JAMA Network Open

Original Investigation | Surgery

Generalized Pairwise Comparisons to Support Shared Decision-Making in the CODA Trial

Samuel Salvaggio, PhD; Sarah E. Monsell, MSc; Patrick J. Heagerty, PhD; Mickael De Backer, PhD; Emilie Barré, MSc; Jean-Christophe Chiem, PhD; Everardo D. Saad, MD; Marc Buyse, ScD; David R. Flum, MD, MPH

Abstract

IMPORTANCE Shared decision-making (SDM) can be made difficult by the multifaceted nature of outcome assessment. A rigorous method for analyzing results from multiple outcomes is called generalized pairwise comparisons (GPC), which could assist in SDM.

OBJECTIVE To examine whether GPC can be useful in SDM by using individual-patient data from the Comparison of Outcomes of Antibiotic Drugs and Appendectomy (CODA) trial.

DESIGN, SETTING, AND PARTICIPANTS This comparative effectiveness study used data from participants in the multicenter US CODA trial (conducted between May 2016 and March 2020). All possible pairs of patients (one from each arm) were formed to analyze each of 7 outcomes of interest sequentially. Data were analyzed between February 2020 and early 2024.

EXPOSURES Three scenarios of priorities related to a different order of outcomes were considered. The first scenario came from a consensus exercise with patients that favored antibiotics, whereas the other 2 were arbitrarily chosen to illustrate the range of possible outcomes depending on prioritizations. Scenario 2 favored neither treatment, and scenario 3 favored appendectomy.

MAIN OUTCOMES AND MEASURES The primary outcome was the net treatment benefit (NTB), a formal measure of benefit-risk, which is the net probability that a randomly selected patient from the antibiotic-assigned arm would have a more favorable outcome than a randomly selected patient from the appendectomy-assigned arm.

RESULTS A total of 1552 patients were included in the CODA trial, with 776 (mean [SD] age, 38.3 [13.4] years; 286 [37%] female) in the antibiotic arm and 776 (mean [SD] age, 37.8 [13.7] years; 290 [37%] female) in the appendectomy arm. The NTB of antibiotic treatment was 12.8% (95% CI, 7.1% to 18.3%; *P* < .001) for the first scenario, 3.2% (95% CI −2.4% to 8.7%; *P* = .27) for the second, and −14.5% (95% CI. −20.2% to −8.8%; *P* < .001) for the third. These results respectively favored antibiotics, neither treatment, or appendectomy, thus illustrating that benefit-risk varies considerably according to individual priorities.

CONCLUSIONS AND RELEVANCE This comparative effectiveness study of antibiotics and appendectomy illustrates that the GPC method is a flexible yet mathematically rigorous quantitative analysis of benefit-risk balance. This method provides a more exhaustive and nuanced quantitative assessment of the differences between 2 treatment modalities in terms of prioritized outcomes. Furthermore, GPC could support SDM by considering individual prioritizations of the multiple outcomes.

JAMA Network Open. 2025;8(3):e252484. doi:10.1001/jamanetworkopen.2025.2484

Key Points

Question Can generalized pairwise comparisons be used to assist with shared decision-making between patients and clinicians?

Findings This comparative effectiveness study used patient-level data from a randomized clinical trial comparing the outcomes of antibiotics vs appendectomy. Using generalized pairwise comparison, the net treatment benefit significantly favored antibiotics, was neutral, or significantly favored appendectomy, depending on the patient's order of priority.

Meaning This study found that prioritized outcomes are a powerful tool to assess the benefit-risk of a new treatment compared with standard of care in a mathematically rigorous way, providing the outcomes are prioritized to reflect patient-specific choices.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Introduction

Once the purview of physicians alone, clinical decision-making has evolved into a shared process in which a patient's preferences, circumstances, and priorities are elicited to determine a preferred treatment or course of action.[1] Shared decision-making (SDM) improves patient satisfaction, can help avoid decisional regret, may improve adherence to treatment, and most importantly, honors the ethical principle of autonomy in several medical fields, including surgery.[1-4] As a best practice, the SDM process includes a standardized decision aid that provides the needed information about the treatment options and elicits patient preferences.[5,6] A challenge in creating such decision aids is that treatments often have different effects on multiple outcomes and presenting results related to multiple benefits and risks can lead to confusion and cognitive overload.[7] A variety of methods have been identified for comparing patient preferences between specified alternatives; such preferences may relate to an individual's ranking of outcomes as part of the SDM process, and this can be done in a rigorous manner with specific quantitative techniques.[8] In addition, multicriteria decision analysis (MCDA) is a set of powerful tools that allow formal assessment of multiple outcomes by individual and group-level stakeholders.[9,10] However, a common feature of existing quantitative benefit-risk analyses such as MCDA is that multiple outcomes are analyzed separately, and the summary measures of these marginal analyses are then aggregated.

Generalized pairwise comparisons (GPC) are an emerging class of statistical methods for the comparison of 2 samples of patients (eg, in randomized clinical trials) in terms of several outcomes, possibly prioritized.[11] The win ratio and the desirability of outcome ranking (DOOR) are 2 popular instances of such GPC analyses.[12,13] In this article, we use GPC exactly as in a win ratio analysis of prioritized outcomes, but using the net treatment benefit (NTB) instead of the win ratio as a measure of treatment effect, given some limitations in interpreting the win ratio that make it inappropriate for benefit-risk analyses.[14] Briefly, GPC assumes an order of preference among the multiple potential outcomes that may result from 2 competing interventions. Note that this order of preference may be patient dependent, as illustrated later in this article. By doing this, different outcomes can be jointly analyzed using pairs of patients that are formed by taking each patient from the experimental group and comparing them with each patient from the control group. The ordered outcomes are sequentially compared in each pair, from the outcome considered most important to that considered least important. When outcomes are related to benefits and harms, GPC provides a formal benefit-risk analysis of a given treatment.[15] In each pairwise comparison, a pair is classified as favorable to the experimental treatment, unfavorable to the experimental treatment, or neutral. The GPC method has been used in cardiology for several conditions, including amyloid heart disease,[16] heart failure,[17] and myocardial infarction,[18] with the win ratio as the measure of treatment effect.[13,19] Here we use the GPC method to assess the net benefit of competing interventions in terms of the NTB, an absolute measure of effect that has a more intuitive interpretation to patients and is adequate to combine treatment effects on multiple outcomes having different baseline risks (ie, the risks expected in the control group).[11,20-23] Unlike MCDA and similar quantitative methods of benefit-risk analyses, GPC analyses implicitly take the correlation between the outcomes into account, which is useful to distinguish situations in which patients who develop toxic effects are also those more likely to derive benefit from a given treatment, arguably a more desirable situation than if patients with toxic effects would not derive such benefit.[15]

While SDM has most commonly been used for elective procedures, there is an increasing interest in its application to other conditions, including those that present acutely, such as uncomplicated appendicitis.[24] Over the past 20 years, multiple randomized clinical trials (RCTs) have compared antibiotics alone vs appendectomy and found an antibiotic strategy to be safe, effective, and noninferior for overall health status,[25,26] even though doubts remain about the risk of recurrent appendicitis and hospital readmission within 1 year.[27] Since there are many outcomes that are associated with the treatment of appendicitis, many of which with different perceived impacts on different individuals,[28,29] it was hypothesized that the GPC method applied to these outcomes might help future patients identify a treatment strategy based on their personal prioritization of different outcomes. We applied the GPC method to the individual-patient data from the Comparison of Outcomes of Antibiotic Drugs and Appendectomy (CODA) trial, a US multicenter RCT comparing appendectomy vs antibiotics in adults.[26] If the GPC method was effective in distinguishing treatment strategies based on patient prioritization of outcomes, it might have value in complementing existing SDM-based decision aids for appendicitis.

Methods

Data Source

The current work is based on individual patient data from the CODA trial (NCT02800785), an RCT that investigated the noninferiority of a 10-day course of antibiotic therapy as an alternative to surgery for the treatment of uncomplicated appendicitis. The CODA trial met the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline.[26] In total, 1552 patients were stratified by the presence or absence of appendicolith and were randomly assigned in a 1:1 ratio to receive antibiotics (776 patients) or appendectomy (776 patients) at 25 US centers.[26] The primary end point was 30-day health status, which was assessed by the European Quality of Life–5 Dimensions (EQ-5D) questionnaire.[30] Higher scores on this instrument indicate better health status; a minimal clinically important difference of 0.05 points has been established for posttraumatic stress disorder and was used in the CODA trial.[31] Secondary end points of that trial included the total numbers of workdays missed by the patient and by the caregiver within 30 days of enrollment, the duration of hospital stay, symptom resolution (absence of the following: pain in lower right quadrant, tenderness in lower right quadrant when pressed, fever, shaking, and chills) at 2 weeks, any additional overnight hospitalization within 30 days, and any drainage procedure within 30 days (eTable 1 in Supplement 1).[26] No additional institutional review board approval or informed consent were sought given that we reanalyzed already-existing data.

Prioritized Outcomes

We formed all possible pairs of patients, one from each arm of the CODA trial, to analyze sequentially each outcome of interest, starting with the outcome considered highest priority to the one considered lowest priority under different scenarios (eTable 2 in Supplement 1). Each pairwise comparison was evaluated starting from the outcome of highest priority and could result in 3 possible classifications: favorable or unfavorable to antibiotics when a difference was observed (eg, favorable to antibiotics if the patient from the antibiotic arm of a pair had a higher EQ-5D score) or neutral (when it was not possible to determine which patient within the pair had a better outcome). The latter situation could happen when (1) both patients presented the same values for the outcome of interest, (2) there was incomplete information for either patient (due to missing data), or (3) the difference did not reach a predefined threshold of clinical similarity (eg, pairwise comparisons on the EQ-5D score that were within a 0.05 absolute difference were considered neutral). For pairs that were classified as neutral for a given outcome, the pairwise comparisons were carried over to the next prioritized outcome. This sequence was repeated for all pairs until they were either classified as favorable to the antibiotic arm, unfavorable to the antibiotic arm, or all outcomes had been considered (**Figure 1**).

We considered 3 different scenarios among the many that could be envisaged, each related to a different order of priorities of outcomes (eTable 2 in Supplement 1). Scenario 1, which favors antibiotics, stems from responses to a survey on preference ranking provided by 443 of 3066 patients who accessed a decision support website.[32] The preference ranking of these patients used ratings on a 3-point scale for 7 outcomes of appendicitis management they consider most relevant to their well-being: 1 indicated not important; 2, somewhat important; and 3, extremely important (eTable 3 in Supplement 1). Given our goal of illustrating how the GPC method might help future patients with different individual priorities from those arising from the survey, we chose scenarios 2 and 3 as plausible choices of clinically relevant orderings chosen to illustrate that the results could also be neutral (scenario 2) or favor the alternative treatment, ie, appendectomy (scenario 3).

Statistical Analysis

The method of GPC is an extension to multiple outcomes of the Mann-Whitney form of the nonparametric Wilcoxon test.[11] The results from all pairwise comparisons across all outcomes are aggregated in a single statistic, the NTB statistic, which is computed as the proportion of all favorable pairs minus the proportion of all unfavorable pairs. For a more complete description of this method, refer to the eAppendix in Supplement 1. This statistic estimates the NTB, which can be interpreted as the net probability that a randomly selected patient taken from the antibiotic-assigned arm would have a more favorable outcome than a patient taken randomly from the appendectomy-assigned arm, given a specific order of prioritized outcomes. As a net probability (ie, difference between 2 probabilities), NTB ranges from −1 to 1, with 0 indicating no difference between the 2 treatment groups. Univariate GPC analyses were performed to estimate NTB for each outcome individually, assessing their independent impact on the treatment effect. For multivariate analyses, multiplicity was tackled with sequential testing, starting with overall NTB and then proceeding sequentially until a nonsignificant *P* value ($P > .05$) was found. Statistical inference for the NTB was performed using the large-sample distribution of the GPC test statistic, which is a U statistic.[33] All analyses were run using the software package buysetest version 3.0 in R version 4.2.3 (R Project for Statistical Computing). The presence of an appendicolith, solid or calcified material inside the appendix seen on imaging, was used as a stratification factor in the CODA trial because of its association with complications.[26]

Results

A total of 1552 patients from the CODA trial were included in this analysis. There were 776 (mean [SD] age, 38.3 [13.4] years; 286 [37%] female) in the antibiotic arm and 776 (mean [SD] age, 37.8 [13.7] years; 290 [37%] female) in the appendectomy arm.[26]

Univariate Analyses

Results of the univariate analyses are depicted in the **Table**. GPC results were consistent with those from marginal statistical analyses presented in the original publication of the CODA trial.[26]

Multivariate Analyses

Figure 1 shows a schematic view of the multivariate GPC analysis for scenario 1. The overall NTB for the primary prioritization scheme, scenario 1, was positive and statistically significant in favor of

Figure 1. Schematic View of the Multivariate Generalized Pairwise Comparison Analysis of Scenario 1

Prioritized outcomes	Antibiotics better, %	Neutral, %	Surgery better, %	NTB, %
EQ-5D at 30 d	23.9	59.1	17.0	6.9
Symptom resolution	9.3	41.2	8.6	0.6
Any hospitalization	0.6	37.8	2.8	-2.2
Any drainage procedure	0.3	37.2	0.3	0
Days work missed by patient	5.9	28.1	3.2	2.7
Days work missed by caretaker	1.2	26.3	0.6	0.6
Length of hospital stay	15.3	0	11.1	4.1
NTB for all prioritized outcomes (*P* <.001)				**12.8**

Each pair is classified hierarchically through the list of outcomes. The pair is assessed on the first outcome, here the European Quality of Life–5 Dimensions (EQ-5D) score at 30 days. If the result favors or disfavors the treatment, the pair is classified accordingly. Here, 23.9% of all pairs favored antibiotics, while 17.0% of the pairs favored surgery. However, if the pair is neutral, the assessment is carried over to the subsequent outcome for further classification, here symptom resolution. Here 59.1% of all pairs were neutral for EQ-5D at 30 days and were therefore assessed by symptom resolution. The lines in the figure illustrate how neutral or tied matches are resolved at the next level of the hierarchy.

antibiotic treatment (12.8%; 95% CI, 7.1%-18.3%; $P < .001$). **Figure 2** shows the information proportion, which is the weight of each outcome in this analysis (sum of favorable plus unfavorable proportions for each outcome), while the individual contribution provides the contribution of each outcome to NTB (difference of favorable minus unfavorable proportions for each outcome). All outcomes contributed to increasing the NTB in favor of the antibiotic group, except any hospitalization within 30 days and any drainage procedure within 30 days, although the latter only classified a negligible proportion of all possible pairs (0.6%). More than half of the overall NTB (6.9% of 12.8%) was contributed by EQ-5D at 30 days.

Figure 3 shows that the overall NTB for scenario 2 was also positive but not significantly different from zero in favor of antibiotic treatment (3.2%; 95% CI, −2.4% to 8.7%; $P = .27$). Here, the 6.9% contribution to the NTB due to EQ-5D at 30 days was counterbalanced by negative contributions of any hospitalization within 30 days (−4.6%), length of hospital stay (−1.3%), and any drainage procedure within 30 days (−0.2%).

Finally, the overall NTB was negative and statistically significant in favor of appendectomy in scenario 3 (−14.6%, 95% CI, −20.2% to −8.8%; $P < .001$) (**Figure 4**). The major contributors to this negative NTB were length of hospital stay (−8.9%) and any hospitalization within 30 days (−8.3%). Of note, there were no more pairs for the last 3 outcomes in the hierarchy, as the first 4 outcomes classified all pairwise comparisons as favorable or unfavorable.

Discussion

The CODA trial investigated antibiotic therapy as an alternative to surgery for the treatment of appendicitis.[26] The trial contributes to the body of knowledge suggesting that antibiotics are noninferior to appendectomy,[25,26] despite remaining doubts about specific outcomes.[27] However, this type of conclusion is based on group-level data and typically ignores personal prioritization of outcomes. By analyzing data from the CODA trial using GPC, we have illustrated how a formal quantitative analysis of the benefit-risk relationship of antibiotics compared with surgery can provide a more exhaustive and nuanced picture of the differences between the 2 treatment modalities. This approach also takes individual priorities into account. The NTB, an absolute measure of the treatment effect, estimates the net probability that a patient taken randomly from the antibiotic group would have a better outcome than a patient taken randomly from the appendectomy group, given a certain order of priorities of outcomes.[11] It is worth emphasizing that the NTB is a trial-level treatment effect (although potentially based on the personal preferences of a specific patient). As such, it cannot be interpreted as the probability for a specific patient to do better after receiving the treatment condition than the control condition, which would be a causal individual-level treatment effect.[34] GPC allows for an exhaustive benefit-risk analysis by accounting for dependencies between the outcomes (ie, using conditional probabilities), rather than analyzing outcomes at the group level (ie, using marginal probabilities).[15] Of note, thresholds of clinical similarity can be used to define a favorable pairwise comparison for a given outcome.

Table. Univariate NTB for Each Outcome Considered

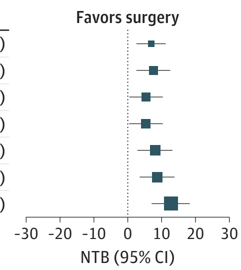
Outcome	Proportion of total pairs			NTB (95% CI)	*P* value
	Favor antibiotics	Favor surgery	Neutral		
EQ-5D at 30 d	0.239	0.170	0.591	0.069 (0.026 to 0.112)	.002
Symptom resolution at 2 weeks	0.179	0.167	0.654	0.012 (−0.029 to 0.053)	.58
Any overnight hospitalization within 30 d	0.026	0.109	0.865	−0.083 (−0.107 to −0.059)	<.001
Any drainage procedure within 30 d	0.008	0.021	0.972	−0.013 (−0.024 to −0.002)	.02
Workdays missed by patient within 30 d	0.218	0.123	0.660	0.095 (0.066 to 0.125)	<.001
Workdays missed by caretaker within 30 d	0.158	0.101	0.742	0.057 (0.029 to 0.086)	<.001
Length of hospital stay	0.584	0.416	0.000	0.168 (0.106 to 0.228)	<.001

Abbreviations: EQ-5D, European Quality of Life–5 Dimensions; NTB, net treatment benefit.

It is expected that patients are differentially impacted by the outcomes included in the analysis and the estimated NTB, based on their selected prioritization, reflects the overall benefit-risk that they would face by choosing one treatment over the other. The usefulness of individual prioritization rests on the assumption that individuals have a good understanding of the outcomes being considered. Thus, before conducting a ranking exercise with as part of SDM, patient training and discussions with a physician should take place, especially considering the large number of possibilities regarding prioritization. In the presence of 7 different outcomes, a total of 5040 permutations of outcomes are possible, each reflecting one order of priorities. For the present GPC

Figure 2. Multivariate Generalized Pairwise Comparison Analysis of Scenario 1

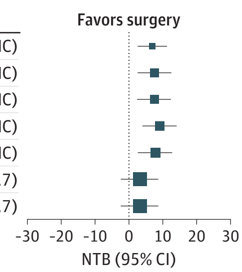
Prioritized outcomes	Total pairs, %	Neutral pairs, %	Antibiotics better, %	Surgery better, %	IP (CPC)	IC (CPC)	NTB (95% CI)	*P* value
EQ-5D at 30 d	100.0	59.1	23.9	17.0	40.9 (40.9)	6.9 (6.9)	6.9 (2.6-11.2)	.002
Symptom resolution at 2 wk	59.1	41.2	9.3	8.6	17.9 (58.8)	0.7 (7.6)	7.6 (2.6-12.5)	.003
Any hospitalization within 30 d	41.2	37.8	0.6	2.8	3.4 (62.2)	-2.2 (5.4)	5.4 (0.3-10.4)	.04
Any drainage procedure within 30 d	37.8	37.2	0.3	0.3	0.6 (62.8)	0 (5.4)	5.4 (0.3-10.4)	.04
Days of work missed by patient within 30 d	37.2	28.1	5.9	3.2	9.1 (71.9)	2.7 (8.1)	8.1 (2.9-13.2)	.002
Days of work missed by caretaker within 30 d	28.1	26.3	1.2	0.6	1.8 (73.7)	0.6 (8.7)	8.7 (3.5-13.8)	.001
Length of hospital stay	26.3	0.0	15.2	11.1	26.3 (100.0)	4.1 (12.8)	12.8 (7.1-18.3)	<.001



Information proportion (IP) was calculated as the percentage of pairs favoring antibiotics plus those favoring surgery; information contribution (IC), percentage of pairs favoring antibiotics minus those favoring surgery; net treatment benefit (NTB), sum of ICs. The size of the squares is proportional to the cumulative pairs classified (CPC). EQ-5D indicates European Quality of Life–5 Dimensions.

Figure 3. Multivariate Generalized Pairwise Comparison Analysis of Scenario 2

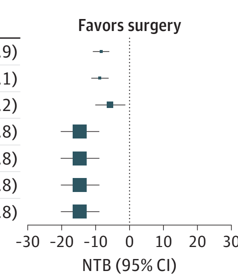
Prioritized outcomes	Total pairs, %	Neutral pairs, %	Antibiotics better, %	Surgery better, %	IP (CPC)	IC (CPC)	NTB (95% CI)	*P* value
EQ-5D at 30 d	100.0	59.1	23.9	17.0	40.9 (40.9)	6.9 (6.9)	6.9 (NC to NC)	NC
Symptom resolution at 2 wk	59.1	41.3	9.3	8.5	17.8 (58.7)	0.8 (7.7)	7.7 (NC to NC)	NC
Any drainage procedure within 30 d	41.3	40.5	0.3	0.5	0.8 (59.5)	-0.2 (7.5)	7.5 (NC to NC)	NC
Days of work missed by caretaker within 30 d	40.5	33.1	4.5	2.9	7.4 (66.9)	1.6 (9.1)	9.1 (NC to NC)	NC
Length of hospital stay	33.1	31.0	0.4	1.7	2.1 (69.0)	-1.3 (7.8)	3.2 (NC to NC)	NC
Any hospitalization within 30 d	31.0	0	13.2	17.8	31.0 (100.0)	-4.6 (3.2)	3.2 (-2.4 to 8.7)	.27
Days of work missed by patient within 30 d	0	0	0	0	0 (100.0)	0 (3.2)	3.2 (-2.4 to 8.7)	.27



Information proportion (IP) was calculated as the percentage of pairs favoring antibiotics plus those favoring surgery; information contribution (IC), percentage of pairs favoring antibiotics minus those favoring surgery; net treatment benefit (NTB), sum of ICs. The size of the squares is proportional to the cumulative pairs classified (CPC). The size of the squares is proportional to the cumulative number of pairs classified. NC indicates not calculated. EQ-5D indicates European Quality of Life–5 Dimensions.

Figure 4. Multivariate Generalized Pairwise Comparison Analysis of Scenario 3

Prioritized outcomes	Total pairs, %	Neutral pairs, %	Antibiotics better, %	Surgery better, %	IP (CPC)	IC (CPC)	NTB (95% CI)	*P* value
Any hospitalization within 30 d	100.0	86.5	2.6	10.9	13.5 (13.5)	-8.3 (-8.3)	-8.3 (-10.7 to -5.9)	<.001
Any drainage procedure within 30 d	86.5	85.3	0.4	0.8	1.2 (14.7)	-0.4 (-8.7)	-8.7 (-11.2 to -6.1)	<.001
Symptom resolution at 2 wk	85.3	56.9	15.7	12.7	28.4 (43.1)	3.0 (-5.7)	-5.7 (-10.0 to -1.2)	.01
Length of hospital stay	56.9	0	24.0	32.9	56.9 (100.0)	-8.9 (-14.6)	-14.6 (-20.2 to -8.8)	<.001
EQ-5D at 30 d	0	0	0	0	0 (100.0)	0 (-14.6)	-14.6 (-20.2 to -8.8)	<.001
Days of work missed by patient within 30 d	0	0	0	0	0 (100.0)	0 (-14.6)	-14.6 (-20.2 to -8.8)	<.001
Days of work missed by caretaker within 30 d	0	0	0	0	0 (100.0)	0 (-14.6)	-14.6 (-20.2 to -8.8)	<.001



Information proportion (IP) was calculated as the percentage of pairs favoring antibiotics plus those favoring surgery; information contribution (IC), percentage of pairs favoring antibiotics minus those favoring surgery; net treatment benefit (NTB), sum of ICs. The size of the squares is proportional to the cumulative pairs classified (CPC). The size of the squares is proportional to the cumulative number of pairs classified. EQ-5D indicates European Quality of Life–5 Dimensions.

analysis, we first considered the prioritized outcomes that were endorsed as most relevant in a survey of patients (scenario 1). The corresponding NTB for patients receiving antibiotic therapy, compared with patients undergoing surgery, was estimated at 12.8%, the difference between the probability of a better outcome for antibiotics (56.4%) and the probability of a better outcome for surgery (43.6%) using this order of priorities. As an additional interpretation, the inverse of the NTB can be translated to the number needed to treat. Here, approximately 8 patients (1 of 12.8% = 7.8 patients; 95% CI, 5.5-14.3 patients) would need to be treated on average for 1 patient to benefit from the antibiotic treatment.

Two other possible scenarios of outcomes ordering were used as illustrative sensitivity analyses. In scenario 2, raising drainage procedures and the length of hospital stay in the outcome hierarchy reduced the NTB to 3.2% and made it nonstatistically significant. Hence, for patients with this individualized prioritization of the outcomes, there is no overall statistical advantage of choosing one strategy over the other. Finally, scenario 3 placed hospitalizations within 30 days and drainage procedures as the outcomes of highest priority. With this prioritization, the NTB favored appendectomy, with a statistically significant negative NTB of −14.5% (corresponding to a number needed to treat of approximately 7 patients). Although these are only 3 of many possible scenarios, they reflect a range of variation that clearly shows that marginal comparisons between 2 interventions answer only part of a much more nuanced question about treatment benefit, one that can be explored in a rigorous and informative way using GPC.

Limitations

This study has limitations, including that this approach may not be consistent with the more nuanced way in which people actually make decisions, often in a gestalt-oriented manner that extends beyond rational definitions of priorities.[35] In addition, a prerequisite for the usefulness of GPC as a support to SDM would be willingness to undergo antibiotics or appendectomy. Such willingness is essential for SDM (and potential support from GPC) to be relevant.[29,36] A second limitation is that we could not include appendectomy as an outcome, since almost all patients in the appendectomy arm received surgery by design; nevertheless, salvage appendectomy is an important outcome for a patient choosing antibiotics, and our analysis could not take that outcome into account. Third, some of our results are heavily influenced by the EQ-5D; in scenario 1, approximately half of the net benefit (6.9% of 12.8%) from antibiotics was attributable to an EQ-5D higher by at least 0.05 at 30 days. The use of EQ-5D as a patient-reported outcome warrants caution, as its relevance may vary by clinical context. This also highlights a fourth limitation, of a more general nature, which is the fact that with GPC, the analysis of the first outcome is identical to a univariate analysis of this outcome, while the contributions of other outcomes are conditional on all previous outcomes being neutral. Similarly, the hierarchical nature of GPC may lead outcomes with lower probabilities of equivalence to disproportionately influence the analysis, potentially affecting interpretability of the overall NTB. Finally, it is worth noting that the application of GPC to indirect comparisons, like matching adjusted indirect comparisons for cross-trial analyses, generally require population adjustments which were not required in our reanalysis of the CODA trial.[37]

Conclusions

In this benefit-risk analysis of antibiotics vs appendectomy, the NTB was used as a tool to collaboratively engage patients in SDM using statistical summaries tailored to their individual preferences. There is considerable interest in the development of support tools for SDM in uncomplicated appendicitis.[7,24,38] We surmise that the GPC methodology could contribute to decision support tools by incorporating a multivariate dimension to such tools. To fully explore the potential of adding GPC to decision support tools for SDM, work remains to be done toward communicating probabilities to patients, a key undertaking in medicine in general, and in the setting of appendicitis in particular.[39]

ARTICLE INFORMATION

Accepted for Publication: January 28, 2025.

Published: March 31, 2025. doi:10.1001/jamanetworkopen.2025.2484

Corresponding Author: Samuel Salvaggio, PhD, One2Treat, 25 Bd Baudouin 1er, 1348 Louvain-la-Neuve, Belgium (samuel.salvaggio@one2treat.com).

Author Affiliations: One2Treat, Louvain-la-Neuve, Belgium (Salvaggio, De Backer, Barré, Chiem, Buyse); University of Washington, Seattle (Monsell, Heagerty, Flum); IDDI (International Drug Development Institute), Louvain-la-Neuve, Belgium (Saad, Buyse); I-BioStat, Hasselt University, Hasselt, Belgium (Buyse).

Author Contributions: Dr Flum had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Salvaggio, Monsell, Heagerty, Chiem, Buyse, Flum.

Acquisition, analysis, or interpretation of data: Salvaggio, Monsell, Heagerty, De Backer, Barre, Saad, Flum.

Drafting of the manuscript: Salvaggio, Monsell, Saad, Buyse.

Critical review of the manuscript for important intellectual content: Salvaggio, Monsell, Heagerty, De Backer, Barre, Chiem, Flum.

Statistical analysis: Salvaggio, Monsell, Heagerty, De Backer, Barre, Chiem, Buyse.

Obtained funding: Heagerty, Flum.

Administrative, technical, or material support: Salvaggio, Monsell, Chiem, Flum.

Supervision: Salvaggio, Monsell, Heagerty, Chiem.

Conflict of Interest Disclosures: Dr Salvaggio reported receiving grants from BioWin during the conduct of the study and being an employee of One2Treat. Dr Monsell reported receiving grants from the Patient-Centered Outcomes Research Institute (PCORI) during the conduct of the study. Dr Heagerty reported receiving grants from the National Institutes of Health during the conduct of the study. Dr Saad reported having a patent for 18/653,133 pending and being an employee of IDDI during the conduct of the study. Dr Buyse reported stock ownership in IDDI and One2Treat during the conduct of the study and outside the submitted work. No other disclosures were reported.

Funding/Support: This work was supported in part by the Government of Wallonia, Belgium (BioWin Consortium Agreement No. 7979). The CODA Trial was funded by a PCORI Award (No. 1409-240099).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Elwyn G, Frosch D, Thomson R, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med*. 2012;27(10):1361-1367. doi:10.1007/s11606-012-2077-6

2. Mitropoulou P, Grüner-Hegge N, Reinhold J, Papadopoulou C. Shared decision making in cardiology: a systematic review and meta-analysis. *Heart*. 2022;109(1):34-39. doi:10.1136/heartjnl-2022-321050

3. Niburski K, Guadagno E, Abbasgholizadeh-Rahimi S, Poenaru D. Shared decision making in surgery: a meta-analysis of existing literature. *Patient*. 2020;13(6):667-681. doi:10.1007/s40271-020-00443-6

4. Niburski K, Guadagno E, Mohtashami S, Poenaru D. Shared decision making in surgery: a scoping review of the literature. *Health Expect*. 2020;23(5):1241-1249. doi:10.1111/hex.13105

5. Stacey D, Volk RJ, IPDAS Evidence Update Leads (Hilary Bekker, Karina Dahl Steffensen, Tammy C. Hoffmann, Kirsten McCaffery, Rachel Thompson, Richard Thomson, Lyndal Trevena, Trudy van der Weijden, and Holly Witteman). The International Patient Decision Aid Standards (IPDAS) Collaboration: evidence update 2.0. *Med Decis Making*. 2021;41(7):729-733. doi:10.1177/0272989X211035681

6. Agoritsas T, Heen AF, Brandt L, et al. Decision aids that really promote shared decision making: the pace quickens. *BMJ*. 2015;350:g7624. doi:10.1136/bmj.g7624

7. Rosen JE, Flum DR, Davidson GH, Liao JM. Randomized pilot test of a decision support tool for acute appendicitis: decisional conflict and acceptability in a healthy population. *Ann Surg Open*. 2022;3(4):e213. doi:10.1097/AS9.0000000000000213

8. Soekhai V, Whichello C, Levitan B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(7):1324-1331. doi:10.1016/j.drudis.2019.05.001

9. Thokala P, Devlin N, Marsh K, et al. Multiple criteria decision analysis for health care decision making–an introduction: report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2016;19(1):1-13. doi:10.1016/j.jval.2015.12.003

10. Dodgson JS. Multi-criteria analysis: a manual. Department for Communities and Local Government. January 2009. Accessed December 24, 2024. https://assets.publishing.service.gov.uk/media/5a790545e5274a2acd18b975/1132618.pdf

11. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med*. 2010;29(30):3245-3257. doi:10.1002/sim.3923

12. Evans SR, Rubin D, Follmann D, et al. Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). *Clin Infect Dis*. 2015;61(5):800-806. doi:10.1093/cid/civ495

13. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33(2):176-182. doi:10.1093/eurheartj/ehr352

14. Butler J, Stockbridge N, Packer M. Win ratio: a seductive but potentially misleading method for evaluating evidence from clinical trials. *Circulation*. 2024;149(20):1546-1548. doi:10.1161/CIRCULATIONAHA.123.067786

15. Buyse M, Saad ED, Peron J, et al. The net benefit of a treatment should take the correlation between benefits and harms into account. *J Clin Epidemiol*. 2021;137:148-158. doi:10.1016/j.jclinepi.2021.03.018

16. Maurer MS, Schwartz JH, Gundapaneni B, et al; ATTR-ACT Study Investigators. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med*. 2018;379(11):1007-1016. doi:10.1056/NEJMoa1805689

17. Voors AA, Angermann CE, Teerlink JR, et al. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial. *Nat Med*. 2022;28(3):568-574. doi:10.1038/s41591-021-01659-1

18. Berwanger O, Pfeffer M, Claggett B, et al. Sacubitril/valsartan versus ramipril for patients with acute myocardial infarction: win-ratio analysis of the PARADISE-MI trial. *Eur J Heart Fail*. 2022;24(10):1918-1927. doi:10.1002/ejhf.2663

19. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med*. 1999;18(11):1341-1354. doi:10.1002/(SICI)1097-0258(19990615)18:11<1341::AID-SIM129>3.0.CO;2-7

20. Bodemer N, Meder B, Gigerenzer G. Communicating relative risk changes with baseline risk: presentation format and numeracy matter. *Med Decis Making*. 2014;34(5):615-626. doi:10.1177/0272989X14526305

21. Péron J, Lambert A, Munier S, et al. Assessing long-term survival benefits of immune checkpoint inhibitors using the net survival benefit. *J Natl Cancer Inst*. 2019;111(11):1186-1191. doi:10.1093/jnci/djz030

22. Péron J, Roy P, Ozenne B, Roche L, Buyse M. The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials. *JAMA Oncol*. 2016;2(7):901-905. doi:10.1001/jamaoncol.2015.6359

23. Deltuvaite-Thomas V, De Backer M, Parker S, et al. Generalized pairwise comparisons of prioritized outcomes are a powerful and patient-centric analysis of multi-domain scores. *Orphanet J Rare Dis*. 2023;18(1):321. doi:10.1186/s13023-023-02943-8

24. Patterson KN, Deans KJ, Minneci PC. Shared decision-making in pediatric surgery: an overview of its application for the treatment of uncomplicated appendicitis. *J Pediatr Surg*. 2023;58(4):729-734. doi:10.1016/j.jpedsurg.2022.10.009

25. Sallinen V, Akl EA, You JJ, et al. Meta-analysis of antibiotics versus appendicectomy for non-perforated acute appendicitis. *Br J Surg*. 2016;103(6):656-667. doi:10.1002/bjs.10147

26. Flum DR, Davidson GH, Monsell SE, et al; CODA Collaborative. A randomized trial comparing antibiotics with appendectomy for appendicitis. *N Engl J Med*. 2020;383(20):1907-1919. doi:10.1056/NEJMoa2014320

27. Herrod PJJ, Kwok AT, Lobo DN. Randomized clinical trials comparing antibiotic therapy with appendicectomy for uncomplicated acute appendicitis: meta-analysis. *BJS Open*. 2022;6(4):6. doi:10.1093/bjsopen/zrac100

28. Knaapen M, de Wind A, van der Lee JH, et al. Implementing nonoperative treatment strategy for simple pediatric appendicitis: a qualitative study. *J Surg Res*. 2022;279:218-227. doi:10.1016/j.jss.2022.06.011

29. Hanson AL, Crosby RD, Basson MD. Patient preferences for surgery or antibiotics for the treatment of acute appendicitis. *JAMA Surg*. 2018;153(5):471-478. doi:10.1001/jamasurg.2017.5310

30. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9

31. Le QA, Doctor JN, Zoellner LA, Feeny NC. Minimal clinically important differences for the EQ-5D and QWB-SA in post-traumatic stress disorder (PTSD): results from a doubly randomized preference trial (DRPT). *Health Qual Life Outcomes*. 2013;11:59. doi:10.1186/1477-7525-11-59

32. Appendicitis treatment options. Accessed February 19, 2025. https://appyornot.org/

33. Bebu I, Lachin JM. Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics*. 2016;17(1):178-187. doi:10.1093/biostatistics/kxv032

34. Fay MP, Brittain EH, Shih JH, Follmann DA, Gabriel EE. Causal estimands and confidence intervals associated with Wilcoxon-Mann-Whitney tests in randomized experiments. *Stat Med*. 2018;37(20):2923-2937. doi:10.1002/sim.7799

35. Gladwell M. *Blink: The Power of Thinking Without Thinking*. Black Bay Books; 2005.

36. Rosen JE, Agrawal N, Flum DR, Liao JM. Willingness to undergo antibiotic treatment of acute appendicitis based on risk of treatment failure. *Br J Surg*. 2021;108(11):e361-e363. doi:10.1093/bjs/znab280

37. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*. 2012;15(6):940-947. doi:10.1016/j.jval.2012.05.004

38. Minneci PC, Cooper JN, Leonhart K, et al. Effects of a patient activation tool on decision making between surgery and nonoperative management for pediatric appendicitis: a randomized clinical trial. *JAMA Netw Open*. 2019;2(6):e195009. doi:10.1001/jamanetworkopen.2019.5009

39. Rosen JE, Agrawal N, Flum DR, Liao JM. Verbal descriptions of the probability of treatment complications lead to high variability in risk perceptions: a survey study. *Ann Surg*. 2023;277(4):e766-e771. doi:10.1097/SLA.0000000000005255

SUPPLEMENT 1.

eAppendix. Additional Methods
eReferences.
eTable 1. Outcomes Investigated Using GPC and Desirable Values for a Positive NTB
eTable 2. Scenarios Considered in the 3 GPC Analyses
eTable 3. Mean Preference Rankings of Outcomes for 443 Patients Who Used the AppyOrNot Website

SUPPLEMENT 2.

Data Sharing Statement