PCR103

Evaluating the Generalizability of Preference Estimates: Implementing Weighting Strategies in Stated Preference Studies

Divya Mohan[1], Marco Boeri[1,2], Josh Coulter[3], Byron Jones[4], Brett Hauber[3]

[1]OPEN Health, London, UK [2]Queen's University Belfast, UK [3]Pfizer Inc, New York, NY, US [4]Novartis Pharma AG, Basel, CH

INTRODUCTION

- Quantitative Patient Preference Information (PPI) is increasingly used by regulatory agencies like the U.S. FDA to inform benefit-risk assessments[1-3].
- Surveys used in these studies must have a sample that is representative of the intended patient population for the findings to be generalizable[4].
- Achieving sample representativeness is difficult in preference studies, as they often rely on non-probabilistic strategies (like consumer panels or convenience sampling), which can lead to overrepresentation of certain groups (e.g., white, female, or highly educated individuals)[5].
  - Researchers often use purposive sampling to increase diversity on demographic and clinical characteristics for subgroup analysis which may require over sampling of certain groups.
- Sample weighting can be used to test whether a given sample is likely to be generalizable to a specific target population.
- We explore the application of sample weighting strategies to enhance the generalizability of preference study findings by adjusting for imbalances in observable demographic or clinical characteristics between the sample and target population in three use cases.

METHODS

Sample weighting

- The goal of the sample weighting in this study was to adjust the study sample to reflect the known characteristics of a target population (e.g., clinical trial participants or a broader patient population).
  - Two strategies were considered for weighting the sample to reflect the characteristics of a target population in 3 use cases.
    - **Weighting strategy 1** (pre-estimation): weights were applied directly to participants' choice data, modifying the log-likelihood function to influence the estimation of utility parameters.
    - **Weighting strategy 2** (post-estimation): individual-specific preference parameters were estimated using unweighted data, and weights were then applied to these parameters to calculate a weighted sample average of preferences.
    - The resulting coefficients were used to calculate relative attribute importance (RAI) and maximum acceptable risk (MAR) estimates for both the sample and the target population in each use case.
    - For use cases 1 and 2, the weights were generated using iterative proportional fitting (IPF) on a set of demographic and clinical characteristics.
    - For use case 3, weights were generated using stratification based on age and gender.

USE CASES

- Data from two previously conducted preference studies were used to demonstrate these strategies. The study samples and target samples for these use cases are presented in **Table 1**.
  - The first study used a discrete-choice experiment (DCE) to estimate preferences for osteoarthritis (OA) treatments in the United States (US)[(6)] and the United Kingdom (UK)[7].
    - The US study consisted of 602 participants and the UK study consisted of 437 participants.
    - For the original analysis, random parameter logit (RPL) and latent class (LC) models were estimated.
  - The second study[8], used a DCE to collect data to estimate preferences for treatments to reduce the risk of rheumatoid arthritis (RA) using a sample of UK residents.
    - A total of 982 participants from the general UK population completed the survey.
    - The original analysis employed RPL models to estimate preference weights.

Table 1: Use Cases

Use Case	Study Sample	Target Population	Application
1	US study sample (patients with OA) (Turk et al., (2020))	Clinical trial participants (Schnitzer et al. (2019))	Preference results from one study population to a clinical trial sample
2	US study sample (patients with OA) (Turk et al., (2020))	UK study sample (patients with OA) (Walsh et al. (2021))	Preference results from one country to another country
3	UK study sample (UK general population) (DiSantostefano et al. (2024))	UK general population (UK Office of National Statistics)	Preference results from a convenience sample to the general population

Weighting parameters

- In each use case, the characteristics used to weight the study sample to reflect the target population were identified.
- For the first and second use cases, both using the OA treatment preference data from Turk et al. (2021), the distributions of gender, race, age, and disease duration were used to weight the study sample to reflect the target populations.
  - The distributions of these characteristics were determined for in the US clinical study target populations[8] and UK patient preference study population[7].
- For the third use case, age and gender were used to weight the study sample to reflect the UK general population.
  - The distributions of these characteristics in the target population were derived from data on the number of people in England and Wales who fell into the corresponding gender-age categories as reported by the UK Office of National Statistics (ONS).

Table 2: Distributions of Characteristics in the Study Samples and Target Populations

Use Case	Sub Categories	Use case 1 N=201		Use case 2 N=171		Use case 3 N=982	
		Sample Distribution – Preference study population	Target Distribution – Clinical study population	Sample Distribution – US study population	Sample Distribution – US study population	Sample Distribution – General Population	Target Distribution – General Population
		(Turk et al. 2020)	(Schnitzer et al. 2019)	(Turk et al. 2020)	(Walsh et al. 2022)	(DiSantostefano et al. 2024)	(ONS)
Gender	Female/Male	0.59/0.41	0.67/0.33	0.59/0.41	0.62/0.38	0.66/0.34	0.52/0.48
Race	White/ Non-white	0.94/0.06	0.81/0.19	0.94/0.06	0.97/0.03	Not used	Not used
Age	Under 63/ Over 63	0.59/0.41	0.5/0.5	0.59/0.41	0.66/0.34	Not used	Not used
	18-29/Not 18-29	Not used	Not used	Not used	Not used	0.22/0.78	0.19/0.81
	30-39/Not 30-39					0.21/0.79	0.17/0.83
	40-49/Not 40-49					0.23/0.77	0.16/0.84
	50-59/Not 50-59					0.21/0.79	00.17/0.83
	60-69/Not 60-69					0.09/0.91	0.14/0.86
	70+/Not 70+					0.03/0.97	0.17/0.83
Disease duration (OA)	<5 years/>5 years	0.68/0.33	0.5/0.5	0.68/0.33	0.59/0.41	N/A	N/A

- In use case 1 and 2, MAR was calculated as the maximum acceptable incremental treatment-related risk of heart attack risk. In use case 3 MAR was calculated for the risk of serious infection.
- Differences in RAI and MAR estimates between the sample and target populations in each use case were compared using the confidence intervals for each measure.
  - The 95% confidence intervals for RAI and MAR were calculated using the delta strategy [(10)] in the unweighted RPL model and in the RPL model weighted using weighting strategy 1.
  - In weighting strategy 2, the mean preference coefficients (at the sample level) were used to calculate the RAI and MAR. The RAI and MAR estimates for the sample and target population in each use case were compared.

RESULTS

Relative attribute importance

- Generally, the relative importance rankings remained consistent in the original and weighted samples, with a few exceptions.
- The number of rank reversals is small and the confidence limits for the RAI estimates overlap for each attribute between the study sample and weighting strategy 1.
- There were some cases where the RAI estimate for each attribute derived using weighting strategy 2 (which does not have a confidence interval) does not lay within the confidence intervals for the corresponding estimate from the study sample and the RAI estimate form weighting strategy 1 (**highlighted in table 3**)
- In all use cases, the rank order of the attribute is generally consistent between the study sample and both weighting strategies.
- In use case 3 where we can compare the results from the weighting strategy to the results from a known population (i.e., the UK preference study population), the confidence intervals for the RAI estimate for each attribute in the known population overlap with the corresponding confidence intervals for RAI estimated using both the study sample and weighting strategy 1

Table 3: Relative attribute importance estimates in all three use cases

Attribute	RAI Estimate (Unweighted Sample)	RAI rank	RAI Estimate (Weighted Sample, Strategy 1)	RAI rank	RAI Estimate (Weighted Sample, Strategy 2)	RAI rank	RAI Estimate (True Sample)	RAI rank
Use case 1								
Symptom control (patient global assessment)	25.712 [22.087 ; 29.337]	2	23.183 [19.006 ; 27.359]	2	**29.424**	2	N/A	N/A
Incremental risk of severe rapidly progressive joint problems	4.292 [2.188 ; 6.397]	6	5.712 [3.393 ; 8.03]	6	4.901	6	N/A	N/A
Incremental risk of heart attack	5.658 [3.247 ; 8.069]	5	7.54 [4.281 ; 10.800]	5	5.838	4	N/A	N/A
Treatment-related risk of physical dependence	20.34 [17.394 ; 23.286]	3	20.025 [16.291 ; 23.759]	3	21.637	3	N/A	N/A
Mode and frequency of administration	7.241 [4.338 ; 10.144]	4	7.959 [4.814 ; 11.104]	4	5.830	5	N/A	N/A
Personal (out-of-pocket) monthly cost X Ln(Income)	36.757 [32.962 ; 40.551]	1	35.582 [31.309 ; 39.854]	1	32.328	1	N/A	N/A
LL	-1490.03		-1669.05		-1490.03*		N/A	N/A
Use case 2								
Symptom control (patient global assessment)	25.712 [22.087 ; 29.337]	2	25.523 [20.906 ; 30.14]	2	**30.104**	2	23.127 [17.997 ; 28.256]	2
Incremental risk of severe rapidly progressive joint problems	4.292 [2.188 ; 6.397]	6	5.075 [2.302 ; 7.848]	6	4.827	6	3.552 [0.096 ; 7.008]	6
Incremental risk of heart attack	5.658 [3.247 ; 8.069]	5	5.293 [2.411 ; 8.175]	5	5.702	5	6.19 [2.71 ; 9.67]	4
Treatment-related risk of physical dependence	20.34 [17.394 ; 23.286]	3	18.85 [14.783 ; 22.917]	3	21.634	3	18.222 [14.028 ; 22.416]	3
Mode and frequency of administration	7.241 [4.338 ; 10.144]	4	9.049 [5.544 ; 12.554]	4	5.719	4	4.239 [0.116 ; 8.361]	5
Personal (out-of-pocket) monthly cost X Ln(Income)	36.757 [32.962 ; 40.551]	1	36.21 [30.757 ; 41.663]	1	32.015	1	44.67 [38.394 ; 50.946]	1
LL	-1490.03		-1372.17		-976.17*		-976.17	
Use case 3								
Chance of developing RA reduced from 60% to …	28.71 [25.68 ; 31.75]	1	25.97 [23.26 ; 28.68]	2	**29.52**	1	N/A	N/A
How the treatment is taken	28.05 [25.41 ; 30.7]	2	28.38 [25.87 ; 30.9]	1	**25.47**	2	N/A	N/A
How often the medication has to be taken	14.92 [11.97 ; 17.86]	3	16.03 [13.38 ; 18.68]	3	15.05	3	N/A	N/A
Chance of reversible mild side effects	6.07 [4.34 ; 7.79]	6	5.23 [3.66 ; 6.79]	6	6.08	6	N/A	N/A
Chance of a serious infection	12.57 [10.72 ; 14.42]	4	13.5 [11.79 ; 15.21]	4	12.71	4	N/A	N/A
Chance of a serious, potentially irreversible side effect	9.68 [7.73 ; 11.63]	5	10.89 [9.11 ; 12.67]	5	11.16	5	N/A	N/A
LL	-11942		-11898		-11942*		N/A	N/A

Note: weights were based on gender, race, age and time since diagnosis for use case 1 and 2 and gender and age for use case 3. As in RAI Estimate (Weighted Sample, Strategy 2) the model estimated using the unweighted sample is used to retrieved conditional posterior individual parameters to which the weights are applied, the LL of the two approaches is the same.
LL = loglikelihood; RAI = relative attribute importance.

Maximum acceptable risk results

- MAR estimated remained largely consistent across models before and after weighting, with only minor variations.
- Overlapping confidence intervals suggest that any observed differences are not statistically significant, and the mean value for weighting strategy 2 is within the confidence intervals for the results from both the unweighted sample and estimated derived using weighting strategy 1.

Table 4. Maximum acceptable risk (MAR) estimates

Attribute	From level	To level	MAR Estimate (Unweighted Sample)	MAR Estimate (Weighted Sample, Strategy 1)	MAR Estimate (Weighted Sample, Strategy 2)	MAR Estimate (True Sample)
Use case 1						
Symptom control (patient global assessment)	Poor	Very Good	2.507 [1.287 ; 3.727]	1.962 [1.026 ; 2.898]	2.423	N/A
	Poor	Good	2.191 [1.142 ; 3.239]	1.767 [0.952 ; 2.581]	2.115	N/A
	Poor	Fair	1.510 [0.837 ; 2.184]	1.145 [0.653 ; 1.637]	1.436	N/A
Use case 2						
Symptom control (patient global assessment)	Poor	Very Good	2.507 [1.287 ; 3.727]	2.47 [0.971 ; 3.969]	2.653	2.367 [0.491 ; 4.243]
	Poor	Good	2.191 [1.142 ; 3.239]	2.085 [0.895 ; 3.276]	2.305	1.885 [0.42 ; 3.35]
	Poor	5	1.510 [0.837 ; 2.184]	1.487 [0.686 ; 2.287]	1.565	1.458 [0.397 ; 2.519]
Use case 3						
Chance of developing RA reduced from 60% to …	10%	40%	12.022 [10.105 ; 13.939]	9.829 [8.483 ; 11.175]	10.409	N/A
	20%	40%	7.178 [5.944 ; 8.412]	6.150 [5.241 ; 7.060]	6.043	N/A
	30%	40%	5.073 [4.039 ; 6.107]	4.637 [3.822 ; 5.453]	4.599	N/A

Note: weights were based on gender, race, age and time since diagnosis for use cases 1 and 2 and gender and age for use case 3. As in RAI Estimate (Weighted Sample, Strategy 2) the model estimated using the unweighted sample is used to retrieved conditional posterior individual parameters to which the weights are applied, the LL of the two approaches is the same.

CONCLUSIONS

- The generalizability of preference findings is uncertain because the target population is often not known, and key influencing factors (e.g., prior treatment experience, sociodemographic variables) are inconsistently available in the literature, making this an empirical question.
- Sample weighting is a promising methodological tool to address some of the concerns on generalizability and support the regulatory acceptance of patient preference data.
- The results in this study suggest that, at least in the contexts tested, preference estimates derived from DCEs are relatively stable and sufficiently generalizable to inform decision-making, even when there are differences in the characteristics of the sample and the target population.
- Weighting provides supplementary, empirical evidence to contextualize findings and address concerns about sample representativeness.
- Limitation: Weighting can only adjust for observed variables, and its utility depends on the availability of comparable, high-quality data for both the study sample and the target population.

REFERENCES

1. Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling. 2020. **2.** Center for Devices and Radiological Health. Incorporating Voluntary Patient Preference Information over the Total Product Life Cycle. 2024. **3.** U.S. Department of Health and Human Services, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research. Benefit-Risk Assessment for New Drug and Biological Products. 2023. **4.** Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. International journal of epidemiology. 2013;42(4):1012-4. **5.** Boyle JM, Fakhouri TH, Freedner-Maguire N, Iachan R. Characteristics of the population of internet panel members. Survey Practice. 2017;10(4). **6.** Turk D, Boeri M, Abraham L, Atkinson J, Bushmakin A, Cappelleri JC, et al. Patient preferences for osteoarthritis pain and chronic low back pain treatments in the United States: a discrete-choice experiment. Osteoarthritis and cartilage. 2020;28(9):1202-13. **7.** Walsh DA, Boeri M, Abraham L, Atkinson J, Bushmakin AG, Cappelleri JC, et al. Exploring patient preference heterogeneity for pharmacological treatments for chronic pain: a latent class analysis. European Journal of Pain. 2022;26(3):648-67. **8.** Schnitzer TJ, Easton R, Pang S, Levinson DJ, Pixton G, Viktrup L, et al. Effect of tanezumab on joint pain, physical function, and patient global assessment of osteoarthritis among patients with osteoarthritis of the hip or knee: a randomized clinical trial. Jama. 2019;322(1):37-48. **9.** DiSantostefano RL, Simons G, Englbrecht M, Humphreys JH, Bruce IN, Bywall KS, et al. Can the General Public Be a Proxy for an "At-Risk" Group in a Patient Preference Study? A Disease Prevention Example in Rheumatoid Arthritis. Medical decision making. 2024;44(2):189-202.

Presented at: ISPOR Europe, 9-12 November 2025, Glasgow, Scotland