Uncertainty around the Incremental Cost Utility Ratio Accounting for Mapping Prediction: Application to Hepatitis C *

Carole Siani [†a] **Christian de Peretti** [b] **Christel Castelli** [c]
Gérard Duru [d] **Jean-Pierre Daurès** [c,e]

December 9, 2011

[a] Research Laboratory in Knowledge Engineering (ERIC, EA3083), Institut des Sciences Pharmaceutiques et Biologiques (ISPB), University Claude Bernard Lyon 1, University of Lyon, France.

[b] Laboratory of Actuarial and Financial Sciences (SAF, EA2429), Institute of Financial and Insurance Sciences (ISFA School), University Claude Bernard Lyon 1, University of Lyon, France.

[c] Département Biostatistique Epidemiologie clinique Santé Publique et Information Médicale (BESPIM), Centre Hospitalier Universitaire, Nîmes, France.

[d] Cyklad, France.

[e] Laboratoire Epidémiologie et Biostatistique, Institut Universitaire de Recherche Clinique, Montpellier, France.

Abstract

In cost-utility analysis, one or more medical treatment(s) are compared with a standard treatment on the two-fold basis of their cost and their utility. However, the health utility values are rarely available and are generally predicted by extrapolating (using a "mapping" function) a known clinical questionnaire. In the literature this mapping is not accounted for when uncertainty is handled, leading to wrong decision-making with serious consequence on the patient's health. The purpose of this paper is to build a confidence region around the incremental cost-utility ratio (or the incremental cost per QALY ratio), accounting for the uncertainty coming from the questionnaire extrapolation. Analytic and nonparametric bootstrap procedures are then proposed.

J.E.L. Classification: C13, C15, [C44]; I19.

Keywords: Mapping, Incremental Cost-Utility Ratio, uncertainty, bootstrap, EQ-5D, Hepatitis C.

*The authors would like to thank IReSP and DGS, DREES-MiRe, InVS, HAS, CANAM, AFSSAPS, Inserm for funding this project.

†Correspondence to: Carole Siani, Bâtiment d'odontologie, 11 Rue Guillaume Paradin, 69372 Lyon cedex 08, France. Tel: +33 (0)4 78 77 10 50. Fax: +33 (0)4 72 43 10 44 . Email: carole.siani@univ-lyon1.fr. Web: recherche.univ-lyon2.fr/eric/82-Carole-Siani.html.

1 Introduction

In cost-effectiveness analysis (CEA), one or more medical treatment(s) are compared with a standard treatment on the two-fold basis of their cost and their medical effectiveness by decision-makers. Only recently, the utility was taken into account instead of the sole effectiveness. Since the utility measure is rarely available, it is often extrapolated from a technical questionnaire. In practice, the utility is measured on a sample in a reference study, as well as a clinical questionnaire is collected. A mapping function is then estimated between the utility measure and the questionnaire on the reference sample. Then, for a new sample, only the clinical questionnaire is collected, and the utility measure is predicted for the patients by extrapolating the questionnaire using the estimated mapping. For instance, in Rheumatoid Arthritis, the EuroQol-5D measure is extrapolated from the Health Assessment Questionnaire (HAQ) or the Disease Activity Score (DAS28) (see Ariza-Ariza et al. [2006], Nord et al. [1992], Longworth et al. [2005]). For other examples, see among others Torrance [1976], Krabbe et al. [1997], Dolan and Sutton [1997], O'Leary et al. [1995], Torrance et al. [1996].

In Tsuchiya et al. [2002], the authors convert Asthma Quality of Life Questionnaire (AQLQ) into EQ5D indices. They propose a simple transformation (linear), multi-linear regressions over the various domains or items. The Generalized Linear Model (GLM) can also be cited to transform the dependent variable into an *s*-shaped. The dependent variable is transformed into an s-shaped non-linear variable approaches 1, but does not reach it. The logit transformation can also be applied. The obvious shortcoming of this in the context of Tsuchiya et al. [2002] is that there are many responses with observed EQ-5D index of 1.00, and the transformation will imply dropping these observations (because the transformed values approach infinity). Tsuchiya et al. [2002] adapted this procedure by standardizing the raw EQ-5D scores to the range [0,1], based on an artificial range, say, [-0.5, +1.1], and then transforming this. In their paper, Tsuchiya et al. [2002], given the additional complication of their model, the arbitrary nature of the standardization and the transformation, and the fact that the maximum predicted EQ-5D scores of the simple linear models hardly exceed 1.00, the associated benefits of GLM do not seem to outweigh its costs. Therefore Tsuchiya et al. [2002] decided not to use GLM. In Stevens et al. [2006], Shmueli [2007], the authors propose mapping between Visual Analogue Scale and Standard Gamble data using Power models. The authors assume that the transformation function does not vary across individuals and that there is independence across observations. Linear, quadratic, cubic and power functions are estimated. The quadratic and cubic models use the value form and are constrained to pass through 0 and 1.

In their paper, Stevens et al. [2006], Shmueli [2007], Salomon and Murray [2004], Longworth et al. [2005] report the prediction of their models for the mean utility. However, no confidence interval of this prediction is given. The absolute error of the model is provided, but it not the estimate error of the mean estimator. In their paper, Rivero-Arias et al. [2009] also report the mean utility on the out-of-sample. They provide a confidence region for the mean utility. They argue that "In terms of predicting uncertainty around EQ-5D mean estimates, their model estimated tighter 95% CIs when compared to the actual 95% CIs." However, these confidence regions are under-estimated (this will be shown theoretically in subsection 3.3 and empirically in subsection 4.2).

Consequently, in this paper, we propose a method to handle the statistical uncertainty around the cost and utility simultaneously. More precisely, the purpose of this

paper is to build a confidence region around the incremental cost-utility ratio (ICUR), accounting for the uncertainty coming from the questionnaire extrapolation, commonly called "mapping". We show that if the extrapolated utility values are used to compute a confidence region as if they were the observed values, this procedure dramatically decreases the confidence region so that the conclusion is not reliable since the uncertainty is largely underestimated. This spurious decrease in uncertainty is not accounted for in the studies of the literature as well as in the CEAs conducted in the pharmaceutical industry. Furthermore, decision-making coming from these results, can be misleading.

In this paper, analytic and nonparametric bootstrap procedures are proposed to build the confidence region for the ICUR accounting for the questionnaire extrapolation procedure. The performance of the methods is assessed using Monte Carlo experiments for various sample sizes and for various models (linear, random coefficients, non-Gaussian, non-linear). Then, the methods are made valid from real data issued from a randomized controlled clinical trial dealing with Hepatitis C, whose objective is to measure the impact of the therapeutic education in the undertaking. And finally, an out-of-sample validation is carried out to check the performance of the various methods on real data.

The remainder of this paper is organized as follows. section 3 proposes several methods to handle uncertainty around the ICUR accounting for mapping extrapolation. section 4 provides the Monte Carlo results about the performance of the methods. section 5 provides an application to Hepatitis C as well as a cross sample validation. Finally, section 6 concludes.

2 Background: the incremental cost-utility ratio

2.1 Definition and estimation

In economic evaluations, an ICUR statistic in which a new therapy ($T = 1$) is compared with a standard therapy ($T = 0$) is defined by:

$$ICUR = \frac{\mu_{C_1} - \mu_{C_0}}{\mu_{U_1} - \mu_{U_0}}, \tag{1}$$

where μ is the true mean value of (subscripts) costs (C) and utility (U) for treatments number 1 and number 0. Since the true means corresponding to the theoretical population are not known, the ICUR can be estimated as follows, on the basis of data collected from two groups of patients, each undergoing one of the forms of therapy (group number 1, consisting of n_1 individuals, underwent treatment ($T = 1$) and group number 0, consisting of n_0 individuals [1], underwent treatment ($T = 0$)):

$$\widehat{ICUR} = \frac{\overline{C_1} - \overline{C_0}}{\overline{U_1} - \overline{U_0}} = \frac{\Delta\bar{C}}{\Delta\bar{U}}, \tag{2}$$

where $\overline{C_1}$, $\overline{C_0}$ are the sample mean of the costs and $\overline{U_1}$, $\overline{U_0}$ are the sample mean of utility in the two treatment arms.

[1] n_1 is generally different from n_0.

2.2 Assumptions and statistical properties

Assumption 1: Utility distribution

It is assumed that the utilities of treatment $T = 0, 1$ follow independent random variables with $[0, 1]$ support and with distribution D_{U_T} as follows:

$$U_{T,i} \sim D_{U_T}(\mu_{U_T}, \sigma^2_{U_T}), \tag{3}$$

where i denotes the individual.

Since the support is finite, all the moments of the distribution are finite.

Assumption 2: Existence of the mapping

It is assumed that the utility measure can be explained by some variables X.

$$U_{T,i} = f(X_{T,i}, \varepsilon_{T,i}; \beta_T), \tag{4}$$
$$\varepsilon_{T,i} \sim i.i.d.N(0, \sigma^2_{\varepsilon_T}), \tag{5}$$

where f is a function that depends on a parameter vector β_T, and $\varepsilon_{T,i}$ is independent of X_T. It should be noted that β_T is specific to the treatment.

Let us consider a *reference sample* of size n_T^{ref} where the utility measure is known and the *current sample* where the utility is unknown. The aim of mapping is to assess the current sample mean utility:

$$\mu_{U_T} = E(U_{T,i}) = E(f(X_{T,i}, \varepsilon_{T,i}; \beta_T)).$$

Assumption 3: Cost distribution

It is assumed that the costs of treatment $T = 0, 1$ follow independent random variables with $(0, +\infty)$ support and with distribution D_{C_T} as follows:

$$C_{T,i} \sim D_{C_T}(\mu_{C_T}, \sigma^2_{C_T}), \tag{6}$$

where $\sigma^2_{C_T} < \infty$. i denotes the individual.

Assumption 4: Cost-Utility dependence

It is assumed that the utility measure and the cost are correlated for individual i and treatment T:

$$Cov(U_{T,i}, C_{T,i}) = \sigma_{UC_T} \tag{7}$$

Although the data in question do not follow normal distributions, we can generally apply the Central Limit Theorem (CLT), partly thanks to the fact that each sequence of pairs of random variables $(C_{1,i}, U_{1,i})_{i=1,\cdots,n_1}$, $(C_{0,i}, U_{0,i})_{i=1,\cdots,n_0}$ are independent and identically distributed (because the data were obtained in a randomized trial). Therefore, $\overline{C_1}$, $\overline{C_0}$, $\overline{U_1}$ and $\overline{U_0}$ are asymptotically normally distributed, and the same applies for $\Delta\bar{C}$ and $\Delta\bar{E}$ as the difference between normally distributed variables. [2]

[2]The estimated ICUR does not necessarily have a defined mean or a defined variance, mainly owing to the fact that the denominator of the ratio can be statistically close to zero. In this case, the estimated ICUR will be very large (so that it is statistically close to infinite) or indeterminate, depending on whether $\Delta\bar{C}$ is also statistically close to zero, and the distribution of the estimated ICUR will be close to a Cauchy distribution (whose moments are infinite).

2.3 Linear approximation of the mapping

2.3.1 Utility modeling and estimate

If a first order approximation of the function f in Equation 4 was calculated, it can be assumed that the utility measure is approximated by the following equation:

$$U_{T,i} = X_{T,i}\beta_T + \varepsilon_{T,i}, \tag{8}$$

where the constant term is assumed to be contained in X_T. Thus,

$$\mu_{U_T} = E(X_{T,i})\beta_T.$$

The estimator usually used is:

$$\hat{\mu}_{U_T} = \frac{1}{n_T}\sum_{i=1}^{n_T}\hat{U}_{T,i} = \frac{1}{n_T}\sum_{i=1}^{n_T}X_{T,i}\hat{\beta}_T^{\text{ref}} = \bar{X}_T\hat{\beta}_T^{\text{ref}} \tag{9}$$

where $\hat{U}_i$ are the predicted values for the utility measure, $\hat{\beta}_T^{\text{ref}} = (X_T^{\text{ref}'}X_T^{\text{ref}})^{-1}X_T^{\text{ref}'}U_T^{\text{ref}}$ the estimate of β_T in the reference sample. It should be noted that:

$$\hat{\mu}_{U_T} = \frac{1}{n_T}\iota'_{n_T}X_T\beta_T + \frac{1}{n_T}\iota'_{n_T}X_T(X_T^{\text{ref}'}X_T^{\text{ref}})^{-1}X_T^{\text{ref}'}\varepsilon_T^{\text{ref}}, \tag{10}$$

where $\iota_{n_T} = (1,\ldots,1)'$. Then, we have the following properties: $E(\hat{\mu}_{U_T}) = E(X_T)\beta_T$. The estimator is unbiased, and can be used to assess the mean utility. However, to take a decision, the uncertainty has also to be accounted for. The problem, which will be handle in section 3, is to estimate the variance of $\hat{\mu}_{U_T}$.

2.3.2 Cost estimate

The mean cost can be estimated as follows:

$$\bar{C}_T = \frac{1}{n_T}\sum_{i=1}^{n_T}C_{T,i}. \tag{11}$$

Under Assumption 4, and applying the Central Limit Theorem (CLT) as $n_T^{\text{ref}} \longrightarrow \infty$, we obtain:

$$\bar{C}_T \sim N\left(\mu_{C_T}, \frac{1}{n_T}\sigma^2_{C_T}\right). \tag{12}$$

The variance of $C_{T,i}$ can be estimated as follows:

$$\hat{\sigma}^2_{C_T} = \frac{1}{n_T - 1}\sum_{i=1}^{n_T}\left(C_{T,i} - \bar{C}_T\right). \tag{13}$$

2.3.3 Utility-Cost dependence modeling and estimate

$$\sigma_{UC_T} = \text{Cov}(X_{T,i,1}, C_{T,i})\beta_{T,1} + \cdots + \text{Cov}(X_{T,i,K}, C_{T,i})\beta_{T,K} + \text{Cov}(\varepsilon_{T,i}, C_{T,i}), \tag{14}$$

where K is the number of explanatory variables, X_0 corresponds to the constant term. Let us denote $\text{Cov}(X_{T,i,k}, C_{T,i}) = \gamma_{T,k}$, $(\gamma_{T,1},\ldots,\gamma_{T,K}) = \gamma_T$, and $\text{Cov}(\varepsilon_{T,i}, C_{T,i}) = \gamma_{\varepsilon_T}$. It should be noted that $\text{Cov}(\varepsilon_{T,i}, C_{T,i})$ is not necessarily equal to 0, since for an individual

i, if its health status is deteriorated, it will also increase (in general) the corresponding cost. The covariance vector can be estimated as follows:

$$\hat{\gamma}_T = \frac{1}{n_T}\sum_{i=1}^{n_T}(X_{T,i} - \iota_{n_T}\bar{X}_T)'(C_{T,i} - \bar{C}_T), \tag{15}$$

where $\iota_{n_T} = (1, \ldots, 1)$ and $\bar{X}_T = \frac{1}{n_T}\sum_{i=1}^{n_T} X_{T,i}$.

3 Confidence regions for the ICUR accounting for mapping

First, we recall here the Fieller's method which was used in the context of cost-effectiveness analysis, thus with no mapping. This method will be used latter in the context of mapping. Second, we propose three methods for handling mapping.

3.1 Recall of the Fieller's method (case of no mapping)

We briefly recall the general context of Fieller's theorem Fieller [1954] (see also Heitjan [2000]). It is assumed here that X_1 and X_2 are two random normally distributed variables such that:

$$X \sim N(\eta, \Omega) \text{ with } X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}, \tag{16}$$

and it is proposed to determine a $(1-\alpha)$ confidence region for $\frac{\eta_1}{\eta_2}$. For this purpose, we draw up the (statistic) $Z = X_1 - \rho X_2$ and we note that:

$$Z \sim N(0, \omega_1^2 + \rho^2\omega_2^2 - 2\rho\,\omega_{12}) \text{ under the assumption that } \rho \text{ is equal to } \frac{\eta_1}{\eta_2}.$$

Therefore, we have:

$$\frac{Z^2}{\omega_1^2 + \rho^2\omega_2^2 - 2\rho\,\omega_{12}} \sim \chi^2(1), \tag{17}$$

$$\Rightarrow P\left(\frac{(X_1 - \rho X_2)^2}{\omega_1^2 + \rho^2\omega_2^2 - 2\rho\,\omega_{12}} \leq k_{1-\alpha}\right) = 1 - \alpha\ ,$$

where $k_{1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-squared distribution with one degree of freedom. To find the $(1-\alpha)$ confidence region for $\frac{\eta_1}{\eta_2}$, the following inequation must be solved:

$$Q(\rho) \leq 0, \tag{18}$$

where

$$Q(\rho) = x\rho^2 + y\rho + z, \tag{19}$$

with $x = X_2^2 - k_{1-\alpha}\omega_2^2$, $y = 2(k_{1-\alpha}\omega_{12} - X_1X_2)$ and $z = X_1^2 - k_{1-\alpha}\omega_1^2$.

We assume that (C_i^T, E_i^T) are independent random 2-vectors with mean (μ_{C_T}, μ_E^T), variance $\left((\sigma_C^T)^2, (\sigma_E^T)^2\right)$ and covariance σ_{CE}^T for $T = 0, 1$. The variables used in Fieller's

method correspond to the following values:

$$\begin{aligned} X_1 &= \Delta\bar{C}, \\ X_2 &= \Delta\bar{E}, \\ \omega_1^2 &= {\sigma_C^0}^2/n_0 + {\sigma_C^1}^2/n_1, \\ \omega_2^2 &= {\sigma_E^0}^2/n_0 + {\sigma_E^1}^2/n_1, \\ \omega_{12} &= \sigma_{CE}^0/n_0 + \sigma_{CE}^1/n_1. \end{aligned}$$

After solving the inequation 18, $(1-\alpha)$ confidence regions for the ICER can have different forms. Depending on the sign of x, defined in Equation 19, and depending on the sign of the discriminant Δ of the polynomial function Q, the various forms of the confidence region obtained with Fieller's method are shown in Table 1.

Table 1: Form of the confidence region

	$\Delta < 0$	$\Delta = 0$	$\Delta > 0$
$x > 0$ Q convex	impossible case	impossible case	$\left[R^L, R^U\right]$
$x = 0$ Q linear	impossible case	$\mathbb{R}$	$\left(-\infty, R^U\right]$ if $y > 0$ $\left[R^L, +\infty\right)$ if $y < 0$
$x < 0$ Q concave	$\mathbb{R}$	$\mathbb{R}$	$(-\infty, R^U] \cup [R^L, +\infty)$

R^L and R^U are the roots of the polynomial function Q, given by the following formulas:

$$R^L = \frac{X_1X_2 - k_{1-\alpha}\omega_{12} - \sqrt{(k_{1-\alpha}\omega_{12} - X_1X_2)^2 - (X_2^2 - k_{1-\alpha}\omega_2^2)(X_1^2 - k_{1-\alpha}\omega_1^2)}}{X_2^2 - k_{1-\alpha}\omega_2^2}, \quad (20)$$

$$R^U = \frac{X_1X_2 - k_{1-\alpha}\omega_{12} + \sqrt{(k_{1-\alpha}\omega_{12} - X_1X_2)^2 - (X_2^2 - k_{1-\alpha}\omega_2^2)(X_1^2 - k_{1-\alpha}\omega_1^2)}}{X_2^2 - k_{1-\alpha}\omega_2^2}. \quad (21)$$

For a detailed analyze of all the cases, see Siani and de Peretti [2003]. [3]

3.2 "Naive" confidence region

A naive, and wrong, way to compute the variance of $\hat{\mu}_{U_T}$ –that can be found in some papers of the literature– would be:

$$\widehat{V}^{\text{Naive}}\left(\hat{\mu}_{U_T}\right) = \frac{1}{n_T}\left[\frac{1}{n_T - 1}\sum_{i=1}^{n_T}\left(\hat{U}_{T,i} - \hat{\mu}_{U_T}\right)^2\right], \quad (22)$$

[3] Siani and Moatti [2003] analyzed all the method of the literature for calculating a confidence region for the incremental cost effectiveness ration (ICER). They found that the only two methods that are reliable are Fieller's method and the "re-ordered" bootstrap method. Siani et al. [2004], Siani and de Peretti [2010] then focused on the performance of Fieller's and "reordered" bootstrap methods in the problematic cases, frequently occurring in practice, of the difference between average effects of the two treatments approaching statistically zero or of the (mean costs difference, mean effects difference) pair also approaching statistically zero using Monte Carlo simulations. Their Monte Carlo simulations show that the non-reordered bootstrap method performs worse than Fieller's method in these problematic cases.

and the covariance between $\hat{\mu}_{U_T}$ and $\bar{C}_T$:

$$\widehat{Cov}^{\text{Naive}}\left(\hat{\mu}_{U_T}, \bar{C}_T\right) = \frac{1}{n_T}\left[\frac{1}{n_T - 2}\sum_{i=1}^{n_T}\left(\hat{U}_{T,i} - \hat{\mu}_{U_T}\right)\left(C_{T,i} - \bar{C}_T\right)\right]. \tag{23}$$

The standard deviation in Equation 22 seems to be used in Rivero-Arias et al. [2009] to compute the confidence interval for the mean utility. A naive way to provide a $(1-\alpha)$-confidence region for the ICUR is to provide the following moments to the Fieller's method:

$$\begin{aligned}
X_1 &= \Delta\bar{C} = \bar{C}_1 - \bar{C}_0, && (24)\\
X_2 &= \Delta\hat{U} = \hat{\mu}_{U_1} - \hat{\mu}_{U_0}, && (25)\\
\hat{\omega}_1^2 &= \hat{\sigma}_{C_1}^2/n_1 + \hat{\sigma}_{C_0}^2/n_0, && (26)\\
\hat{\omega}_2^2 &= \widehat{V}^{\text{Naive}}\left(\hat{\mu}_{U_1}\right) + \widehat{V}^{\text{Naive}}\left(\hat{\mu}_{U_0}\right), && (27)\\
\hat{\omega}_{12} &= \widehat{Cov}^{\text{Naive}}\left(\hat{\mu}_{U_1}, \bar{C}_1\right) + \widehat{Cov}^{\text{Naive}}\left(\hat{\mu}_{U_0}, \bar{C}_0\right). && (28)
\end{aligned}$$

The confidence region for the ICUR is then given by Table 1, Equation 20, and Equation 21 when using the moments above.

3.3 Analytic confidence region

In the case of a linear approximation, the variance of $\hat{\mu}_{U_T}$ can be estimated as follows:

$$\hat{V}\left(\hat{\mu}_{U_T}\right) = \frac{1}{n_T}\hat{\beta}_T^{\text{ref}\prime}\hat{\Omega}_{X_T}\hat{\beta}_T^{\text{ref}} + \hat{\sigma}^2_{\varepsilon_T^{\text{ref}}}\bar{X}_T(X_T^{\text{ref}\prime}X_T^{\text{ref}})^{-1}\bar{X}_T{}'. \tag{29}$$

See proof in Appendix, subsection A.2. It should be noted that

$$\hat{V}\left(\hat{\mu}_{U_T}\right) = \widehat{V}^{\text{Naive}}\left(\hat{\mu}_{U_T}\right) + \hat{\sigma}^2_{\varepsilon_T^{\text{ref}}}\bar{X}_T(X_T^{\text{ref}\prime}X_T^{\text{ref}})^{-1}\bar{X}_T{}'. \tag{30}$$

In Equation 22, the term $\hat{\sigma}^2_{\varepsilon_T^{\text{ref}}}\bar{X}_T(X_T^{\text{ref}\prime}X_T^{\text{ref}})^{-1}\bar{X}_T{}'$ is missing. The covariance between $\hat{\mu}_{U_T}$ and $\bar{C}_T$ can be estimated as follows:

$$\widehat{Cov}\left(\hat{\mu}_{U_T}, \bar{C}_T\right) = \frac{1}{n_T}\hat{\gamma}_T{}'\hat{\beta}_T^{\text{ref}}. \tag{31}$$

See proof in Appendix, subsection A.3. The confidence region for the ICUR is then given by Table 1, Equation 20, and Equation 21 when using the following moments:

$$\begin{aligned}
X_1 &= \Delta\bar{C} = \bar{C}_1 - \bar{C}_0, && (32)\\
X_2 &= \Delta\hat{U} = \hat{\mu}_{U_1} - \hat{\mu}_{U_0}, && (33)\\
\hat{\omega}_1^2 &= \hat{\sigma}_{C_1}^2/n_1 + \hat{\sigma}_{C_0}^2/n_0, && (34)\\
\hat{\omega}_2^2 &= \widehat{V}\left(\hat{\mu}_{U_1}\right) + \widehat{V}\left(\hat{\mu}_{U_0}\right), && (35)\\
\hat{\omega}_{12} &= \widehat{Cov}\left(\hat{\mu}_{U_1}, \bar{C}_1\right) + \widehat{Cov}\left(\hat{\mu}_{U_0}, \bar{C}_0\right). && (36)
\end{aligned}$$

These analytic confidence region is restricted to a linear framework with Gaussian error terms. It can easily be extended to a nonlinear framework, using an Edgeworth expansion, but it will be an approximation. Consequently, we prefer to propose a bootstrap methodology, which will account for nonlinear specification, which may be used (such as logistic specification for instance). In addition, using nonparametric bootstrap, error terms with non-Gaussian distribution can also be accounted for.

3.4 Nonparametric bootstrap confidence region

In this subsection, we propose a methodology for building a confidence region based on the nonparametric bootstrap technique to compute the moments of the estimators. For a general presentation of the percentile-t method, see Hall [1992], Davidson and MacKinnon [1993], Efron and Tibshirani [1993], Hjorth [1994], and Shao and Tu [1995]. A mapping model is chosen:

$$\begin{aligned} U_{T,i} &= f(X_{T,i}, \varepsilon_{T,i}; \beta_T), && (37)\\ C_{T,i} &= g(U_{T,i}, \nu_{T,i}; \theta_T), && (38) \end{aligned}$$

where X_T is a regressor matrix, and the functions f and g are known and are parametric in the sense that they depend on a parameter vector β_T or θ_T. $\varepsilon_{T,i}$ and $\nu_{T,i}$ are not assumed to be Gaussian. $V(\varepsilon_{T,i}) = \sigma^2_{\varepsilon_T}$, and $V(\nu_{T,i}) = (\sigma_{\nu_T})^2$. The confidence region is built as follows:

1. Equation 37 and Equation 38 are estimated on the reference sample, providing $\hat{\beta}_T^{\text{ref}}$, $\hat{\varepsilon}_T^{\text{ref}}$, $(\hat{\sigma}_{\varepsilon_T}^{\text{ref}})^2$, $\hat{\theta}_T^{\text{ref}}$, $\hat{\nu}_T^{\text{ref}}$, and $(\hat{\sigma}_{\nu_T}^{\text{ref}})^2$.

2. A bootstrap Data Generating (DGP) has to be defined. It may be either parametric or semiparametric, characterized by $\hat{\beta}_T^{\text{ref}}$, $\hat{\theta}_T^{\text{ref}}$, and by any other relevant estimates that may be needed. In a general case, we propose:

$$\begin{aligned} X_{T,i}^{\text{ref},b} &\sim \text{ i.i.d. uniform distribution on } (X_{T,i}^{\text{ref}})_i, && (39)\\ \varepsilon_{T,i}^{\text{ref},b} &\sim \text{ parametric or nonparametric specification based on } \hat{\varepsilon}_T^{\text{ref}}, && (40)\\ U_{T,i}^{\text{ref},b} &= f(X_{T,i}^{\text{ref},b}, \varepsilon_{T,i}^{\text{ref},b}; \hat{\beta}_T^{\text{ref}}), && (41)\\ \nu_{T,i}^{\text{ref},b} &\sim \text{ parametric or nonparametric specification based on } \hat{\nu}_T^{\text{ref}}, && (42)\\ C_{T,i}^{\text{ref},b} &= g(U_{T,i}^{\text{ref},b}, \nu_{T,i}^{\text{ref},b}; \hat{\theta}_T^{\text{ref}}), && (43)\\ X_{T,i}^{b} &\sim \text{ i.i.d. uniform distribution on } (X_{T,i})_i, && (44)\\ \hat{\hat{\beta}}_T^{\text{ref},b} &= \text{ estimate of } \beta \text{ in } U_{T,i}^{\text{ref},b} = f(X_{T,i}^{\text{ref},b}, \varepsilon_i; \beta), && (45)\\ \hat{U}_{T,i}^{b} &= f(X_{T,i}^{b}, 0; \hat{\hat{\beta}}_T^{\text{ref},b}), && (46)\\ & && (47) \end{aligned}$$

   for $T = 1, 0$ and $i = 1, \ldots, n_T^{\text{ref}}$. The distribution of ε_i^b will be discussed later. f and g have to be chosen. For simplicitys sake, a linear model is chosen here:

$$\begin{aligned} U_{T,i}^{b} &= X_{T,i}^{b}\hat{\beta}_T^{\text{ref}} + \varepsilon_{T,i}^{b}, && (48)\\ C_{T,i}^{b} &= U_{T,i}^{b}\hat{\theta}_T^{\text{ref}} + \nu_{T,i}^{b}, && (49) \end{aligned}$$

   where X_T is assumed to contain the constant term, but a more specific nonlinear model can be chosen in practice in accordance with the data.

3. B bootstrap samples are generated:

$$(X_{T,i}^{b})_{i=1}^{n_T}, (U_{T,i}^{b})_{i=1}^{n_T^{\text{ref}}}, (C_{T,i}^{b})_{i=1}^{n_T},$$

   for $T = 1, 0$ and $b = 1, \ldots, B$.

4. For each of these samples, we compute $\hat{\beta}_T^{\text{ref}}$ -denoted $\hat{\beta}_{\text{in}}^{T,b}$-, $\Delta\hat{U}^b = \hat{\mu}_{U^{1,b}_{\text{out}}} - \hat{\mu}_{U^{0,b}_{\text{out}}} = \bar{X}_{\text{out}}^{1,b}\hat{\beta}_{\text{in}}^{1,b} - \bar{X}_{\text{out}}^{0,b}\hat{\beta}_{\text{in}}^{0,b}$ and $\Delta\bar{C}^b = \bar{C}_{\text{out}}^{1,b} - \bar{C}_{\text{out}}^{0,b}$.

5. The variance-covariance matrix of $X_1 = \Delta\bar{C} = \bar{C}_1 - \bar{C}_0$ and $X_2 = \Delta\hat{U} = \hat{\mu}_{U_1} - \hat{\mu}_{U_0}$ is then computed as follows:

$$\hat{\omega}_1^2 = \frac{1}{B}\sum_{b=1}^{B}\left(\Delta\bar{C}^b - \Delta\bar{C}\right)^2, \tag{50}$$

$$\hat{\omega}_2^2 = \frac{1}{B}\sum_{b=1}^{B}\left(\Delta\hat{U}^b - \Delta\hat{U}\right)^2, \tag{51}$$

$$\hat{\omega}_{12} = \frac{1}{B}\sum_{b=1}^{B}\left(\Delta\bar{C}^b - \Delta\bar{C}\right)\left(\Delta\hat{U}^b - \Delta\hat{U}\right). \tag{52}$$

6. The confidence region is obtained by applying Fieller's method to the moments computed using bootstrap techniques.

We consider the following way of generating the bootstrap residuals $\varepsilon_{T,i}^b$ and $\nu_{T,i}^b$ (see Weber Weber [1984]). The $\varepsilon_{T,i}^b$ are generated by independent uniform draws with replacement among the vector with the typical element $\tilde{\varepsilon}_{T,i}$ constructed as follows:

1. Calculate $(P_{X_T})_{i,i}$, $i = 1, \ldots, n_T^{\text{ref}}$, the diagonal elements of the projection matrix on X_T.
2. Calculate $\frac{\hat{\varepsilon}_{T,i}}{\sqrt{1-(P_{X_T})_{i,i}}}$, $\forall i = 1, \ldots, n_T^{\text{ref}}$.
3. Recenter the vector that results.
4. Rescale it so that it has the variance $(\hat{\sigma}_\varepsilon^T)^2$.

This permits to correct the heteroskedasticity in the residuals due to the regressors. The same procedure is applied for $\nu_{T,i}^b$.

4 Performance of the methods: Monte Carlo experiments

Data Generating (DGP) are use to generate simulated data samples. The methods are applied to each simulated sample $j = 1, \ldots, S$, and it is examined if each confidence region j contains or not the true value $ICUR$ of the ratio (which is known, since the DGP is known, conversely to real data). The coverage c of the confidence regions can be estimated as follow:

$$\hat{c} = \frac{1}{S}\sum_{j=1}^{S} I\left(\mu_U \in Interval_j\right). \tag{53}$$

The standard deviation of this Monte Carlo estimate of the coverage is $\sqrt{\frac{1}{S}c(1-c)}$, where c is the true coverage.

In our Monte Carlo experiments, we choose the confidence level $1-\alpha = 0.95$. The number of bootstrap replications is $B = 999$. The number of Monte Carlo replications is $S = 10,000$. If the true coverage $c = 0.95$, the standard deviation of the Monte Carlo estimate of the coverage is 0.002179. At most (where $c = 0.5$) the standard deviation is 0.005. Several values for n_T^{ref} and n_T are chosen. Small values for n_T^{ref} and large values for n_T are first allow to reflect the case where the utility is assessed only on a small subsample, and then extrapolated to the other patients.

4.1 Data Generating Process

A variety of DGP are proposed to check the robustness of the methods: linear, nonlinear, with non-Gaussian error terms.

4.1.1 Linear Case

$$
\begin{aligned}
U_{T,i} &= \beta_0^T + \beta_{T,1} \cdot X_{1i}^T + \beta_2 \cdot X_{2i}^T + \varepsilon_{T,i}, && (54)\\
C_{T,i} &= \beta_{C,0}^T + \beta_{C,1}^T \cdot U_{T,i} + \nu_{T,i}, && (55)\\
X_{1i}^T &\sim i.i.d.U([0,1]), && (56)\\
X_{2i}^T &\sim i.i.d.B(0.5,3), && (57)\\
\varepsilon_{T,i} &\sim i.i.d.N(0,(\sigma_\varepsilon^T)^2), && (58)\\
\nu_{T,i} &\sim i.i.d.N(0,(\sigma_\nu^T)^2). && (59)
\end{aligned}
$$

The parameters values are set to:
$\beta_0^1 = 0.2$, $\beta_1^1 = 0.7$, $\beta_2^1 = 0.3$, $\sigma_{\varepsilon_1} = 0.15$, $\beta_{C,0}^1 = 0.35$, $\beta_{C,1}^1 = 0.5$, $\sigma_\nu^1 = 0.15$.
$\beta_0^0 = 0$, $\beta_1^0 = 0.5$, $\beta_2^0 = 0.1$, $\sigma_{\varepsilon_0} = 0.15$, $\beta_{C,0}^0 = 0.35$, $\beta_{C,1}^0 = 0.5$, $\sigma_\nu^0 = 0.15$.
We have $E(U_i^1) = 1$, $E(C_i^1) = 0.85$, $E(U_i^0) = 0.4$, $E(C_i^0) = 0.55$, $ICUR = 0.5$.

4.1.2 Random Linear Case

The model is the same, but the parameters vary randomly across the Monte Carlo replications:

$$
\begin{aligned}
\beta_0^1 \sim i.i.d.N(0.2,0.2^2) \quad &, \quad \beta_0^0 \sim i.i.d.N(0,0^2), && (60)\\
\beta_1^1 \sim i.i.d.N(0.7,0.7^2) \quad &, \quad \beta_1^0 \sim i.i.d.N(0.5,0.5^2), && (61)\\
\beta_2^1 \sim i.i.d.N(0.3,0.3^2) \quad &, \quad \beta_2^0 \sim i.i.d.N(0.1,0.1^2), && (62)\\
\sigma_{\varepsilon_1} \sim i.i.d.U([0.15\cdot 0.5, 0.15\cdot 1.5]) \quad &, \quad \sigma_{\varepsilon_0} \sim i.i.d.U([0.15\cdot 0.5, 0.15\cdot 1.5]), && (63)\\
\beta_{C,0}^1 \sim i.i.d.N(0.35,0.35^2) \quad &, \quad \beta_{C,0}^0 \sim i.i.d.N(0.35,0.35^2), && (64)\\
\beta_{C,1}^1 \sim i.i.d.N(0.5,0.5^2) \quad &, \quad \beta_{C,1}^0 \sim i.i.d.N(0.5,0.5^2), && (65)\\
\sigma_\nu^1 \sim i.i.d.U([0.15\cdot 0.5, 0.15\cdot 1.5]) \quad &, \quad \sigma_\nu^0 \sim i.i.d.U([0.15\cdot 0.5, 0.15\cdot 1.5]). && (66)
\end{aligned}
$$

We have $E(U_{T,i}|\beta_T) = (1, 0.5, 1.5)\beta_T$, $E(C_{T,i}|\beta_T, \theta_T) = (1, E(U_{T,i}|\beta_T))\,\theta_T$.

4.1.3 Non-Gaussian Random Linear Case

The model is the same as in Equation 54–Equation 57, but the error terms follow the uniform distribution:

$$
\begin{aligned}
\varepsilon_{T,i} &\sim i.i.d.U([-2,2]) * \sigma_\varepsilon^T, && (67)\\
\nu_{T,i} &\sim i.i.d.U([-2,2]) * \sigma_\nu^T. && (68)
\end{aligned}
$$

The parameters follow the same distributions as in Equation 60–Equation 66. Then, we also have $E(U_{T,i}|\beta_T) = (1, 0.5, 1.5)\beta_T$, $E(C_{T,i}|\beta_T, \theta_T) = (1, E(U_{T,i}|\beta_T))\,\theta_T$.

4.1.4 Nonlinear Case

$$U_{T,i} = F\left[\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \varepsilon_i\right] \quad (69)$$
$$C_{T,i} = \left(\beta_{C,0}^T + U_{T,i}\beta_{C,1}^T\right)\chi^2(1) \quad (70)$$

F is the cumulative distribution function of a normal variable. The parameters values are set to:
$\beta_0^1 = -0.5$, $\beta_1^1 = 0.7$, $\beta_2^1 = 0.3$, $\sigma_{\varepsilon_1} = 0.3$, $\beta_{C,0}^1 = 0.25$, $\beta_{C,1}^1 = 0.5$.
$\beta_0^0 = -0.7$, $\beta_1^0 = 0.5$, $\beta_2^0 = 0.1$, $\sigma_{\varepsilon_0} = 0.3$, $\beta_{C,0}^0 = 0.25$, $\beta_{C,1}^0 = 0.5$.
We have $E(U_i^1) = 0.61791142$, $E(C_i^1) = 0.55895571$, $E(U_i^0) = 0.38208858$, $E(C_i^0) = 0.44104429$, $ICUR = 0.5$.

4.2 Performance

The coverage of the various confidence regions, computed using Monte Carlo experiments for various samples sizes, are presented in Table 2. The sizes are chosen to correspond to a mapping assessment on a small subsample, and then an extrapolation of the utility values to the remaining sample of sizes that can be encountered in practice. The results show that the coverage of the "naive" confidence region is small (down to 60% depending on the sample sizes) with respect to the confidence level (95%), whereas both the analytic and bootstrap confidence regions perform correctly.

5 Application to Hepatitis C

5.1 Data description

The data are issued from a randomized controlled clinical trial dealing with hepatitis C treatment, whose objective is to measure the impact of the therapeutic education in the undertaking. The compared strategies are thus: treatment alone versus treatment added to a consultation of education by a male nurse. Data of direct and indirect costs are available. The utility values are assessed with the EuroQol EQ-5D questionnaire validated in France The functional questionnaire of Nottingham Health Profiles (NHP) is available. The data are collected at 0 weeks (at the inclusion), at 4, 8, 12, 24, 36 and 48 weeks. For some patients, the treatment duration is 24 weeks, and for the others the duration is 48 weeks. The data are also collected 36 weeks and 48 weeks after the end of the treatment is the treatment duration is 24 weeks, and 60 weeks and 72 weeks after the end of the treatment is the treatment duration is 48 weeks.

Because of the missing data, the global scores of utility, NHP and cost are computed as follows: an average is computed on the non-missing observation, weighted by the duration between two observations. Table 3 presents some descriptive statistics.

5.2 Mapping

A linear regression is estimated to explain the utility by the NHP. The results are presented in Table 4.

Table 2: Coverage and mean length of the 95% confidence intervals

		Coverage			Mean angle		
n_T^{ref}	n_T	Naive	Analytic	Bootstrap	Naive	Analytic	Bootstrap
Linear Data Generating Process							
40	400	0.9239	0.9818	0.9818	0.0929	0.1309	0.1275
30	600	0.7892	0.9620	0.9567	0.0761	0.1322	0.1273
20	800	0.6659	0.9598	0.9471	0.0659	0.1501	0.1403
Random Linear Data Generating Process							
80	40	0.9593	0.9679	0.9685	0.4240	0.4538	0.4518
100	100	0.9417	0.9616	0.9616	0.2848	0.3104	0.3102
40	400	0.8348	0.9667	0.9605	0.1324	0.2337	0.2265
30	600	0.7495	0.9620	0.9523	0.1101	0.2520	0.2406
20	800	0.6349	0.9540	0.9437	0.0826	0.2519	0.2319
Non-Gaussian Random Linear DGP							
40	400	0.8000	0.9538	0.9495	0.1457	0.2666	0.2591
30	600	0.7057	0.9577	0.9498	0.1010	0.2624	0.2511
20	800	0.6112	0.9461	0.9381	0.0872	0.3170	0.2886
Nonlinear Data Generating Process							
40	400	0.9512	0.9580	0.9321	0.7086	0.7383	0.7164
30	600	0.9382	0.9559	0.9171	0.5827	0.6292	0.5989
20	800	0.9230	0.9483	0.9207	0.5119	0.5967	0.5608

n_T^{ref} is the *in sample* size used to assess the mapping. n_T is the *out of sample* size where the utility values are predicted.

The NHP has a good power of explanation: the R^2 is about 0.4 for the standard treatment, and about 0.6 for the alternative treatment. The is can be used to do mapping to extrapolate utility values.

5.3 Cross validation

A sub-sample of various sizes is drawn from the real dataset with replacement. More precisely, the triplet (X_i, U_i, C_i) is drawn, to keep the correlation structure. The remaining sample is also drawn from the real dataset with replacement. Then, the mapping is applied from the sub-sample to the remaining sample. The procedure is repeated 10,000 times to get the coverage at the 95% confidence level of the confidence regions. The mean angle is also provided. The number of bootstrap replications is $B = 999$. The performance is provided in Table 5.

Again, the performance shows that the analytic and the bootstrap methods perform better than the naive method. The analytic method performs a little bit better than the bootstrap method.

Table 3: Descriptive statistics

Treatment	Sample size	Missing data
Standard	57	15
Alternative	63	9

Average variable per week	Treatment	Mean	Standard deviation	Min.	Max.
EuroQol EQ-5D	Standard	0.70	0.15	0.31	0.92
	Alternative	0.76	0.14	0.38	1.00
NHP	Standard	193.07	115.53	6.97	427.42
	Alternative	176.64	114.84	0.00	401.98
Cost	Standard	172.40	53.68	77.34	369.16
	Alternative	165.70	45.56	72.54	399.26

6 Conclusion

In this paper, the case where the utility is measured on a subsample of patients is dealt woth. Thus, the link between some explanatory variables and the utility (often called mapping) can be estimated in the aim to calculate the mean utility on the entire sample. But the variability of this link can also be estimated and can be used to handle the uncertainty around the ICUR (to build a the confidence region). However, in many studies, only the explanatory variables are available, such as technical or medical questionnaires, and the authors predict the utility using mapping techniques and value for the mapping parameters given by some experts in reference articles or reports. These values allow to compute the mean utility, but it is impossible to compute any confidence region around this ICUR, and it is impossible to take any decision-making on the basis of this kind of study. In this paper, we propose two methods –one is analytic, the other one is bootstrap– accounting for the mapping extrapolation and providing accurate confidence regions, allowing a reliable decision-making. A cross validation on data are issued from a randomized controlled clinical trial dealing with hepatitis C confirms these results.

Appendix

Table 4: Utility of each treatment with respect to the NHP

Alternative treatment

Variable	Estimate	Standard Error	t-value	Prob $>\|t\|$	Standardized Estimate	Cor with Dep Var
Constant	536.5009	46.5257	11.5312	0.000***	—	—
NHP	-468.5412	63.4275	-7.3870	0.000***	-0.6464	-0.6464

Dependent variable: EQ-5D; Valid cases: 78; R^2=0.418; $\bar{R}^2$=0.410; F(1,76)=54.569; Probability of F=0.000***

Standard treatment

Variable	Estimate	Standard Error	t-value	Prob $>\|t\|$	Standardized Estimate	Cor with Dep Var
Constant	572.7864	33.3699	17.1647	0.000***	—	—
NHP	-528.0491	43.2324	-12.2141	0.000***	-0.7737	-0.7737

Dependent variable: EQ-5D; Valid cases: 102 R^2=0.599; $\bar{R}^2$=0.595; F(1,76)=149.186; Probability of F=0.000***

*: significant at 5%; **: significant at 2%; ***: significant at 1%.

Table 5: Coverage and mean angle of the 95% confidence intervals

Sample Sizes		Coverage			Mean angle		
$n_1 = n_0$	$m^1 = m^0$	Naive	Analytic	Bootstrap	Naive	Analytic	Bootstrap
40	200	0.9112	0.9437	0.9302	3.4089	4.6246	4.5043
30	300	0.8917	0.9461	0.9336	3.1969	4.5627	4.3657
20	400	0.8416	0.9437	0.9243	3.0314	4.5871	4.2805

Mean n_T^{ref} is the *in sample* size (in average) in proportion of the original sample size used to assess the mapping. Mean n_T is the *out of sample* (in average) in proportion of the original sample size size where the utility values are predicted.

A Proof of analytic confidence interval

Assume the following linear model:

$$U_i = X_i\beta + \varepsilon_i.$$

Let the in-sample be denoted: $i = 1, \ldots, n$, and the out-of-sample be denoted: $i = n+1, \ldots, n+m$.

The aim is to assess the out-of-sample mean utility:

$$\mu_{U_{\text{out}}} = E(U_{\text{out},i}) = E(X_{\text{out},i})\beta.$$

The estimator of $\mu_{U_{\text{out}}}$ is:

$$\hat{\mu}_{U_{\text{out}}} = \bar{X}_{\text{out}}\hat{\beta}_{\text{in}} = \frac{1}{m}\iota'_{m^T} X_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}U_{\text{in}},$$

where $\iota_{m^T} = (1, \ldots, 1)'$. It can be noted that:

$$\hat{\mu}_{U_{\text{out}}} = \frac{1}{m}\iota_{m^T} X_{\text{out}}\beta + \frac{1}{m}\iota_{m^T} X_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}. \tag{71}$$

Then, the estimator has the following properties.

A.1 Bias

$E(\hat{\mu}_{U_{\text{out}}}) = E(X_{\text{out},i})\beta$, the estimator is unbiased.

A.2 Variance

$$V(\hat{\mu}_{U_{\text{out}}}) = V(\bar{X}_{\text{out}}\hat{\beta}_{\text{in}}).$$

From Equation 71, we have:

$$\begin{aligned}
V\left(\hat{\mu}_{U_{\text{out}}}\right) &= V\left[\bar{X}_{\text{out}}\beta + \bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\right] && (72)\\
&= V\left[\bar{X}_{\text{out}}\beta\right] + V\left[\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\right] \\
&\quad +2\cdot cov\left[\bar{X}_{\text{out}}\beta, \bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\right] && (73)\\
&= A + B + 2\cdot C && (74)\\
A &= \frac{1}{m}\beta' V(X_{\text{out},i}\beta) = \frac{1}{m}\beta'\Omega_{X_{\text{out}}}\beta && (75)\\
B &= E\left[\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\varepsilon'_{\text{in}}X_{\text{in}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}}\right] \\
&\quad -\left\{E\left[\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\right]\right\}^2 && (76)\\
&= E\left\{E\left[\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\varepsilon'_{\text{in}}X_{\text{in}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}}|X\right]\right\} \\
&\quad -\left\{E\left\{E\left[\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}|X\right]\right\}\right\}^2 && (77)\\
&= E\left\{\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}E\left[\varepsilon_{\text{in}}\varepsilon'_{\text{in}}|X\right]X_{\text{in}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}}\right\} \\
&\quad -\left\{E\left\{\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}E\left[\varepsilon_{\text{in}}|X\right]\right\}\right\}^2 && (78)
\end{aligned}$$

Since $E\left[\varepsilon_{\text{in}}\varepsilon'_{\text{in}}|X\right] = \sigma_\varepsilon^2 I_n$ and $E\left[\varepsilon_{\text{in}}|X\right] = 0$, then

$$\begin{aligned} B &= \sigma_\varepsilon^2 E\left\{\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}}\right\} && (79)\\ C &= E\left[\bar{X}_{\text{out}}\beta\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\right] && (80)\\ &\quad -E\left[\bar{X}_{\text{out}}\beta\right]\cdot E\left[\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}X'_{\text{in}}\varepsilon_{\text{in}}\right] && (81)\\ &= 0 && (82)\\ V\left(\hat{\mu}_{U_{\text{out}}}\right) &= \frac{1}{m}\beta'\Omega_{X_{\text{out}}}\beta + \sigma_\varepsilon^2 E\left\{\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}}\right\}. && (83)\end{aligned}$$

It should be noted that $\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}} = O\left(\frac{1}{n}\right)$. Then, as $n \longrightarrow \infty$ and $m \longrightarrow \infty$, $V\left(\hat{\mu}_{U_{\text{out}}}\right) \longrightarrow 0$.

In practice, this variance can be estimated as follows:

$$\hat{V}\left(\hat{\mu}_{U_{\text{out}}}\right) = \frac{1}{m}\hat{\beta}'_{\text{in}}\hat{\Omega}_{X_{\text{out}}}\hat{\beta}_{\text{in}} + \hat{\sigma}^2_{\varepsilon_{\text{in}}}\bar{X}_{\text{out}}(X'_{\text{in}}X_{\text{in}})^{-1}\bar{X}'_{\text{out}},$$

where $\hat{\Omega}_{X_{\text{out}}}$ is estimated on the whole sample rather than only on X_{out} to increase the precision.

A.3 Covariance with mean cost

$Cov\left(\hat{\mu}_{U^T_{\text{out}}}, \bar{C}^T_{\text{out}}\right)$

$$\begin{aligned} &= Cov\left(\frac{1}{m^T}\iota'_{m^T}X^T_{\text{out}}\beta^T + \frac{1}{m^T}\iota'_{m^T}X^T_{\text{out}}(X^{T\prime}_{\text{in}}X^T_{\text{in}})^{-1}X^{T\prime}_{\text{in}}\varepsilon^T_{\text{in}}, \frac{1}{m^T}\sum_{i=n^T+1}^{n^T+m^T}C^T_i\right) && (84)\\ &= Cov\left(\frac{1}{m^T}\iota'_{m^T}X^T_{\text{out}}\beta^T, \frac{1}{m^T}\sum_{i=n^T+1}^{n^T+m^T}C^T_i\right),\\ &\quad + Cov\left(\frac{1}{m^T}\iota'_{m^T}X^T_{\text{out}}(X^{T\prime}_{\text{in}}X^T_{\text{in}})^{-1}X^{T\prime}_{\text{in}}\varepsilon^T_{\text{in}}, \frac{1}{m^T}\sum_{i=n^T+1}^{n^T+m^T}C^T_i\right), && (85)\\ &= \frac{1}{(m^T)^2}\sum_{i=n^T+1}^{n^T+m^T}Cov\left(\iota'_{m^T}X^T_{\text{out}}\beta^T, C^T_i\right),\\ &\quad + \frac{1}{(m^T)^2}\sum_{i=n^T+1}^{n^T+m^T}Cov\left(\iota'_{m^T}X^T_{\text{out}}(X^{T\prime}_{\text{in}}X^T_{\text{in}})^{-1}X^{T\prime}_{\text{in}}\varepsilon^T_{\text{in}}, C^T_i\right), && (86)\\ &= \frac{1}{(m^T)^2}\sum_{i=n^T+1}^{n^T+m^T}\iota'_{m^T}Cov\left(X^T_{\text{out}}, C^T_i\right)\beta^T,\\ &\quad + \frac{1}{(m^T)^2}\sum_{i=n^T+1}^{n^T+m^T}\iota'_{m^T}Cov\left(X^T_{\text{out}}(X^{T\prime}_{\text{in}}X^T_{\text{in}})^{-1}X^{T\prime}_{\text{in}}\varepsilon^T_{\text{in}}, C^T_i\right). && (87)\end{aligned}$$

Since $Cov\left(X^T_{j,k,\text{out}}, C^T_i\right) = \begin{cases} 0 \text{ if } j \neq i, \\ \gamma^T_k \text{ if } j = i. \end{cases}$, then

$$\iota'_{m^T}Cov\left(X^T_{\text{out}}, C^T_i\right)\beta^T = \gamma^{T\prime}\beta^T \text{ for all } i > n^T. \quad (88)$$

In addition, we have:
$Cov\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T}, C_i^T\right)$

$$= \quad E\left[\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T} - E(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T})\right)\left(C_i^T - \mu_C^T\right)\right]. \quad (89)$$

Since $E(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T}) = E(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'})E(\varepsilon_{\text{in}}^{T}) = 0$, then
$Cov\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T}, C_i^T\right)$

$$= \quad E\left[\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T}\right)\left(C_i^T - \mu_C^T\right)\right], \quad (90)$$

$$= \quad E\left\{E\left[\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T}\right)\left(C_i^T - \mu_C^T\right)|X\right]\right\}, \quad (91)$$

$$= \quad E\left\{\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\right)E\left[\varepsilon_{\text{in}}^{T}\left(C_i^T - \mu_C^T\right)|X\right]\right\}. \quad (92)$$

Since $i > n^T$ and since for ε_{in} the individuals j correspond to $j \leq n^T$, we have:

$$E\left[\varepsilon_{\text{in}}^{T}\left(C_i^T - \mu_C^T\right)|X\right] = E\left[\varepsilon_{\text{in}}^{T}\right]E\left[\left(C_i^T - \mu_C^T\right)|X\right] = 0. \quad (93)$$

By replacing expression of Equation 93 in Equation 92, we get:

$$Cov\left(X_{\text{out}}^{T}(X_{\text{in}}^{T'}X_{\text{in}}^{T})^{-1}X_{\text{in}}^{T'}\varepsilon_{\text{in}}^{T}, C_i^T\right) = 0. \quad (94)$$

By replacing expressions in Equation 88 and in Equation 94 into Equation 87, we get:

$$Cov\left(\hat{\mu}_{U_{\text{out}}^{T}}, \bar{C}_{\text{out}}^{T}\right) = \frac{1}{m^T}\gamma^{T\prime}\beta^{T}. \quad (95)$$

It should be noted that $Cov\left(\hat{\mu}_{U_{\text{out}}^{T}}, \bar{C}_{\text{out}}^{T}\right) \longrightarrow 0$ as $m^T \longrightarrow \infty$. $Cov\left(\hat{\mu}_{U_{\text{out}}^{T}}, \bar{C}_{\text{out}}^{T}\right)$ can be estimated as follows:

$$\widehat{Cov}\left(\hat{\mu}_{U_{\text{out}}^{T}}, \bar{C}_{\text{out}}^{T}\right) = \frac{1}{m^T}\hat{\gamma}_{\text{out}}^{T\prime}\hat{\beta}_{\text{in}}^{T}. \quad (96)$$

References

R. Ariza-Ariza, B. Hernández-Cruz, B. Carmona, M. D. Ruiz-Montesinos, J. Ballina, F. Navarro-Sarabia, The Costs, and Quality of Life in Rheumatoid Arthritis Study Group. Assessing utility values in rheumatoid arthritis: A comparison between time trade-off and the euroqol. *Arthritis and Rheumatism (Arthritis Care and Research)*, 55(5):751–756, October 15 2006. DOI 10.1002/art.22226, American College of Rheumatology. 1

R. Davidson and J. G. MacKinnon. *Estimation and inference in economics*. Oxford University Press, New York, 1993. 3.4

P. Dolan and M. Sutton. Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. *Soc Sci Med*, 44(10):1519–1530, 1997. 1

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 1993. 3.4

E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B*, 16:175–183, 1954. 3.1

P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992. 3.4

D.F. Heitjan. Fieller's method and net health benefits. *Health Economics*, 9:327–335, 2000. 3.1

J. S. U. Hjorth. *Computer Intensive Statistical Methods*. Chapman and Hall, London, 1994. 3.4

P. F. Krabbe, M. L. Essink-Bot, and G. J. Bonsel. The comparability and reliability of five health-state valuation methods. *Soc Sci Med*, 45(11):1641–1652, 1997. 1

L. Longworth, M. Buxton, M. Sculpher, and D. H. Smith. Estimating utility data from clinical indicators for patients with stable angina. *Eur J Health Econom*, 6:347–353, 2005. 1

E. Nord, J. Richardson, and K. Macarounas-Kirchmann. Social evaluation of health care versus personal evaluation of health states: Evidence on the validity of four health state scaling instruments using norvegian and australian survey data. Centre for Health Program Evaluation (CHPE), Working Paper 23, September 1992. 1

J. F. O'Leary, D. L. Fairclough, M. K. Jankowski, and J. C. Weeks. Comparison of time trade-off utilities and rating scale values of cancer patients and their relatives: evidence for a possible plateau relationship. *Med Decis Making*, 15(2):132–137, 1995. 1

O. Rivero-Arias, M. Ouellet, A. Gray, J. Wolstenholme, P. M. Rothwell, and R. Luengo-Fernandez. Mapping the modified rankin scale (mrs) measurement into the generic euroqol (eq-5d) health outcome. In *7th World Congress on Health Economics*, volume Harmonizing Health and Economics, Beijing, China, July 12–15 2009. International Health Economics Association. Poster presentation. 1, 3.2

J. A. Salomon and C. J. L. Murray. A multi-method approach to measuring health-state valuations. *Health Economics*, 13:281–290, 2004. Published online 20 June 2003 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/hec.834. 1

J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer-Verlag, New York, 1995. 3.4

A. Shmueli. It might be premature to reject the assumption of a power curve relationship between vas and sg data: Three comments on stevens, mccabe and braziers mapping between vas and sg data; results from the uk hui index 2 valuation survey. *Health Economics*, 16(in press):755–758, 2007. Published online 18 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hec.1188. 1

C. Siani and C. de Peretti. Is fieller's method applicable in all the situations? *GREQAM working paper*, 03A33, 2003. 3.1

C. Siani and C. de Peretti. Are fieller's and bootstrap methods really equivalent for calculating confidence regions for ratios. *Health, Decision and Management*, forthcoming, 2010. 3

C. Siani, C. de Peretti, and J. P. Moatti. Rehabilitating the confidence region for icers and which method to use: Fieller or re-ordered bootstrap? *Applied Health Economics and Health Policy*, 3(1):S62–S63, 2004. supplement. 3

C. Siani and J.-P. Moatti. Quelles méthodes de calcul des régions de confiance du ratio coût-efficacité incrémental choisir ? *Revue d'Epidémiologie et de Santé Publique*, 51: 255–276, 2003. 3

K. J. Stevens, C. J. McCabe, and J. E. Brazier. Mapping between visual analogue scale and standard gamble data; results from the uk health utilities index 2 valuation survey. *Health Econ., Health Economics Letters*, 15:527–533, 2006. Published online 3 January 2006 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/hec.1076. 1

G. W. Torrance. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-Econ Planning Sci*, 10(3):129–136, 1976. 1

G. W. Torrance, D. H. Feeny, W. J. Furlong, R. D. Barr, Y. Zhang, and Q. Wang. Multiattribute utility function for a comprehensive health status classification system. health utilities index mark 2. *Med Care*, 34(7):702–722, 1996. 1

A. Tsuchiya, J. Brazier, E. McColl, and D. Parkin. Deriving preference-based single indices from non-preference based condition-specific instruments: Converting aqlq into eq5d indices. Sheffield Health Economics Group, Discussion Paper Series Ref: 02/1; The University of Sheffield, ScHARR, School of Health and Related Research, May 2002. 1

N.C. Weber. On resampling techniques for regression models. *Statistics and Probability Letters*, 2:275–278, 1984. 3.4