Eur J Health Econ
DOI 10.1007/s10198-013-0558-0

ORIGINAL PAPER

Output correlations in probabilistic models with multiple alternatives

Klemen Naveršnik

Received: 4 July 2013 / Accepted: 18 December 2013

Abstract A comprehensive cost-effectiveness decision model will often go beyond a one-to-one comparison and will include a number of competing alternatives. Only a simultaneous assessment of all relevant treatment alternatives avoids comparing average cost-effectiveness ratios and allows a truly incremental analysis. Two issues arise if the analysis is probabilistic, namely, the occurrence of output correlations and difficulty in presenting the results. I have examined the role of output correlations using a screening model with eight alternatives and have shown that specifically cost–cost and quality-adjusted life years (QALY)–QALY correlations between alternatives have a major impact on decision uncertainty, as measured by the probability of the cost-effectiveness and expected value of perfect information. In particular, the latter strongly depends on between-alternative output correlations. This analysis shows that both the expected value of perfect information plots and acceptability curves/frontiers are sensitive to output correlations and thus appropriate for presentation of multiple alternatives. To avoid confusing statistical significance and economic importance I propose that acceptability curves be augmented by incremental net-benefit density plots at a given willingness to pay threshold.

Keywords Output correlations · Multiple alternatives · Uncertainty analysis · Probabilistic model

JEL Classification C5 · I1

K. Naveršnik (✉)
Lek Pharmaceuticals d.d., Sandoz Development Center, Verovskova 57, 1000 Ljubljana, Slovenia
e-mail: klemen.naversnik@gmail.com; klemen.naversnik@sandoz.com

Introduction

Probabilistic models are commonly used to provide information on resource allocation within healthcare systems. Input parameters are described by probability distributions and propagated through the health economic model by means of second order Monte Carlo simulation. The outputs of a cost-effectiveness model are effectiveness [e.g. quality-adjusted life years (QALYs)] and costs of all relevant alternatives. QALYs and costs are combined to produce summary measures, such as incremental cost-effectiveness ratios (ICER), cost-effectiveness acceptability curves (CEAC) and expected value of perfect information (EVPI). The usual approach is to base decision-making on the means of the output distributions and to estimate uncertainty using the variance in the outputs [1]. The focus of my work is correlations among model outputs (QALYs and costs of all alternatives) and how they relate to decision uncertainty. Output correlations occur due to model structure and should not be confused with correlations between input parameters, which are specified prior to simulation. The aim of my work is first to provide an overview of output correlations and then to use a published model to create scenarios with varying output correlation structures in order to see how they influence CEAC and EVPI.

It is common to express results incrementally when two treatments are compared (e.g. new treatment vs. current practice), resulting in two outputs: incremental QALYs and incremental costs. These can be visualized on the cost-effectiveness plane, thereby allowing easy identification of the optimal alternative and showing decision uncertainty (the spread of simulation results). It is a known fact that output correlation between QALYs and costs has a direct influence on decision uncertainty [2]. Positive correlation

Published online: 04 January 2014

Table 1 Outputs of the screening model

Outputs	Treatment alternatives							
	0	33	29	30	26	27	28	16
Cost (€)								
Mean	283	435	491	522	602	614	632	826
SD	117	111	118	122	137	139	142	180
QALY								
Mean	22.97	23.01	23.02	23.02	23.04	23.04	23.04	23.04
SD	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05
NB (×1,000 €)								
Mean	688.9	690.5	690.5	690.5	690.5	690.1	690.2	689.8
SD	1.2	1.4	1.4	1.4	1.4	1.3	1.4	1.3
ICER (€/QALY)	NA	4,300	5,400	5,800	7,700	8,500	13,000	42,000

SD Standard deviation, *NB* net benefit at 30,000 €/quality-adjusted life years (QALY), *ICER* incremental cost-effectiveness ratio

narrows the ICER interval, may increase the probability of cost-effectiveness and has a significant effect on EVPI, sample size and power calculations [3].

More than two alternatives are often compared within a multiple technology appraisal. In fact, only a simultaneous comparison of all relevant treatment alternatives avoids comparing average CE ratios and allows a truly incremental analysis [4]. More than one treatment is often available for the disease in question, and there is also more than one relevant comparator technology (e.g. "do nothing", "current practice" or "best alternative practice"). Alternative treatment sequences or combinations should also be considered, as they are mutually exclusive. Synthesis of data across multiple alternatives is facilitated by network meta-analyses, mixed treatment comparison or other Bayesian evidence synthesis approaches, allowing greater flexibility than traditional head-to-head comparison [5, 6]. It is therefore logical that a comprehensive analysis of a decision-making problem will involve multiple treatment alternatives, which expands the number of output correlations from the single QALY–cost correlation to a matrix of output correlations, as described herein in a following section.

In the analysis reported here I show how decision-making can be affected by output correlations. The results of a previously published probabilistic model with eight alternatives are presented in Table 1. In this model, monetary net benefit (NB) is calculated for each of the eight alternatives in every simulation cycle ($N = 10{,}000$). Overlaid densities of NB are plotted in Fig. 1a. It is clear from the results shown that all alternatives produced very similar NB values with mean at around 690,000 € at a given threshold of 30,000 €/QALY. With a standard deviation of about 1,400 €, the NB densities of all alternatives are practically indistinguishable and show the absence of any "statistically significant differences". However, this would only hold if the NB values of the eight alternatives were independent (uncorrelated). In contrast, the actual outputs are highly correlated. How this affects the decision-making process only becomes apparent if NBs are calculated incremental to the optimal alternative (one with the highest expected NB) and simulation results are plotted as incremental net benefit (INB) densities (Fig. 1b). From this plot one can see that in the majority of simulations, the INB of the strategies competing with the optimal one is negative. This provides strong evidence in favor of the optimal alternative and is a direct result of output correlations. Not only are the output means important for the decision-making process, but also their correlations.

Output correlations

A model with k alternatives will have $2k$ outputs (i.e. QALYs and the costs of each alternative). These are viewed as random variables with posterior distributions, obtained by Bayesian updating or a result of a forward Monte Carlo simulation. Most often these outputs will not be independent but correlated to some degree. Dependency between the outputs can be quantified by a square correlation matrix (Σ) of dimension $2k$ with columns (or rows) representing k QALYs and k costs. The matrix is symmetric and both off-diagonal parts contain identical information.

Off-diagonal blocks of Σ include:

- QQ correlations among QALYs (Q_I) and CC correlations among costs (C_I) between alternatives;
- QC correlations between QALYs and costs between alternatives;
- QC correlations between QALYs and costs within alternatives.

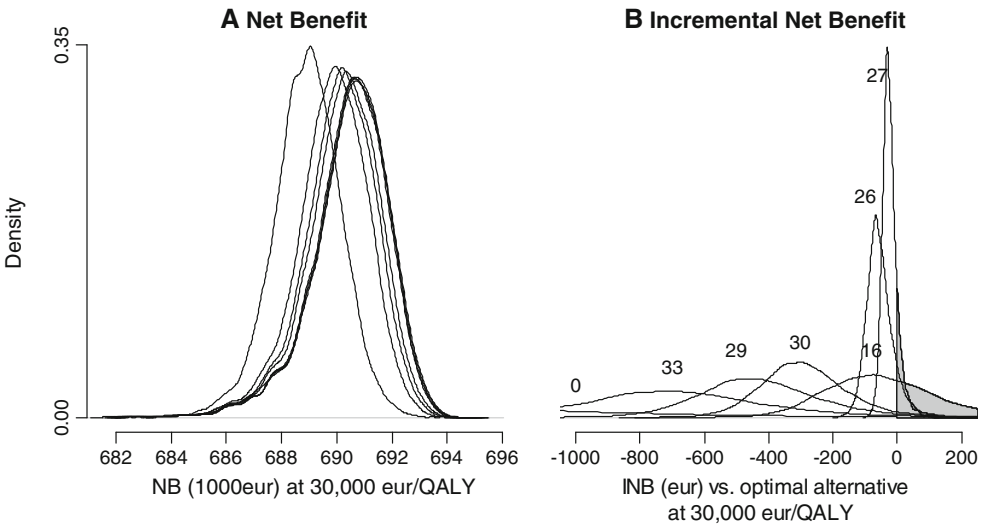

Fig. 1 Screening example (baseline) results. **a** Net benefit (*NB*) density plots of eight alternatives at 30,000 €/quality-adjusted life years (*QALY*), **b** Incremental net benefit (*INB*) (vs. optimal alternative) density plots for seven alternatives at 30,000 €/QALY. *Numbers above lines* denote alternatives (see Table 1)

The choice of an optimal alternative depends only on the marginal distributions of the NB. Under risk neutrality, the cost-effectiveness of competing alternatives is ranked based on the expected values of each marginal NB (Eq. 1), where λ is the willingness to pay threshold (€/QALY).

$$E(NB_I) = E(\lambda Q_I - C_I) = \lambda E(Q_I) - E(C_I) \quad (1)$$

Since NB is a linear function of costs and QALYs, its expectation is not influenced by QC correlations. Correlations thus do not have any impact on the choice of optimal alternative. NB variance, on the other hand, depends on QC correlations within alternatives (Eq. 2). Note that correlation is just a scaled version of covariance.

$$\mathrm{Var}(NB_I) = \lambda^2 \mathrm{Var}(Q_I) + \mathrm{Var}(C_I) - 2\lambda \mathrm{Cov}(Q_I, C_I) \quad (2)$$

Correlation can have a dramatic effect on the confidence interval for the cost-effectiveness ratio, EVPI and sample size/power calculations [3]. This has long been known for single technology appraisal [2] and plays an equally critical role in cases with multiple alternatives [7].

Here, I have shown that the choice of optimal alternative will not depend on QQ or CC correlations. In contrast, the uncertainty in decision-making will depend on the choice of optimal alternative, and this is explored in the following section.

Examining the influence of correlation structure

In this analysis a previously published breast cancer screening model [8] was used to show the influence of output correlations on cost-effectiveness measures. The screening model compares eight alternative screening sequences (annotated: 0, 16, 26, 27, 28, 29, 30 and 33; Table 1) which differ in screening frequency and patient age. A probabilistic uncertainty analysis was performed by Monte Carlo simulation to address parametric uncertainty. Parameter values were drawn from 38 independent input distributions and run through a 14-stage Markov model. Simulations plotted on the cost-effectiveness plane (Fig. 2a) revealed small differences in mean costs and QALYs among alternatives and considerable variability. The resulting CEAC and the cost-effectiveness acceptability frontier (CEAF) of the baseline analysis are shown in Fig. 2b. The outputs are strongly correlated (CC and QQ correlations between alternatives, see Table 2, baseline case) because all the alternatives are simulated simultaneously with the same set of input parameters and differ only in screening frequency.

The influence of output correlations was assessed to expand this field of research [9–11]. The output set was reshuffled using the Iman–Conover method [12] in three scenarios (denoted 1, 2 and 3) to produce alternative output correlation structures. The resulting output correlations (Table 2) were designed so that different types of output correlations could be addressed separately. The aim was to emphasize that the simulated output values (and thus marginal output distributions) were unchanged and therefore the choice of optimal alternative (NB maximization) would remain as in the baseline model.

In scenario 1 all output correlations were removed to create a situation of apparent independence. The shape of CEAC changed dramatically—acceptability of competing alternatives shifted towards a common value at around 20 % (Fig. 2d). aThe EVPI was affected even more (Fig. 2c)—its value increased by more than one order of magnitude. This change is a direct result of reducing the QQ and CC correlations between alternatives to zero.

Scenario 2 was designed to examine the influence of QC correlations within alternatives. To this end, these

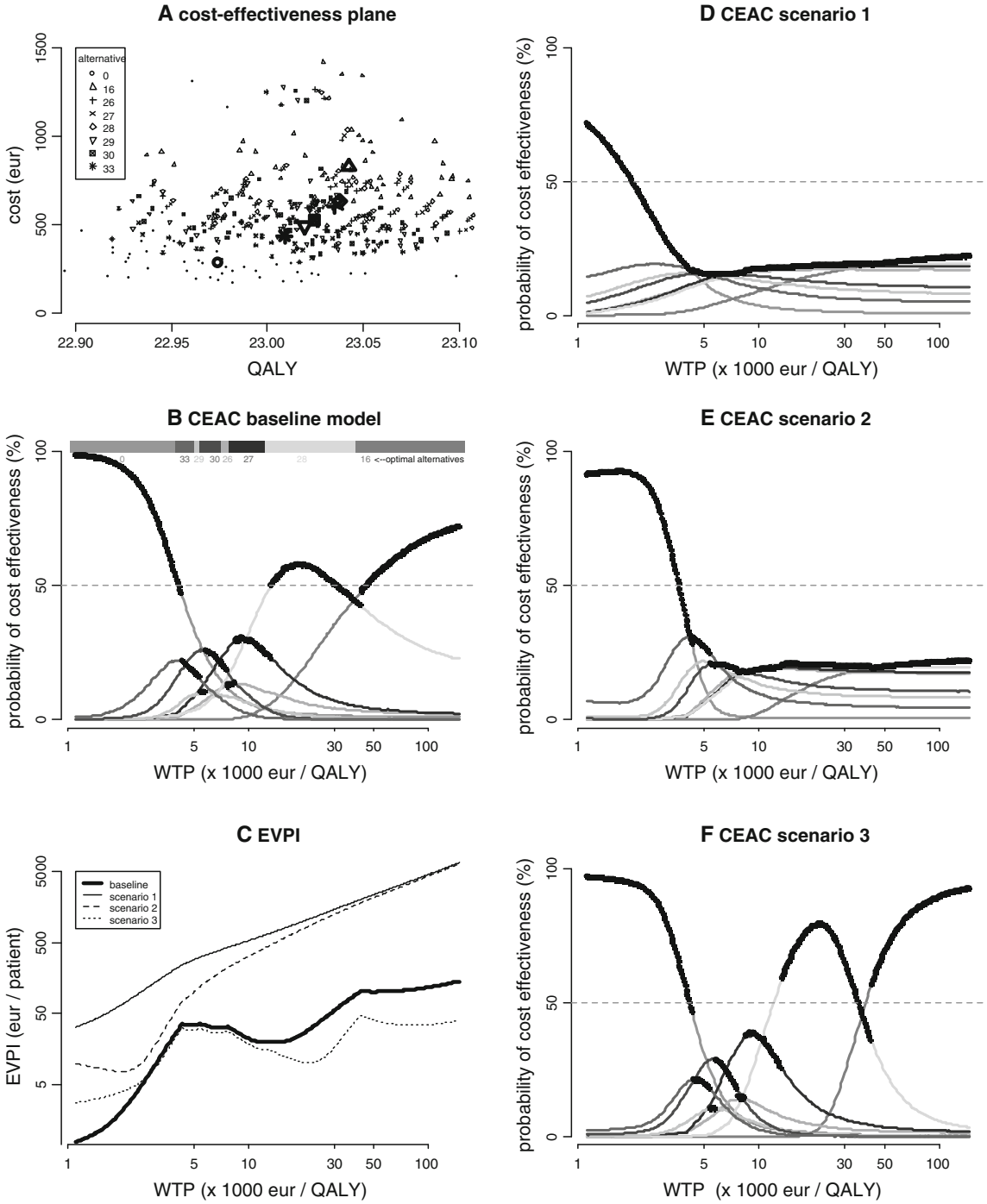

Fig. 2 Results of the baseline model (*left*) and three alternative scenarios (*right*). **a** Cost-effectiveness plane with (non-incremental) cost/QALY pairs of eight alternatives. **b** Acceptability curve for the baseline model (*thick segments of graph* show the acceptability frontier). **c** Expected value of perfect information (*EVPI*) of the baseline model and three alternative scenarios (*logarithmic axes*). **d**–**f** Cost-effectiveness acceptability curves (*CEAC*) for three alternative scenarios: **d** scenario 1—complete independence, **e** scenario 2—increased correlations within alternatives, **f** scenario 3—increased correlations within and between alternatives. *WTP* Willingness-to-pay

correlations were increased to the maximum possible extent (correlation was 0.7 on average due to non-elliptic joint distributions), which resulted in decreased individual NB variances (Eq. 2) but had only a very small effect on the CEAC (Fig. 2e), with the latter remaining very similar to that of the independent scenario 1. Compared to the EVPI of scenario 1, the EVPI of scenario 2 was also reduced only by a fraction, indicating that within-treatment QC correlations are of lesser importance.

Table 2 Correlations and probability of cost-effectiveness for the baseline case and three alternative scenarios[a]

Correlations and probability of cost-effectiveness	Scenario			
	Baseline	Scenario 1	Scenario 2	Scenario 3
Average QQ and CC correlation between alternatives	0.9	0.0	0.0	0.9
Average QC correlation between alternatives	−0.1	−0.1	−0.1	−0.8
Average QC correlation within alternatives	−0.1	−0.1	0.7	−0.8
Probability of cost-effectiveness for the optimal alternative at 30,000 €/QALY	51	19	20	65

CC Correlations among costs, *QQ* correlations among QALYs

[a] See Fig. 2 for description of the different scenarios

The conclusion drawn from scenario 2 was confirmed in scenario 3, where all QC correlations (within and between alternatives) were increased to either the maximum positive value or the minimum negative value. Both cases produced virtually identical results (only the negative case is shown in Fig. 2f), namely, a slight decrease in decision-making uncertainty, as compared to the baseline scenario. The shape of CEAC was changed in terms of a higher probability and the EVPI was reduced. The change in uncertainty was smaller than in the case of CC and QQ correlations (baseline vs. scenario 1). As expected, none of the scenarios changed the choice of the optimal alternative. The presence of output correlations in general decreased decision-making uncertainty, and the correlation between (as opposed to within) alternatives had the highest impact in the examples. For further information, those interested are directed to a published template [13], where the influence of output dependency structure can be analyzed for their own model.

The four scenarios discussed here clearly show that, although irrelevant to the choice of optimal (NB maximizing) alternative under risk neutrality, output correlations considerably influence the shape of CEAC and to an even larger extent the EVPI. I argue that output correlations are not a special feature of these examples but are to be expected in multiple treatment models [11]. Treatment alternatives will undoubtedly share a set of identical inputs, such as all-cause mortality, disease-related costs, QALY weights, among others, and some related inputs are likely to be correlated between alternatives [5, 14]. These will, propagated through the model, cause output correlation. When model outputs are analyzed, care should be taken to retain their correlation structure, and any methodological work will also benefit from a wider applicability if outputs are not assumed to be independent. A rule has been developed, for example, that when an alternative in a multiple CEAC exceeds 50 % it is optimal over all others [7]. Since this rule was designed based on the assumption of output independence, its usefulness is questionable (acceptability of two options under scenario 3 exceeds 50 %, yet they are not necessarily optimal).

Graphical presentation of results when multiple alternatives are compared

The choice between multiple alternatives under conditions of uncertainty presents an interesting decision-making problem because when the number of alternatives is large, clinicians and policy-makers tend to prefer status quo—"the baseline option" [15]. Within cost-effectiveness decision rules under conditions of risk neutrality, the optimal alternative is based on the expected values of output distributions. Output variance, although irrelevant to the choice of optimal alternative, is relevant to the question of whether more information should be acquired [1]. The ideal graphical presentation would:

- indicate or be sensitive to output correlations,
- allow an estimation of decision-making uncertainty,
- show the precision and magnitude of differences in NB between alternatives, possibly for a range of willingness-to-pay (WTP) values.

Several graphical presentations have been published, and each serves a distinct purpose. Not all of them may be suitable for use with multiple alternatives. Here, I have reviewed the following options:

- Cost-effectiveness plane [2]
- NB plot [16]
- CEAC with cost-effectiveness acceptability frontier (CEAF) [17]
- EVPI plot [4]
- Incremental benefit curve [18]
- Stochastic dominance plot [19, 20]

The strong and weak points of these options are presented in Table 3.

The use of EVPI and CEACs is well suited to problems with multiple alternatives. They summarize the dependency structure of the joint distribution of NBs into a single numerical value (at a given WTP). A very useful feature of EVPI and CEACs is that they are sensitive to output correlations and thus provide an appropriate graphical presentation when multiple alternatives are being compared.

Table 3 Properties of various graphical presentations

Properties	Cost-effectiveness plane	NB density	INB density	Stochastic dominance	CEAC	EVPI	Incremental benefit curve
Useful for many alternatives	−	0	+	0	+	+	0
Indicates/is sensitive to output correlations	−	−	+	−	+	+	+
Indicates decision uncertainty	0	0	0	0	+	+	0
Shows the magnitude of differences in NB between alternatives	−	+	+	+	−	−	+
Requires a WTP value to be specified	No	Yes	Yes	Yes	No	No	Yes

CEAC Cost-effectiveness acceptability curve, *EVPI* expected value of perfect information. *WTP* willingness-to-pay

+, Yes; −, no; 0, to some extent

A CEAC for the baseline screening example is shown in Fig. 2b. Regions of optimality are displayed above the graph and by plotting the CEAF. The optimal interventions in the screening example do not necessarily have the highest probability of cost-effectiveness for two reasons: (1) skewness of NB and (2) the presence of output correlations [11]. This was acute in our model where the probability of cost-effectiveness for the optimal alternative (NB maximization) at 5,000 €/QALY was as low as 10 % (see Fig. 2b). In our model, there were regions of CEAC where competing alternatives had significantly higher acceptability than the optimal one, indicating the important role of CEAF in the proper interpretation of the CEAC.

A known shortcoming of the CEAC is that it conflates the precision of the NB estimates with their magnitude [20]. It has been argued that this is not specific to CEACs but rather more an educational issue of confusing statistical significance and clinical importance [21]. The solution to this problem is very simple: the magnitude of differences in NBs and their uncertainty can be easily identified by plotting INB densities (Fig. 1b) next to a CEAC.

Discussion

The question commonly asked within a resource allocation decision is "is a new treatment cost-effective relative to a chosen comparator." More often, the more appropriate question to ask would be "which treatment is most cost-effective for a given disease," as this approach avoids the problem of finding the appropriate comparator and enables a more comprehensive analysis. A probabilistic model with more than two alternatives will result in QALYs and costs for each alternative, which are likely to be correlated. In my analysis, I have shown that output correlations, particularly between alternatives, had a significant influence on decision-making uncertainty in the screening model and should be accounted for. Care should be taken to analyze all alternatives in the model using identical realizations from distributions of input parameters and to keep the joint distribution unchanged when outputs are analyzed.

The value of perfect information is strongly influenced by output correlations. The example used here showed that there was a one order of magnitude increase in EVPI when output correlations were ignored. Intuitively, a reduction in EVPI for correlated outputs is expected since correlation means that each output provides some information on the other outputs. This implies that a correct estimate of decision-making uncertainty will only be relevant when all alternatives relevant to the appraisal in question are analyzed concurrently within a comprehensive decision analysis.

A proper presentation of the analysis cannot be achieved by presenting alternatives individually; rather, their joint distribution (and the correlation structure it contains) needs to be summarized. The CEAC is appropriate for this purpose but needs to include the CEAF to identify the optimal (NB maximization) alternative. The CEAC is likely to be highly dependent on the choice of alternatives when these have been selected from a very large set [9]. For a large decision set, the probability of cost-effectiveness for the optimal alternative could be small enough to raise doubts concerning the optimal choice even in the mind of a risk-neutral, expected NB-maximizing decision-maker. Output correlations can further complicate matters (decreasing or increasing the probability of cost-effectiveness), making it difficult to draw clear conclusions. One example of this is a review of seven anti-depressants where a low probability of cost-effectiveness (<30 %) of the optimal alternative was quoted as a major limitation and "although indicative, is rather low and inadequate to lead to a safe conclusion on … superiority" [22]. An additional indication of the size of differences between alternatives (especially relative to the optimal one) could lead to a different conclusion. This could easily have been done by plotting INB densities.

Acceptability curves have been shown here to be a valuable tool for presenting results when multiple alternatives are being analyzed since these curves identify the

optimal alternative (by use of CEAF), indicate the joint decision-making uncertainty (probability of cost-effectiveness) and are sensitive to output correlations (as they should be). However, they lack the ability to indicate the magnitude of differences in NB. Since WTP thresholds are made explicitly, I propose that a NB density plot (incremental to the optimal alternative) be plotted to augment the CEAC where it is lacking in order to present the differences in the NB of the alternatives. The usefulness of presenting NB density plots along with acceptability curves is particularly evident in cases with multiple alternatives, where the low acceptability of the optimal alternative and the presence of output correlations blur the choice of the optimal decision.

Further implications of output correlations are expected in decision-making with risk aversion because the latter introduces non-linearity in the utility equation (with respect to QALYs and costs). One option of specifying preferences over NB under risk aversion is the use of Eq. 3, where r represents the level of risk aversion [23]. Since the equation for utility also involves a variance term, it will depend on QC correlations within alternatives (see Eq. 2).

$$\text{Utility}(NB_I) = E(NB_I) - (r/2)\text{Var}(NB_I) \quad (3)$$

Under the condition of risk aversion, output correlations have a direct influence not only on decision-making uncertainty, but also on the choice of the optimal alternative. This opens up a new field of further research on the importance of output correlations.

References

1. Claxton, K.: The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J. Health. Econ. **18**(3), 341–364 (1999)
2. Briggs AH, Gray AM: Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess. **3**(2) (1999)
3. Glick, H.A.: Sample size and power for cost-effectiveness analysis (part 1). Pharmacoeconomics. effectiveness analysis (part 1). Pharmacoeconomics. **29**, 189–198 (2011)
4. National Institute for Health and Clinical Excellence: Guide to the methods of technology appraisal, June 2008. Available at: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf. Accessed 19, Sept 2012
5. Cooper, N.J., Peters, J., Lai, M.C.W., et al.: How valuable are multiple treatment comparison methods in evidence-based healthcare evaluation? Value Health **14**, 371–380 (2011)
6. Saramago, P., Manca, A., Sutton, A.J.: Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. Value Health **15**, 639–649 (2012)
7. Jakubczyk, M., Kamiński, B.: Cost-effectiveness acceptability curves—caveats quantified. Health Econ. **19**, 955–963 (2010)
8. Rojnik, K., Naveršnik, K., Mateović-Rojnik, T., et al.: Probabilistic cost-effectiveness modeling of different breast cancer screening policies in Slovenia. Value Health **11**(2), 139–148 (2008)
9. Barton, P.: What happens to value of information measures as the number of decision options increases? Health Econ. **20**(7), 853–863 (2010)
10. Koffijberg, H., de Wit, G.A., Feenstra, T.L.: Communicating uncertainty in economic evaluations: verifying optimal strategies. Med. Decis. Making **32**, 477 (2012)
11. Sadatsafavi, M., Najafzadeh, M., Marra, C.: Acceptability curves could be misleading when correlated strategies are compared. Med. Decis. Making **28**, 306–307 (2008)
12. Iman, R.L., Conover, W.J.: A distribution free approach to inducing rank correlation among input variables. Commun. Stat. Simulat. **B11**(3), 311–334 (1982)
13. Barton, G.R., Briggs, A.H., Fenwick, E.: Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). Value Health **11**(5), 886–897 (2008)
14. Naveršnik, K., Rojnik, K.: Handling input correlations in pharmacoeconomic models. Value Health **15**, 540–549 (2012)
15. Redelmeier, D.A., Shafir, E.: Medical decision making in situations that offer multiple alternatives. JAMA **273**(4), 302–305 (1996)
16. Stinnett, A.A., Mullahy, J.: Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med. Decis. Making. **18**[Suppl 2] S68–S80 (1998)
17. Fenwick, E., Claxton, K., Schulper, M.: Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ. **10**, 779–787 (2001)
18. Bala, M.V., Zarkin, G.A., Mauskopf, J.: Presenting results of probabilistic sensitivity analysis: the incremental benefit curve. Health Econ. **17**, 435–440 (2008)
19. Leshno, M., Levy, H.: Stochastic dominance and medical decision making. Health Care Manag. Sci. **7**, 207–215 (2004)
20. Koerkamp, B.G., Hunink, M.G.M., Stijnen, T., et al.: Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. Med. Decis. Making **27**(2), 101–111 (2007)
21. Fenwick, E., Briggs, A.H.: Cost-effectiveness acceptability curves in the dock: case not proven? Med. Decis. Making **27**(2), 93–95 (2007)
22. British Psychological Society: Schizophrenia: core interventions in the treatment and management of schizophrenia in primary and secondary care (update). National Collaborating Centre for Mental Health (UK): Leicester (UK). (2009)
23. Zivin, J.G.: Cost-effectiveness analysis with risk aversion. Health Econ. **10**, 499–508 (2001)