical School, the French Ministry of Health, the Assurance Maladie, PromTime, the French College of Patient Outcomes, Johnson & Johnson, Zeiss, Alcon, and Bausch & Lomb.

Gregory Katz received funding for the study from Université Paris Cité Medical School, the French Ministry of Health, the Assurance Maladie, PromTime, the French College of Patient Outcomes, Bausch & Lomb, Johnson & Johnson Vision, Zeiss, and Alcon.

Cécile Rousseau is an employee of PromTime, which conducted this study.

Jens Deerberg-Wittram received personal consulting fees from Statista Healthcare, Resilience Care, The Boston Consulting Group, and Bertelsmann Foundation unrelated to the work described in this manuscript. Jens Deerberg-Wittram is an *NEJM Catalyst Innovations in Care Delivery* editorial board member.

Zirui Song received the following, which are unrelated to this article: research grants from the Agency for Healthcare Research and Quality, The Commonwealth Fund, Arnold Ventures, and the National Institutes of Health; honoraria from Google Ventures; payment for expert testimony from Greylock McKinnon Associates; travel support from the National Academy of Medicine, the AHIP Institute, Weill Cornell Medical College, and the Council of Medical Specialty Societies. He was also the Chair of the 2024 Annual Meeting Program for the Society of General Internal Medicine.

References

1. Staiger RD, Schwandt H, Puhan MA, Clavien P-A. Improving surgical outcomes through benchmarking. *Br J Surg* 2019;106:59-64. DOI: [10.1002/bjs.10976](https://doi.org/10.1002/bjs.10976).
2. Willmington C, Belardi P, Murante AM, Vainieri M. The contribution of benchmarking to quality improvement in healthcare. A systematic literature review. *BMC Health Serv Res* 2022;22:139. DOI: [10.1186/s12913-022-07467-8](https://doi.org/10.1186/s12913-022-07467-8).
3. Matulewicz RS, Tosoian JJ, Stimson CJ, et al. Implementation of a surgeon-level comparative quality performance review to improve positive surgical margin rates during radical prostatectomy. *J Urol* 2017;197:1245-1250. DOI: [10.1016/j.juro.2016.11.102](https://doi.org/10.1016/j.juro.2016.11.102).
4. Ali MP, Visser EH, West RL, van Noord D, van der Woude CJ, van Deen WK. Reporting feedback on healthcare outcomes to improve quality in care: a scoping review. *Implementation Sci* 2025;20:14. DOI: [10.1186/s13012-025-01424-9](https://doi.org/10.1186/s13012-025-01424-9).
5. Silver CM, Yang AD, Shan Y, et al. Changes in surgical outcomes in a statewide quality improvement collaborative with introduction of simultaneous, comprehensive interventions. *J Am Coll Surg* 2023;237:128-138. DOI: [10.1097/XCS.0000000000000679](https://doi.org/10.1097/XCS.0000000000000679).
6. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;2012:CD000259. DOI: [10.1002/14651858.CD000259.pub3](https://doi.org/10.1002/14651858.CD000259.pub3).
7. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the Plan-Do-Study-Act method to improve quality in healthcare. *BMJ Qual Saf* 2014;23:290-298. DOI: [10.1136/bmjqs-2013-001862](https://doi.org/10.1136/bmjqs-2013-001862).
8. Rossi T, Romano MR, Iannetta D, et al. Cataract surgery practice patterns worldwide: a survey. *BMJ Open Ophthalmol* 2021;6:e000464. DOI: [10.1136/bmjophth-2020-000464](https://doi.org/10.1136/bmjophth-2020-000464).
9. Mahmud I, Kelley T, Stowell C, et al. A proposed minimum standard set of outcome measures for cataract surgery. *JAMA Ophthalmol* 2015;133:1247-1252. DOI: [10.1001/jamaophthalmol.2015.2810](https://doi.org/10.1001/jamaophthalmol.2015.2810).
10. Segers MHM, Behndig A, van den Biggelaar FJHM, et al. The European Registry of Quality Outcomes for Cataract and Refractive Surgery: benefits and lessons of a multinational registry. *NEJM Catal Innov Care Deliv* 2024;5(11) <https://catalyst.nejm.org/doi/full/10.1056/CAT.24.0029>. DOI: [10.1056/CAT.24.0029](https://doi.org/10.1056/CAT.24.0029).
11. EIT Health. Implementing value-based health care in Europe: handbook for pioneers (Director: Gregory Katz). 2020 (https://eithealth.eu/wp-content/uploads/2020/05/Implementing-Value-Based-Healthcare-In-Europe_web-4.pdf).

12. Katz G, Rouquette A, Lignereux F, Mourgues T, Weber M, Lundström M. Validity of the French version of Catquest-9SF and use of an electronic notepad for entering patient-reported outcome measures. *Eye Vis (Lond)* 2021;8:11. DOI: [10.1186/s40662-021-00233-7](https://doi.org/10.1186/s40662-021-00233-7).
13. Grimfors M, Lundström M, Kugelberg M. Self-assessed visual function outcome in cataract surgery: minimum important difference of the Catquest-9SF questionnaire. *Eye Vis (Lond)* 2022;9:46. DOI: [10.1186/s40662-022-00318-x](https://doi.org/10.1186/s40662-022-00318-x).
14. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA* 2014;312:1342-1343. DOI: [10.1001/jama.2014.13128](https://doi.org/10.1001/jama.2014.13128).
15. Huang FL, Li X. Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behav Res Methods* 2022;54:1181-1199. DOI: [10.3758/s13428-021-01627-0](https://doi.org/10.3758/s13428-021-01627-0).
16. Bonsel JM, Itiola AJ, Huberts AS, Bonsel GJ, Penton H. The use of patient-reported outcome measures to improve patient-related outcomes — a systematic review. *Health Qual Life Outcomes* 2024;22:101. DOI: [10.1186/s12955-024-02312-4](https://doi.org/10.1186/s12955-024-02312-4).
17. Mollazadegan K, Lundström M. A study of the correlation between patient-reported outcomes and clinical outcomes after cataract surgery in ophthalmic clinics. *Acta Ophthalmol.* 2015;93:293-298. DOI: [10.1111/aos.12490](https://doi.org/10.1111/aos.12490).
18. Blyth C, Poole J, Sira-Parfitt L, Letchford R, Davies D, Cahill A. Using outcome data and costs to demonstrate “value” in our cataract service. Reducing variation & using outcomes to support direct care and triage. Poster presented at International Consortium for Health Outcomes Measurement Conference 2019, Rotterdam, the Netherlands, May 2–3, 2019 (https://www.ichom.org/files/conference/posters/Using_outcome_data_and_costs_to_demonstrate_Value_in_our_Cataract_Service.pdf).
19. Tabib CH, Bahler CD, Hardacker TJ, Ball KM, Sundaram CP. Reducing operating room costs through real-time cost information feedback: a pilot study. *J Endourology* 2015;29:963-968. DOI: [10.1089/end.2014.0858](https://doi.org/10.1089/end.2014.0858).
20. Huland H, Graefen M, Deerberg-Wittram J, eds. *Das Martini-Prinzip: Spitzenmedizin durch Spezialisierung, Ergebnistransparenz und Patientenorientierung*. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft, 2018.
21. Kristensen PK, Johnsen SP. Patient-reported outcomes as hospital performance measures: the challenge of confounding and how to handle it. *Int J Qual Health Care* 2022;34:ii59-ii64. DOI: [10.1093/intqhc/mzaco03](https://doi.org/10.1093/intqhc/mzaco03).
22. Cella D, Nolla K, Peipert JD. The challenge of using patient reported outcome measures in clinical practice: how do we get there? *J Patient Rep Outcomes* 2024;8:35. DOI: [10.1186/s41687-024-00711-1](https://doi.org/10.1186/s41687-024-00711-1).