Physician-Specific Maximum Acceptable Risk in Personalized Medicine: Implications for Medical Decision Making

Boeri, M., McMichael, A. J., Kane, J. P. M., O'Neill, F. A., & Kee, F. (2018). Physician-Specific Maximum Acceptable Risk in Personalized Medicine: Implications for Medical Decision Making. *Medical Decision Making*, *38*(5), 593-600. https://doi.org/10.1177/0272989X18758279

Published in:
Medical Decision Making

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Title: Physician-specific maximum acceptable risk in personalised medicine: implications for medical decision making

Short title: Physician-specific maximum acceptable risk

Authors: Marco Boeri[1] PhD, Alan J McMichael[2] BSc, Joseph PM Kane[3], Francis A O'Neill[2] MD, Frank Kee[2] MD

[1] Health Preference Assessment, RTI Health Solutions, Research Triangle Park, NC, USA

[2] UKCRC Centre of Excellence for Public Health, Queen's University Belfast, Royal Victoria Hospital, Belfast, BT12 6BA

[3] Institute of Neuroscience, Newcastle University, Campus of Ageing and Vitality, Newcastle Upon Tyne, NE4 5PL

Corresponding author

Alan McMichael, PhD[1]
Institute for Clinical Sciences Block B,
Queen's University Belfast,
BT12 6BA
Telephone number: +44 (0)28 90635009
Fax number: +44 (0)28 90235900

[1] Financial support for this study was provided by a grant from the Department of Education and Learning. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Abstract

Background: In discrete-choice experiments (DCEs), respondents are presented with a series of scenarios and asked to select their preferred choice. In clinical decision making, DCEs allow one to calculate the maximum acceptable risk (MAR) that a respondent is willing to accept for a one-unit increase in treatment efficacy. Most published studies report the average MAR for the whole sample, without conveying any information about heterogeneity. For a sample of psychiatrists prescribing drugs for a series of hypothetical patients with schizophrenia, this article demonstrates how heterogeneity accounted for in the DCE modelling can be incorporated in the derivation of the MAR. **Methods:** Psychiatrists were given information about a group of patients' responses to treatment on the Positive and Negative Syndrome Scale (PANSS) and the weight gain associated with the treatment observed in a series of 26 vignettes. We estimated a random parameters logit (RPL) model with treatment choice as the dependent variable. **Results:** Results from the RPL were used to compute the MAR for the overall sample. This was found to be equal to 4%, implying that, overall, psychiatrists were willing to accept a 4% increase in the risk of an adverse event to obtain a one-unit improvement of symptoms – measured on the PANSS. Heterogeneity was then incorporated in the MAR calculation, finding that MARs ranged between 0.5 and 9.5 across the sample of psychiatrists. **Limitations:** We provided psychiatrists with hypothetical scenarios, and their MAR may change when making decisions for actual patients. **Conclusions:** This analysis aimed to show how it is possible to calculate physician-specific MARs and to discuss how MAR heterogeneity could have implications for medical practice.

INTRODUCTION

The practice of evidence-based medicine involves a judicious combination of clinical expertise with the best available external clinical evidence from systematic research.[1,2] However, this dynamic process also requires a continuous re-appraisal of the patient's condition and the expertise to revise a plan when the treatment benefits decline or when the treatment is not working as anticipated.[3] For instance, clinicians must judge whether a patient is responding to treatment and often have to trade off the risks and benefits of different treatment options. Prescribing a treatment with risks that the patient is not willing to tolerate may result in poor treatment adherence and low patient satisfaction.[4] It is therefore important to understand whether these trade-offs vary significantly across clinicians, which could, in turn, result in unacceptable outcome variations and inequities in care.

Discrete-choice experiments (DCEs) have become widely employed to investigate patient and clinician preferences for different treatment options.[5–8] In general a DCE survey involves presenting respondents with a series of choices between two or more alternative hypothetical treatments that vary on a range of attributes (characteristics of the treatment such as efficacy, benefits, and risks) and their levels and asking respondents to indicate their preferred choice.[9]

Additionally, in estimating preference weights for different treatment characteristics, DCEs allow researchers to investigate how patients or clinicians trade off the risks and benefits of treatment. In two previous articles,[10,11] we investigated what thresholds psychiatrists use when determining whether a patient is responding to treatment and whether psychiatrists were influenced by the patient's genotype when recommending a treatment. However, DCEs also enable the researcher to derive a marginal rate of substitutions between positively perceived attributes (such as benefits from a treatment) and negatively perceived attributes (such as costs

or risks of adverse events linked to a treatment). Specifically, this article focusses on how to retrieve the maximum acceptable risk (MAR) that a psychiatrist is willing to accept in exchange for a one-unit increase in a benefit associated with the treatment.[12] In this case, MAR represents the highest risk of an adverse event that a psychiatrist perceives would equal the benefits of prescribing the treatment.

Every clinician has a specific MAR threshold, which may differ from that of his or her colleagues. However, most studies only focus on the overall average MAR for the sample, without informing the reader about this possible heterogeneity across clinicians, even though the typical modelling approach adopting a random parameters logit (RPL) model accounts for it.[13–15] This study presents two methods widely used in preference analysis applied to other fields such as transportation and environmental economics to incorporate preference heterogeneity when computing marginal rates of substitution (e.g., MAR). We demonstrate this using data from a previous study with a sample of psychiatrists[10,11] in which we aim to explore the extent of MAR heterogeneity across our sample of psychiatrists clinicians when prescribing treatments to their patients. The idea is to exploit the outcome from the RPL model, which allows us firstly to derive the unconditional distributions for each parameter – and therefore the MAR – and secondly to retrieve individual posterior estimates conditional to the sequence of choices observed for each individual psychiatrist in the sample. Finally, some of the implications for clinical practice and decision making are briefly discussed. This new study uses the same data set that was used in two previous publications.[10,11]

METHODS

Setting and study procedure

The study was developed following good DCE research practice to understand how psychiatrists evaluate different options to treat schizophrenia and to explore the trade-offs they

are willing to accept between benefits and risks of adverse events.[16] The survey instrument was developed and pretested with practising psychiatrists to identify the appropriate attributes and levels to be used in a series of vignettes and to construct realistic choice tasks. The protocol for the study was approved by the Queen's University School of Psychology ethics committee, and each respondent provided written informed consent.

Once the survey instrument was ready for fielding, we used a convenience sample including all psychiatrists who attended their monthly continuous professional development training meetings in three different hospitals across Northern Ireland. Sample size was dictated by the planned meetings but allowed us to interview psychiatrists in person, enhancing data quality and control over the experiment.

During data collection, members of the study team conducted both pre-test and final interviews in person. This allowed the researchers to explain the task if required, without leading the participants. In total, 70 psychiatrists were invited to participate in the study. They were asked to judge a series of hypothetical patient vignettes and rate their confidence in each judgement. One declined to participate, and two did not provide useable data. Therefore, the final sample included 67 psychiatrists. More than half (59%) were male, most (64%) had completed their specialist training, and the average years of clinical experience in their specialty was 10 years (SD = 7.19 years).

Materials

We recently reported on how we developed a series of 26 vignettes[10,11] constructed to depict a patient's pre- and post-treatment score on the positive sub-scale of the Positive and Negative Syndrome Scale (PANSS). The attributes and levels were determined with the help of two practising psychiatrists (Francis O'Neill and Joe Kane). See Figure 1 for an example of the vignette. The positive sub-scale contains seven questions relating to the positive symptoms

of schizophrenia, such as hostility and delusions. Each question is scored on a scale of 1 (symptom absent) to 7 (symptom extreme). Each respondent's pre-treatment score on the PANSS was anchored at 42, indicating severe symptoms before any treatment. Psychiatrists were informed of (1) the patient's change score on the PANSS (range 3-26); (2) whether the patient had a hyper-responsiveness genotype, indicating a hyper-responsiveness to treatment (dummy coded as 0 if hyper-responsiveness genotype was not present and 1 if hyper-responsiveness genotype was present); (3) the number of anticipated acute treatment days in hospital (17-45 days); and (4) the risk of a 10 kg weight gain associated with the treatment (range 30%-70%).

[Table 1 about here]

Psychiatrists were asked to choose which of two treatments they would be willing to prescribe to the patient. Participant information sheets, participant informed consent forms, and a sample survey instrument are included in Appendix 1.

The DCE experimental design was developed using a commonly used D-efficient algorithm to construct a fractional factorial experimental design as implemented in NGENE software.[17–19] We used a semi-Bayesian design, since we applied prior information to build the design but did not update the priors after surveying several clinicians.[20] An efficient design will have a "small" variance-covariance matrix, and D-efficiency is one measure that can be used to compare the relative efficiency of one experimental design with another. D-efficiency ranges between 0 and 100, and higher D-efficiency scores indicate that a particular design is better able than another to capture the trade-offs among attribute levels required to estimate the preference model.[18]

[Figure 1 about here]

Statistical analysis

Preference data were analysed by estimating an RPL model, with treatment choice as the dependent variable and the attributes presented in the DCE as independent variables, employing normal distributions to accommodate preference heterogeneity for each attribute.

The RPL model has been extensively applied in health preference assessment as it allows for heterogeneity across the respondents' preferences.[16,21] Generally, most DCE studies that have focussed on health preference assessment or medical decision making report the mean value of the preference weights – i.e., only reporting the mean of the distribution estimated in the RPL model. Following this trend, we firstly calculate an average MAR across all the psychiatrists by dividing the estimated coefficient for the change score on the PANSS by the estimated coefficient for the risk of gaining weight. The negative of this ratio is the average MAR of weight increase for a one-unit improvement in the PANSS.

However, this average MAR does not account for preference heterogeneity, even if it is retrieved from the RPL model. A second option is to account for the heterogeneity within the RPL model by employing the estimated normal distribution in the computation of the MAR. This can be accomplished with two approaches: (1) by deriving an unconditional distribution based on the estimated mean and standard deviation of both the symptom score and the risk of weight gain or (2) by estimating a specific MAR for each psychiatrist by computing individual posterior parameters conditional on the sequence of their choices and the estimated normal distribution for the relevant parameters.

Following the former approach, implemented through Monte Carlo sampling, we drew 50,000 times from the estimated normal distributions of the change score and the risk, computing the ratios, and we then ordered and plotted the results from the lowest to the highest MAR. This exploits the ability of the RPL to accommodate heterogeneity; however, it does not

allow the researcher to fully use the information contained in the sequence of choices that each respondent made. Indeed, the RPL model facilitates the computation of individual-specific preferences by deriving individuals' conditional parameter distribution based on their choices.[22–25] Those choices can be considered prior knowledge of the individual preferences, and therefore it is possible to invoke Bayes' theorem to derive the expected posterior values of the individual parameters given the observed sequence of their choices:

$$\hat{\beta}_n = \frac{\frac{1}{R}\sum_{r=1}^{R} \widehat{\beta_{x,n}^r}\, L(\widehat{\beta_{x,n}^r}|y_n, \pi_n)}{\frac{1}{R}\sum_{r=1}^{R} L(\widehat{\beta_{x,n}^r}|y_n, \pi_n)} \qquad (1)$$

where $\hat{\beta}_n$ is the individual vector of parameters for respondent *n* for the observed sequence of choices, *r* denotes the generic draw of a random coefficient, and *R* denotes the total number of draws (50,000 in this case). We consider this method particularly suitable to our sample, given the long series of choices presented to the physicians.

The MAR is then computed as the reciprocal of the ratio of the benefit (change score) and the risk parameters. In the results section, we present all the options depicted above and we discuss some differences. Statistical analysis was conducted in Biogeme version 2.3 and, for comparison, with STATA version 12. Individual posterior conditional parameters were computed in R version 3.4 (syntax code is included in Appendix 2).

RESULTS

Results from the RPL model are highlighted in Table 1. For each attribute, the model estimates a mean and a standard deviation of the normal distribution assumed to accommodate preference heterogeneity. So, for example, the normal distribution associated with the change score on the PANSS resulted in a mean estimate of 0.439 and a standard error of 0.190. For every decrease of one unit in symptoms on the PANSS, psychiatrists were significantly more

likely to recommend the associated treatment, with a relatively small but highly significant heterogeneity. Conversely, for every increase of one unit in weight gain or extra days in hospital associated with the treatment, psychiatrists were less likely to recommend the treatment.

Using the mean estimate for change score and risk of weight gain from the RPL model, we first estimated an average MAR. In this case, the MAR was 4.06, implying that, overall, the psychiatrists were willing to accept a 4% increase in the risk of the patient gaining 10 kg of weight to obtain a one-unit decrease in symptoms as measured by the PANSS.

[Table 2 about here]

Our next step was to generate an unconditional distribution of the MAR using the results (both mean and standard deviation) from the RPL model in Table 1. This allows us to better characterise the extent of the heterogeneity in the sample. As can be observed in Figure 2, unconditional MAR estimates based on the normal distribution, estimated with the RPL model, range between minus and plus infinity, a direct consequence of the assumption of a normal distribution for preference heterogeneity. The distribution reported in Figure 2 shows a MAR that varies across the sample between −2 and 18 and above (due to the shape of the normal distribution, the figure includes extreme values).

[Figure 2 about here]

The results are richer when we fully utilise the entire sequence of choices that each psychiatrist made when evaluating the 26 vignettes. Indeed, it is possible to retrieve the posterior conditional parameters for each decision maker using Bayes' theorem (as described previously). As illustrated in Figure 3, using the choices made by each clinician, we retrieved the individual MARs conditional on each clinician's pattern of choices. Inspection of Figure 3

highlights a narrower range of MAR across respondents compared with the reported unconditional distribution in Figure 2. Nevertheless, the physicians most averse to risk had a MAR of about 2, while the least risk-averse psychiatrists had a MAR of about 10.

[Figure 3 about here]

DISCUSSION

DCEs are increasingly used to assess preferences and to help understand how patients and clinicians trade off the risks and benefits associated with different treatment options[15]. DCEs also permit the estimation of the MAR that a decision maker might tolerate to gain a one-unit increase in treatment benefit[26]. However, in many previous studies where it has been estimated in an RPL, the derived MAR represents the average risk that respondents are willing to accept, ignoring the possibility that decision makers may make different benefit-risk trade-offs.[27] Thus our aim was to highlight a method that could be used to understand the extent of MAR variation across samples of decision makers. We have demonstrated a substantial variation in MARs across a sample of psychiatrists in one specialty in one region of the United Kingdom.

In our analysis, we estimated an RPL model to analyse the psychiatrists' preferences for different treatments for patients with schizophrenia. We used the results to compute the MAR at sample level and at individual level. Firstly, we looked at the mean estimate for benefit (change score on the PANSS) and risk of the treatment (probability of 10 kg weight gain) to compute an overall average MAR for the sample, finding a MAR of about 4%, implying that, overall, psychiatrists were willing to accept a 4% increase in the risk of adverse events to obtain a one-unit decrease in symptom score. (The remaining attributes were analysed in our previous articles.)[10,11] This approach is often presented in the literature because it is simple to describe and directly computed from the outputs of the model. However, results from the RPL model

are much richer as they consider the entire sequence of choices and the heterogeneity across clinicians. Therefore, we argue that researchers who assume heterogeneity in preferences (and estimate an RPL model as a consequence) should at least explore the consequences on the MAR of including such heterogeneity in the MAR calculation.

We present two established methods to retrieve MAR incorporating preference heterogeneity: by deriving an unconditional distribution based on the estimated mean and standard deviation of both change score on the PANSS and the risk of weight gain and by computing individual posterior-based MAR for each psychiatrist conditional on the sequence of their choices and the estimated normal distribution for the relevant parameters. The two methods retrieve similar patterns of MAR across the sample, with slightly different outputs but, at least in this case, similar clinical implications. Once the three methods are used to explore MAR and sample heterogeneity, selecting the best representation depends on the aims of the study and the research questions and remains an empirical question. In our case, we think the MAR based on the individual parameters best addresses our question. Overall, we find that physicians in this sample have an individual MAR that ranges between 0.5 and 9.5

These results have some important clinical implications. Champions of evidence-based medicine often advocate for a reduction in unexplained and unjustified variations in clinical care.[2] Though our study focussed on the judgement of response to treatment, this is critical to decisions to continue with an evidence-based treatment, and in such cases, decisions must balance the perceived benefits and risks. Whereas many conceptual models of treatment thresholds are based on perceived benefits and expected utility, equally valid threshold models account for anticipated regret attendant on the risk of treatment.[28] Indeed, these hybrid models predict that the variation in the decision-making styles of clinicians, attendant on their varying benefit-risk trade-offs, is consonant with a duel process model of decision making.[29]

Several previous studies have shown a discordance between the MARs elicited from clinicians and those from patients.[30] Both are important to personalised medicine. If a patient is not willing to tolerate the risks associated with a treatment, this may potentially affect treatment adherence[4] and may lead to the inefficient allocation of health resources.[31] Furthermore Pravettoni et al. have argued that a "P4" paradigm (predictive, personalised, preventive, and participatory) for precision medicine is inadequate if it does not also consider individual psychology, advocating instead for a "P5" approach that accounts for the way patients may make differing judgements of risk and benefits.[32] A similar perspective is offered by Rogowski et al. who have argued that while a greater research effort has so far gone into uncovering the physiological ways to personalise medicine (with genotypes and biomarkers), an equally valid approach is to reveal patient preferences and their benefit-risk tolerability thresholds elicited through DCEs.[33] Though our own study focussed only on the clinicians' thresholds, we recognise that it might be problematic to design vignettes for valid interpretation by patients with psychosis.

Our study does have some limitations.[10,11] For instance, we only looked at the MARs of psychiatrists, and future research is warranted to assess the MARs of clinicians who practice in other areas of medicine. We asked psychiatrists to assess a hypothetical patient only once, whereas in routine clinical practice, psychiatrists often see patients on many occasions, with supplementary information also being used for clinical assessments. Additionally, we did not explicitly tell psychiatrists how long each patient was taking their treatment, which may have influenced the psychiatrists' MAR. Finally, in an ideal study, and following good research practice for DCE, the survey instrument should be focussed and only include one research question. However, to maximise the information gained from this study, we addressed several related research questions in our survey instrument (e.g., when do psychiatrists perceive a patient is responding to a treatment, how do they consider the information when making

treatment decisions, and which trade-offs are they willing to accept), and it would not be possible to reach this sample for a second study to address the excluded research questions. This is not uncommon in social science, but we recognise it as a limitation in some occasions.

The aim of the present article was to show how individual clinicians' specific DCE parameters can be estimated for the MAR. In some contexts, there may be good reasons for why clinicians' or patients' MAR may vary. Risk perception and risk aversion traits have been associated with several other psychological characteristics that were not measured in our study. If a P5 model for personalised medicine gains traction, they merit exploration in future research to see how they might relate to individual MARs. In summary, we have shown how it is possible to estimate a MAR for individual clinicians and have demonstrated a substantial range of MARs for clinicians in one region. Understanding clinician and patient preferences as reflected in individual MARs has implications for personalised medicine, for it is essential that physicians and patients share a common understanding of the risks and benefits associated with treatment.

REFERENCES

1. Sackett D, Rosenberg W, Gray J, et al. Evidence based medicine: what it is and what it isn'thttp://www.bmj.com/content/312/7023/71.short (1996, accessed 19 July 2017).

2. Bhandari M, Giannoudis P V. Evidence-based medicine: what it is and what it is not. *Injury* 2006; 37: 302–306.

3. Klein D, Woods D, Klein G. Can we trust best practices? Six cognitive challenges of evidence-based approaches. *J Cogn*http://journals.sagepub.com/doi/abs/10.1177/1555343416637520 (2016, accessed 4 April 2017).

4. Hauber a B, Mohamed a F, Johnson FR, et al. Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents. *Diabet Med* 2009; 26: 416–424.

5. Vass C, Gray E, Payne K. Discrete choice experiments of pharmacy services: a systematic review. *Int J Clin Pharm* 2015; 38: 620–30.

6. Bien DR, Danner M, Vennedey V, et al. Patients' Preferences for Outcome, Process and Cost Attributes in Cancer Treatment: A Systematic Review of Discrete Choice Experiments. *Patient - Patient-Centered Outcomes Res*. Epub ahead of print 31 March 2017. DOI: 10.1007/s40271-017-0235-y.

7. Michaels-Igbokwe C, MacDonald S, Currie GR. Individual Preferences for Child and Adolescent Vaccine Attributes: A Systematic Review of the Stated Preference Literature. *Patient - Patient-Centered Outcomes Res*. Epub ahead of print 4 May 2017. DOI: 10.1007/s40271-017-0244-x.

8. Vass C, Rigby D, Payne K. The Role of Qualitative Research Methods in Discrete Choice Experiments. *Med Decis Mak* 2017; 37: 298–313.

9. Johnson FR, Hauber a. B, Poulos C. *A Brief Introduction to the Use of Stated-Choice Methods to Measure Preferences for Treatment Benefits and Risks*. 1 September 2009. Epub ahead of print 1 September 2009. DOI: 10.3768/rtipress.2009.rr.0009.0909.

10. McMichael A, Rolison J, Boeri M. How Do Psychiatrists Apply the Minimum

Clinically Important Difference to Assess Patient Responses to Treatment? *MDM Policy*http://mpp.sagepub.com/content/1/1/2381468316678855.short (2016, accessed 30 November 2016).

11. McMichael AJ, Boeri M, Rolison JJ, et al. The Influence of Genotype Information on Psychiatrists' Treatment Recommendations: More Experienced Clinicians Know Better What to Ignore. *Value Heal* 2016; 1–6.

12. Van Houtven G, Johnson FR, Kilambi V, et al. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. *Med Decis Making* 2011; 31: 469–480.

13. Okumura K, Inoue H, Yasaka M, et al. Comparing Patient and Physician Risk Tolerance for Bleeding Events Associated with Anticoagulants in Atrial Fibrillation—evidence from the United States and Japan. *Value Heal Reg Issues* 2015; 6: 65–72.

14. Hauber AB, Mohamed AF, Gonzalez JM, et al. Benefit-risk tradeoff preferences for chronic hand eczema treatments. *J Dermatolog Treat* 2017; 28: 40–46.

15. Johnson FR, Hauber B, Ozdemir S, et al. Are gastroenterologists less tolerant of treatment risks than patients? Benefit-risk preferences in Crohn's disease management. *J Manag Care Pharm* 2010; 16: 616–628.

16. Bridges JFP, Hauber a B, Marshall D, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health* 2011; 14: 403–13.

17. Kuhfeld WF. *Marketing Research Methods in {SAS:} Experimental Design, Choice, Conjoint,and Graphical Techniques*. Cary, NC, USA, 2004.

18. Kuhfeld WF, Tobias RD, Garratt M. Efficient experimental design with marketing research applications. *J Mark Res* 1994; 31: 545–557.

19. Rose, J. M., Collins, A. T., Bliemer, M. C., & Hensher DA. Ngene 1.0 stated choice experiment design software.

20. Ferrini S, Scarpa R. Designs with a-priori information for nonmarket valuation with choice-experiments: a {Monte Carlo} study. *J Environ Econ Manage* 2007; 53: 342–363.

21. Hauber A, González J, Groothuis-Oudshoorn C. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Heal*http://www.sciencedirect.com/science/article/pii/S1098301516302911 (2016, accessed 1 June 2017).

22. Huber J, Train K. On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. *Mark Lett* 2001; 12: 259–269.

23. Greene WH, Hensher DA, Rose JM. Using Classical Simulation-Based Estimators to Estimate Individual {WTP} Values. In: Scarpa R, Alberini A (eds) *Applications of simulation methods in environmental and resource economics*. Dordrecht, The Netherlands: Springer Publisher, 2005, pp. 17–34.

24. Scarpa R, Campbell D, Hutchinson WG. Benefit estimates for landscape improvements: sequential {Bayesian} design and respondents' rationality in a choice experiment study. *Land Econ* 2007; 83: 617–634.

25. Train K. *Discrete Choice Methods with Simulation*. Second edi. New York: Cambridge University Press, 2009.

26. Van Houtven G, Johnson FR, Kilambi V, et al. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. *Med Decis Making* 2011; 31: 469–80.

27. Martin RW, McCallops K, Head AJ, et al. Influence of patient characteristics on perceived risks and willingness to take a proposed anti-rheumatic drug. *BMC Med Inform Decis Mak* 2013; 13: 89.

28. Hozo I, Tsalatsanis A. Expected utility versus expected regret theory versions of decision curve analysis do generate different results when treatment effects are taken into account. *J Eval*http://onlinelibrary.wiley.com/doi/10.1111/jep.12676/full (2016, accessed 4 April 2017).

29. Djulbegovic B, Hozo I. Dual processing model of medical decision-making. *BMC Med*https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-12-94 (2012, accessed 4 April 2017).

30. Sonnenberg A. Patient–physician discordance about benefits and risks in gastroenterology decision-making. *Aliment Pharmacol Ther*http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2036.2004.02004.x/full (2004, accessed 4 April 2017).

31. Marckmann G, Rogowski WH, Grosse SD, et al. Criteria for fairly allocating scarce health-care resources to genetic tests : which matter most ? *Eur J Hum Genet* 2014; 25–31.

32. Pravettoni G, Gorini A. A P5 cancer medicine approach: why personalized medicine cannot ignore psychology. *J Eval Clin Pract*http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2753.2011.01709.x/full (2011, accessed 4 April 2017).

33. Rogowski W, Payne K, Schnell-Inderst P, et al. Concepts of 'Personalization' in Personalized Medicine: Implications for Economic Evaluation. *Pharmacoeconomics* 2015; 33: 49–59.

FIGURES

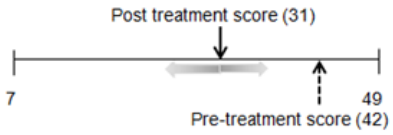
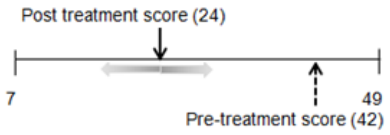
Patient Information	Treatment A	Treatment B
Displayed to the right are the patients' pre- and post-treatment scores on the positive subscale of the PANSS for Treatment A and Treatment B. The numerical values are provided in parenthesis The shaded horizontal arrows represent the 95% confidence intervals for the patients *post treatment* score		
Patient status in respect of *hyper-responsiveness* genotype to this treatment	**Yes**	**No**
Has this patient responded to treatment? (circle as appropriate)	**Yes No**	**Yes No**
How confident are you in your judgement (1=not at all confident, 7=very confident)? (circle as appropriate)	**1 2 3 4 5 6 7**	**1 2 3 4 5 6 7**
Treatment Information	**Treatment A**	**Treatment B**
The cost: acute treatment days in hospital.	**30**	**38**
The percentage probability of a 10kg weight gain in the next six months following the start of treatment	**41%**	**33%**
Based on the information above, which treatment would you recommend? (circle as appropriate)	**Treatment A**	**Treatment B**

Figure 1. Participants had to choose judge whether they believed the patient was responding to treatment, based on the change score on the PANSS. The coefficients for change score on the PANSS and the percentage probability of weight gain were used to derive the average and individual maximum acceptable risks (MARs).

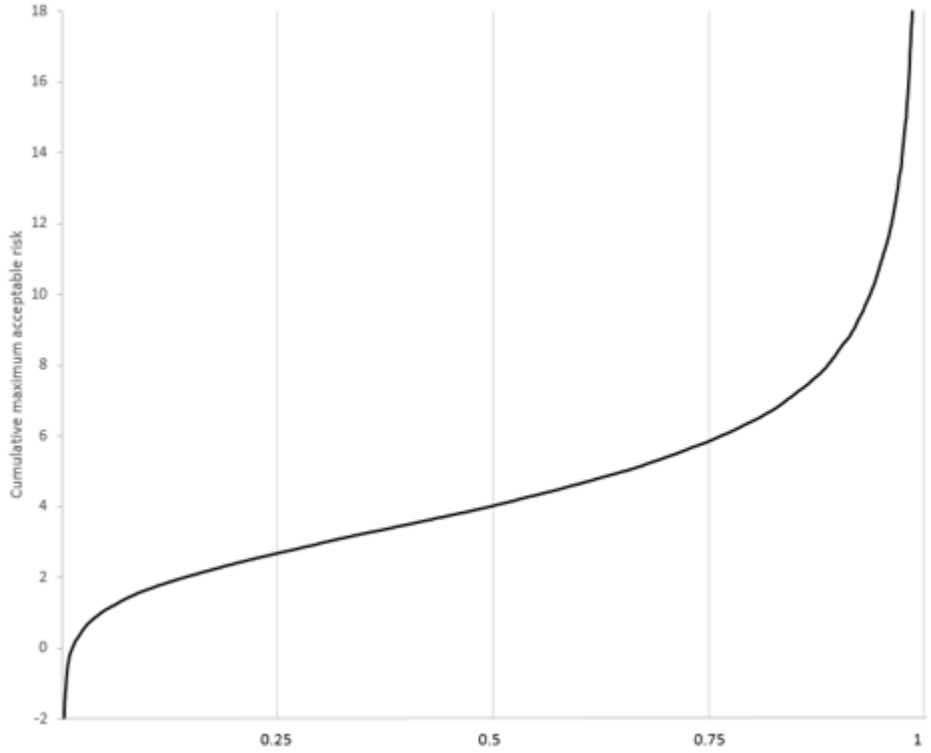


Figure 2. Line graph showing the unconditional distribution of MAR across the sample drawing from the distribution reported in the model's estimates. The draw from the distribution are displayed in increasing order of MAR instead plotted as a Kernel distribution, for simpler comparison with Figure 3.

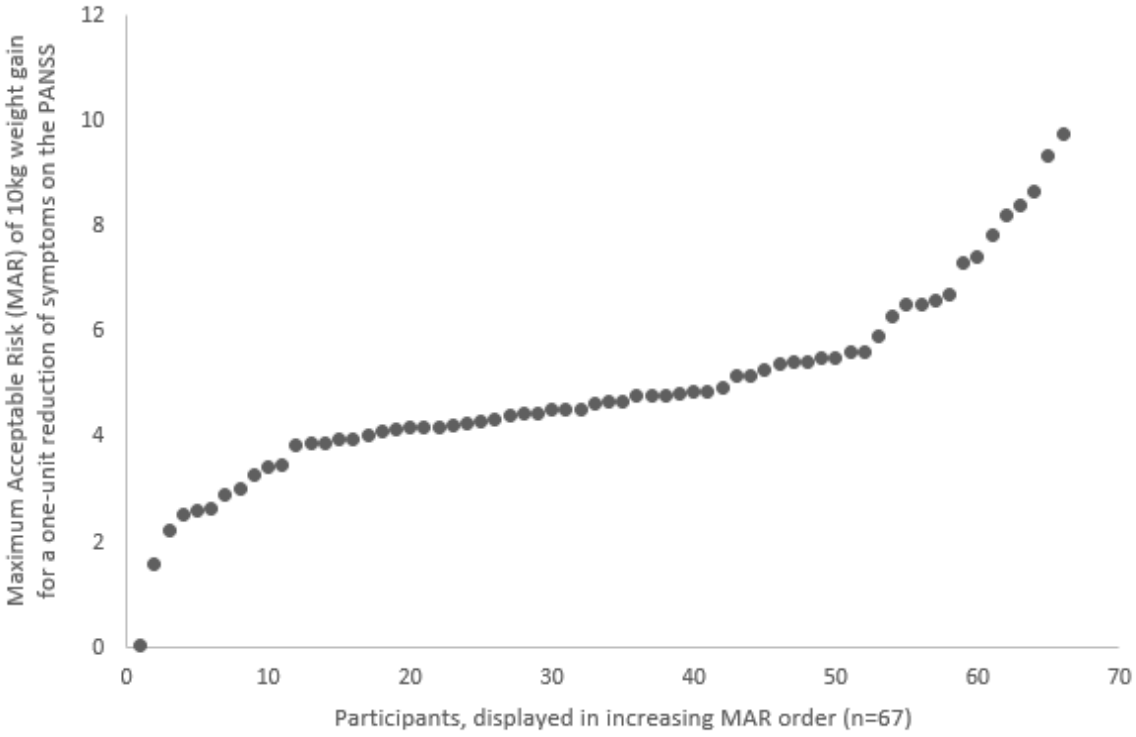


Figure 3. Individual MAR of 10kg weight gain for a one-unit reduction of symptoms on the PANSS scale calculated conditional to the model's estimates and the sequence of choices, displayed in increasing order of MAR.

TABLES

Table 1. Attributes and levels

Attributes	Levels
PANSS change score	• Continuously ranging from 3 to 26 points
Patient have a hyper-responsiveness	• Yes
Genotype	• No
Number of anticipated acute treatment days in hospital	• Continuously ranging from 17 to 45 days
Risk of a 10kg weight gain associated with the treatment	• Continuously ranging from 30% to 70%

Table 2. Random parameters logit with treatment choice as the dependent variable (N = 67, obs. = 1735)

Variable	Estimation	Standard Error	95% confidence interval	
			Lower bound	Upper bound
PANSS change score±	0.439**	0.04	0.3606	0.5174
σ PANSS change score	0.190**	0.03	0.1312	0.2488
Genotype	-0.247	0.17	-0.5802	0.0862
σ Genotype	1.020**	0.17	0.6868	1.3532
Number of acute treatment days in hospital±	-0.106**	0.01	-0.1256	-0.0864
σ Number of acute treatment days in hospital	0.041*	0.02	0.0018	0.0802
Risk of weight gain associated with treatment±	-0.108**	0.01	-0.1276	-0.0884
σ Risk of weight gain associated with treatment	0.038**	0.01	0.0184	0.0576
K (number of parameters)	8			
Log likelihood	-533.073			

*P <.001, **P<.0001, σ = standard deviation (SD) of the normal distribution.

Note: Genotype is a dummy coded variable equal to 1 if hyper-responsiveness genotype was present, zero otherwise. No hyper-responsiveness genotype is the reference category.

± Coefficient refers to a one unit increase in the independent variable, e.g., a one unit increase in acute treatment days in hospital associated with the treatment.