Version provisoire préparée pour les 33èmes Journées des Économistes de la Santé Français

1-2 décembre 2011, CERDI, Université d'Auvergne

Should cost-utility analysis be replaced by cost-benefit analysis for the economic evaluation of health care programs?

A (somewhat) sceptical view

Philippe Tessier[·]

This version, October, 2011

Very preliminary draft

Comments are welcome - please do not quote without author's permission

Abstract. Cost-utility analysis (CUA) and Quality-Adjusted Life-Years (QALYs) are the most widely used methods of economic evaluation of health care programs. Yet, it is often argued in the literature that cost-benefit analysis (CBA) and willingness-to-pay (WTP) are theoretically superior techniques. Unlike QALYs, WTP is explicitly grounded on welfare economics principles and it neither imposes restrictive assumptions on the individual preference structures nor restricts the consequences of an intervention to changes in health and length of life. This paper provides arguments that challenge these alleged advantages of WTP over QALYs. It develops an alternative normative justification supporting a finding originally set out by Adler [*Yale J. Health Policy Law Ethics*, 6, 2006] according to which WTP does not dominate QALYs to represent individual preferences. It further discusses recent researches that lead to question the expected greater ability of WTP estimates to handle non-health outcomes. More specifically, the paper deals with three issues: (i) the determination of a normative framework allowing CBA and WTP to be ethically comparable to CUA and QALYs (ii) the comparison of the validity conditions of QALYs and WTP within this framework and (iii) their respective ability to handle non health-related consequences of health care. We conclude that the capacity of WTP to remedy some limitations of QALYs might reveal quite limited and that, consequently, the expected theoretical benefits of the replacement of CUA with CBA should not be overestimated.

[·] ERSSCa & Biostatistique, recherche clinique et mesures subjectives en santé (EA 4275), Faculté de Médecine, Université de Nantes. Email: philippe.tessier@univ-nantes.fr. Tel: (33) 2 53 48 46 03

1 Introduction

Although cost-benefit analysis (CBA) is commonly employed to assess natural resources management or road safety policies, it is largely supplanted by cost-utility analysis (CUA) to inform resources allocation decisions in the health sector. CUA values health benefits in terms of Quality-Adjusted Life-Years (QALYs). The QALYs gained from an intervention correspond to the sum of remaining life years weighted by quality of life factors placed on a scale where 0 and 1 stand for "death" and "perfect health" respectively (Weinstein and Stason, 1977). By contrast, CBA values health care outcomes in monetary terms which may explain the long reluctance of health professional and decision makers to endorse it (Fuchs, 1980; Weinstein, 1995). The spread of CBA in the health field has also been slow down by concerns about the feasibility and the reliability of contingent valuation methods (Cookson, 2003; Smith & Sachs, 2010). However, since the mid-1990s many efforts have been devoted to refine the measurement of individual willingness-to-pay (WTP) for health interventions (O'Brien and Gafni, 1996; Gafni, 1997; Klöse, 1999; Baker et al., 2010). The choice between CUA and CBA to guide resources allocation within health programs is still an opened question, a large part of it being related to the comparison of QALYs and WTP.

Despite the broad use of QALYs in the health sector, it is widely acknowledged in the literature that WTP is a theoretically superior method, at least from a welfarist point of view. Olsen and Smith (2001) identify three common justifications for this. First, WTP measurement is grounded on welfare economics principles whilst the welfare theoretic foundations of QALYs are still debated (Kenkel, 1997; Liu et al., 2009; Liljas, 2010). Moreover, it has been shown that CUA ranks health interventions consistently with CBA under unrealistic assumptions (Bleichrodt and Quiggin, 1999; Dolan and Edlin, 2002). Second, unlike QALYs, WTP neither requires limiting the consequences of health care to health and longevity nor imposes restrictive conditions on individual preferences (Pauly, 1995; Klöse, 2003; Freeman, 2006; Gafni, 2006). Last, as WTP is expressed in the same monetary unit than costs, CBA can be employed to determine the optimal share of resources that should be allocated to the health care sector (Pauly, 1995; Gafni, 2006; Cohen and Patel, 2009). Unless the social monetary value of a QALY is estimated, CUA cannot address the issue of allocative efficiency (Kenkel, 1997).

The above-mentioned advantages of WTP and CBA over QALYs and CUA are widely recognized (Freeman, 2006; Gafni, 2006) and, setting apart empirical difficulties, they could justify replacing the former with the latter. Yet, they have been partly challenged by Adler (2006). Adler claims that social welfare should be defined according to Harsanyi's utilitarianism (see Harsanyi, 1955). He then shows that, within this framework, there may be cases where QALYs provided a valid cardinal measure of individual welfare whereas WTP fails to do so (Adler, 2006). However, Harsanyi's utilitarianism relies on rather specific normative assumptions that are not free of critics (on this see e.g. Mueller, 2003). This could prevent Adler's analysis of receiving as much as attention it deserves. The first aim of this paper is thus to provide an original argument upon which we derive a result similar to that of Adler. The main benefit of our approach is to support the non-superiority of WTP over QALYs to represent preferences by relying on a more pragmatic and hopefully less controversial normative justification than Harsanyi's utilitarianism. A second aim is to highlight how the findings of recent researches

suggest that the greater inclusiveness of WTP might be less beneficial than expected. In sum, the main contribution of our analyses is to mitigate the alleged theoretical advantages of WTP over QALYs and thus, to minimize some of the expected benefits of a move from CUA to CBA for the economic evaluation of health care programs.

The remainder of the paper is as follows. We start by discussing the foundations of CBA in order to determine under which conditions it can be viewed as an ethically relevant alternative to CUA (section 2). We argue that this requires more than the principle of welfarism with the implication that WTP estimates have to represent the intensity of preferences. Section 3 compares the validity conditions of QALYs and WTP and discusses the possibility to link CUA and CBA within this framework. Lastly, we review arguments suggesting that the definition of QALYs can be broadened to encompass more than health outcomes whereas the inclusiveness of WTP may create inescapable risks of double counting (section 4). Section 5 concludes.

2 CBA and WTP as ethically acceptable alternatives to CUA and QALYs

2.1 Old-fashioned versus modern CBA

Since the ordinalist revolution of the 1930s, economists define utility as a mathematical function representing the order of preferences inferred from an individual's observed or stated choices. Strength of preference measures are either regarded as meaningless or irrelevant for economic analysis purposes (Arrow, 1951; Kolm, 1995). Cardinal (riskless) utility has nevertheless remained present in several fields of research such as optimal taxation or social choice for example. Consequently, economists are sometimes unclear about the properties they grant to utility measures, especially for what concerns WTP estimates (Morey, 1984; Wriglesworth and Gravelle, 1987). For instance, in his classical microeconomics textbook Varian (1993) claims that utility is intrinsically ordinal in nature but he also acknowledges that economists sometimes need measures of utility differences. The interpretation of WTP seems similarly controversial in the health economics literature where it is alternatively held to represent the order (Liljas and Lindgren, 2001; Klöse, 2003; Bayoumi, 2004; Krupnick, 2004; Gafni, 2006) and the intensity of preferences (Donaldson et al., 1997; Olsen and Smith, 2001; Donaldson, 2004). This may be explained by the existence of two competing theoretical bases for CBA.

Historically, CBA is rooted in the Kaldor-Hicks efficiency criterion (KHC) that identifies Potential Pareto Improvements. A policy passes the KHC test if those who gain from it could potentially compensate the losers and still be themselves better off (Bruce and Boadway, 1984). This criterion avoids resorting to a cardinal and interpersonally comparable notion of utility (Arrow, 1951; Fleurbaey and Hammond, 2004). However, it is severely criticized for it is built on a hypothetical compensation mechanism and it can lead to inconsistent orderings of options (Boadway, 1974; Blackorby and Donaldson, 1990). To avoid these shortcomings, modern theorists justify CBA by reference to Bergson-Samuelson social welfare functions (SWF) that allow reflecting various normative ethical judgements about social welfare (Just et al., 2004). Unlike the KHC, a SWF allows a complete ordering of states of affairs. In return, this approach is much more demanding in terms of individual preferences measurement. Defining a SWF whilst escaping the dilemma raised by Arrow's general possibility theorem

(Arrow, 1951) indeed requires a cardinally and intercomparable measure of utility (Mueller, 2003).

CBA is still mostly justified by reference to the Kaldor-Hicks criterion nowadays in the health economics (Garber et al., 1996; Gafni, 2006) as in the wider economics literature (Zerbe & Bellas, 2006). Ranking social states according to the net sum of individual WTP could alternatively be justified by reference to a utilitarian SWF along with an assumption of equal marginal social utility of income across the individuals. The choice between these two justifications may be driven by beliefs about the intrinsic nature of utility and by practical considerations. Indeed, by appealing to Occam's razor one might retain the old-fashioned justification for CBA since, despite its shortcomings, it is far less informational demanding than the SWF based approach. The KHC can thus be invoked to defend CBA when the assumption of equal marginal utility of income does not hold (Liljas and Lindgren, 2001). Hence, to quote Feldman (1998, p. 421), we could follow "the applied economist [who] uses cost-benefit analysis, consumers' surplus measures and the Kaldor-Hicks test to boldly go where the theorist fears to tread". If one follows this line of reasoning, WTP estimates can be interpreted as representing the order of preferences and no appeal to any concept of cardinally and intercomparable utility with which many economists feel uncomfortable is necessary.

2.2 Weighted CBA as a substitute to CUA

CUA can either be justified on a welfarist or on an extra-welfarist basis (Culyer, 1989; Brouwer et al., 2008). Theoretical comparisons between CUA/QALYs and CBA/WTP make sense if they are circumscribed to the principle of welfarism that conceives social welfare as a function of individual preferences only[1] (Bleichrodt and Quiggin, 1999; Dolan and Edlin, 2002). However, welfarism alone might not be sufficient to ensure that CBA is an ethically acceptable alternative to CUA, i.e. to convince those who actually favour CUA to use CBA if WTP proves to be theoretically superior to QALYs to represent preferences. CUA and CBA can be both justified by referring to a utilitarian "unweigthed" SWF (Borghi, 2008). Nonetheless, their underlying objective functions are different. As Brekke (1997) pointed out, if the unweighted sum of net benefits is used as a decision criterion, changing the unit of measurement for individual utility may well change the conclusion of the analysis as well. Therefore, despite their common appeal to utilitarianism a move from CUA to CBA is not equity neutral[2]. Besides, the widespread use of CUA in the health field seems to be primarily attributed not to its adherence to welfarism, as is illustrated by extra-welfarist versions of it, but rather to the fact that by valuing effectiveness without referring to income it does not weigh health care benefits according to individual ability to pay (Gold et al., 1996; Weinstein and Manning, 1997; Brouwer et al., 2008). Accordingly, we believe that for CBA to be a normatively relevant alternative to CUA it should be designed so as to reflect similar ethical concerns.

[1] Extra-welfarist foundations for CUA have been originally proposed by Culyer (1989). In this framework, QALYs are intended to measure health not utility such that the objective of CUA is the maximization of a population's health (Wagstaff, 1991; Brouwer et al., 2008).

[2] This is so since individual utilities are summed up without adjusting for the marginal utility of life years in perfect health and the marginal utility of income. In theory, changing the numéraire does not affect social welfare measurement (see Drèze, 1998; Johansson, 1998).

To be clear, we do not claim that CBA should adhere to the rule of maximizing the sum of QALYs. Our point is that the unweighted sum of WTP hardly provides a near equivalent decision criterion to CUA even if the marginal utility of income is constant between the individual. This is so since the marginal utility of health may depend on individual wealth. This nevertheless leaves a large choice of a morally acceptable form of weighted CBA to compete with CUA. For instance, an approach that has already been put into practice in the health field consists in defining distributive weights that are inversely proportional to income (Donaldson, 1999; Borghi, 2008). Another possibility is to employ weights that reflect a constant social value of a statistical life across the individuals (Johansson-Stenman, 2000; Somanathan, 2006; Baker et al., 2008). Although these approaches do not adjust WTP for individual differences in marginal utility of health and income, they aggregate individual benefits in such a way that can compensate income inequalities. As such, they are more likely to be perceived as ethically substitutable approaches to CUA than traditional CBA.

At least an important objection can be raised against our proposal. Donaldson et al. (2002) observed that CUA is not an 'income free' method. For example, income is related to life expectancy which is used to calculate the number of QALYs a person gains from an intervention. Furthermore, productivity losses due to morbidity that can be included in the numerator of a cost-utility ratio are dependent from the income distribution. Therefore, the authors conclude that "if CBA is to be rejected because of its 'susceptibility' to income distributions, alternatives such as [CUA], which are equally susceptible, cannot be used as a more meaningful approach." (Donaldson et al., 2002, p. 66). This is a strong objection based on hardly debatable observations. Yet, two counterarguments can be raised against it. First, there may be a fundamental difference between observing that the outcomes of health care are partially income-dependent, as is the case for life expectancy or productivity losses, and assuming that the social weights ascribed to measures of preferences can depend on individual ability to pay as traditional CBA allows. Hence it is doubtful that CUA and CBA are equally affected by income distribution. Second, the important point for what concern us is not whether CUA is 'income free' but rather whether it is less income-dependent than CBA and whether they are some forms of weighted CBA that limit the influence of income on distributive weights as CUA tries to do.

The idea we defend is thus that CBA via distribution weights can endorse a value judgement underpinning CUA: that the influence of individual wealth on the weighting of health benefits should be as limited as possible. Accordingly, CBA has to depart from the Hicks-Kaldor criterion if it is to be seen as a possible substitute to CUA since this criterion precisely seeks to avoid the use of distributive weights when aggregating individual WTP. This has two main implications at least. One relates to the possibility to address the issue of allocative efficiency. If a weighted sum of WTP estimates is employed as a decision criterion in the health domain, this possibility holds only if the same distributive scheme is used in all areas where CBA is applied. Otherwise, the estimated net benefits provided by a health program do not reflect the true opportunity costs of the resources diverted from other non-health sectors. Note that this is a practical not a theoretical issue, although probably a very difficult one. Another more fundamental implication, whose consequences are examined in the following section, concerns the interpretation of WTP estimates. If CBA is justified by reference to a SWF, it is ne-

cessary to move from an ordinal to a cardinal and interpersonally comparable notion of utility. The consequences of such a move have been largely ignored to date when studying the validity of WTP measures (Adler, 2006).

3 Is WTP theoretically superior to QALYs to represent preferences intensity?

In this section, we shall compare the validity conditions of QALYs and WTP when the latter is explicitly interpreted as a cardinal utility function. The analysis is set in a context of certainty. Following Broome (1993), we think that risk is a complication that should not be introduced to early in the analysis[3]. We also discuss the implications of our analysis for the interpretation of a link between CBA and CUA.

Let $H_T = (h_1, h_2, \ldots, h_i, \ldots, h_T)$ represent an health profile, *i.e.* a stream of periodic health states h_i over the lifetime T, and U be a utility function defined over the set of all possible health profiles. The QALY model under certainty defines U as an additive function:

$$U(h_1, h_2, ..., h_i, ..., h_T) = \sum_{i=1}^{T} u(h_i) \quad (1)$$

In equation 1, u is an instantaneous utility function for health. The number of QALYs is computed by normalizing u such that $u(h^d) = 0$ and $u(h^p) = 1$ where h^d represents death and h^p stands for 'perfect health'. Two main conditions are crucial to derive this model (Broome, 1993; Bleichrodt, 1995). One is that of mutual preference independence (MPI) between sets of periodic health states: if two health profiles have an outcome in common preferences cannot be reversed when this outcome is changed into another common outcome. This implies that the utility function is additive (Keeney and Raiffa, 1976). A second central condition is that of MPI between health and non-health attributes to ensure the existence of a utility function for health. Adding the assumptions of no discounting and of stable preference over time provides a complete characterization of the QALY model (Bleichrodt, 1995).

Consider now the definition of the WTP for a health improvement. Suppose that the life of an ill person is described by a health profile $H^1 = (h_1^1, ..., h_T^1)$. A medical intervention can allow her to live a preferred life $H^2 = (h_1^2, ..., h_T^2)$. Let's note W^1 the present value of the person's stream of future incomes. Her WTP for the treatment corresponds to the maximum amount of money C she is willing to give up from her present value of incomes after the intervention so that she is indifferent between the two extended life profiles:

$$(H^1, W^1) \sim (H^2, W^2 - C) \quad (2)$$

In relation 2, ~ denotes indifference and C is the person's overall compensating variation. If C is increasing in utility the amounts C represents the individual's order of preferences (Klöse, 2003). This justifies the common claim that the measurement of WTP requires fewer assumptions than for QALYs (Gafni, 1997; Liljas and Lindgren, 2001; Klöse, 2003; Gafni, 2006). Yet, a different conclusion is reached if WTP is intended to measure the strength of preferences.

[3] This contrasts with Adler (2006) who refers to Harsanyi's utilitarianism which implies that he interprets WTP as a von Neumann and Morgenstern (risky) utility function.

Assume for the sake of simplicity, that income and health are the only factors of interest. Using WTP estimates to measure the intensity of preferences over the health attribute supposes that differences in preferences for this attribute can be "priced out" by differences in preferences for the monetary attribute (Farquhar and Keller, 1984). This implicitly assumes that the utility function over health and money is additively separable[4] (Sarin, 1982). Furthermore, a utility function that measures preferences differences is unique up to an affine transformation (Dyer and Sarin, 1979). This means that if money serves as a measure of the strength of preferences it is an affine transformation of a meaningful index of cardinal utility, which implies that the marginal utility of income is constant[5] (Morey, 1984). One can see that these two conditions entail a quasi-linear utility function over health and money such as in relation 3:

$$U(H, W) = u(H) + W \quad (3)$$

Where u is a cardinal utility function over health states and W represent a linear cardinal utility function over income. Interestingly, relation 3 allows retrieving two situations described by Adler (2006) where WTP amounts fail to represent a cardinal utility function which is comparable between the individuals at a constant rate. These situations relate to the possibility of decreasing marginal utility of consumption and of interactions between consumption and health. Thus, in contrast with the case of ordinal preferences, WTP estimates are not consistent with cardinal utility theory unless the individual preference structures satisfy some quite restrictive conditions (see also Wriglesworth and Gravelle, 1987).

According to the above finding, whenever CBA is justified in reference to a social welfare function it cannot be claimed that WTP is a theoretically superior measure of preferences than QALYs. If outcomes are defined in terms of health and wealth, WTP and QALYs require that preferences over these arguments are additively or mutually independent of one another. Hence, the choice between QALYs and WTP is an empirical not a theoretical issue since it depends on the descriptive validity of the assumptions of constant marginal utility of income and of no discounting of life years. Furthermore, if both QALYs and WTP are valid measures of cardinal utility choosing between them might have limited consequences. Klöse (2003) has shown that a wealth-standardized version of the QALY model, where quality weights are defined for a given wealth level, can be defined consistently if the marginal utility of wealth is independent of health. If WTP represents preference differences this condition is fulfilled since the marginal utility of income is constant. It thus turns out that the validity of WTP is a sufficient condition to ensure that wealth-standardized QALYs are consistent with the maximisation of lifetime preferences, i.e. to ground QALYs on welfare theoretic foundations. In the same vein, the impossibility theorem for a link between CBA and CUA proved by Dolan

[4] Formally, the indifference relation $(h_1, w_1) \sim (h_2, w_2)$ indicates that the utility difference $u(h_1) - u(h_2)$ between health states h_1 and h_2 equals the utility difference between incomes w_1 and w_2, e.g. $v(w_1) - v(w_2)$, where u, v are strength of preference functions over health and money respectively. Note that this representation supposes that a utility function over health exists. This condition can be removed by extending the utility function over non-health outcomes other than income.

[5] If income W is a cardinal utility function, one has $W = a + b.U$ where U is a true index of cardinal utility and a, b are constant scalars. One sees that b is the inverse of the marginal utility of money ($\partial U/\partial M$) which is thus constant (see Morey, 1984, p. 166-167).

and Edlin (2002) is slightly weakened when equation 3 holds. The authors show that it is not possible to link CBA and CUA if the axioms of expected utility theory hold, QALYs are valid measures of preferences and illness influences the individuals' ability to enjoy consumption. Again, whilst the latter assumption is descriptively intuitive it is precluded by the validity conditions of WTP. Consequently, provided that we restrict our attention to health and wealth, interpreting WTP as a valid measure of the intensity of preferences may help to ground CUA on welfare economics principles. As a result, CUA may lead to recommendations similar to those provided by CBA for the allocation of a fixed budget[6] (see theorems 2 & 3 in Dolan and Edlin, 2002).

To sum up, if WTP measures cardinal utility it can neither be argued that it is theoretically less restrictive than QALYs nor that the latter cannot be explicitly linked to welfare economics principles. This finding also suggests that the consequences of a move from CUA to CBA might have limited consequences in terms of programs ranking.

4 Is WTP more inclusive than QALYs?

In principle, WTP estimates are more inclusive measures of utility than QALYs since they are not restricted to the valuation of health outcomes. In this section, we build on recent researches to show that (i) when properly interpreted QALYs can be broadened to deal with some non-health consequences of health care and that (ii) in some specific circumstances the inclusiveness of WTP can lead to double counting in economic evaluations.

4.1 Including non-health outcomes in QALYs measurement

According to equation 1, QALYs are a utility function over health states and length of life that does not include non-health factors. Yet, recent empirical studies show that when they are faced to methods of preferences measurement over health states, such as the time trade-off or the standard gamble for instance, some respondents do consider their anticipated income level even if they are not explicitly instructed to do so (Sendi and Brouwer, 2005; Krol et al., 2006; Richardson et al., 2009). Furthermore, Tessier (2011) suggested that restricting QALYs to the assessment of health-related quality of life is more a matter of interpretation than an intrinsic limitation of the model. In order to consistently include the value of income and leisure, he proposes to reinterpret QALYs as some sort of a conditional utility function.

Denote by w^p the income that the individual imagines he could earn if he were in perfect health and by l^p his corresponding leisure time. It seems plausible to assume that this individual anticipates these values given his actual vector e of personal characteristics such as his educational, his professional or his marital status for instance. More precisely, e encompasses all elements that can determine future levels of w and l beyond the direct influence of health. Hence, when asked to imagine living in a hypothetical health state h, a person is supposed to attribute to it unique anticipated amounts of wealth and leisure conditional of her actual of personal characteristics. Tessier (2011) then proposes to reinterpret the QALY model as a conditional utility function V (equation 4).

[6] Dolan and Edlin (2002) note that the possibility of a link either requires that only health and wealth enter the individual utility function or that other non-health factors are held constant.

$$V(H,W,L \mid E) = \sum_{i=1}^{T} v(h_i(c), w_i(h_i,e), l_i(h_i,e)) \quad (4)$$

In equation 4, c denotes the health care consumed by the individual. $w(h, e)$ and $l(h, e)$ are the anticipated amounts of income and leisure corresponding to health state h given the individual's actual vector e of personal characteristics. Relation 4 thus represents conditional preferences in the sense that the utility function v is defined on a restricted domain that depends on e. More precisely, v is defined for feasible values of w and l that an individual can reach when her health varies given her actual profile of non-health characteristics e. The function v can be normalized by posing $v(h^p, w^p, l^p) = 1$ and $v(h^d, 0, 0) = 0$. This function is a broader measure of utility than the weights used to calculate QALYs since it includes income and leisure between its arguments. Note that although one can set a conditional QALY to be of equal value to whomever gets it, the "maximum" values w^p and l^p will not be the same for each individual.

Reinterpreting QALYs as conditional utility functions thus provides a simple way to consistently include income and leisure changes in quality of life weights (Tessier, 2011). This interpretation is also consistent with the suggestion of the Washington Panel of experts on cost-effectiveness (Gold et al., 1996) who once recommended that individual income changes due to illness should be accounted for in CUA when measuring quality of life. If changes in income and leisure are the most important non-health outcomes of health care then the possibility to define QALYs as conditional utilities, and to measure preference-based quality weights accordingly, partly removes one of the comparative advantages of WTP over QALYs.

4.2 Inclusiveness of WTP and risks of double counting

A not much discussed issue when comparing QALYs and WTP is whether the latter is able to replace the former in all circumstances. In particular, we believe that there might be good reasons to be prudent in invoking the broad inclusiveness of WTP as an advantage for economic evaluation. From a welfare theoretic point of view, all costs and benefits accruing to the individuals should be accounted for in economic evaluations (Meltzer, 1997). Yet, this does not mean that inclusive measures of WTP are always the best way to do so.

Consider the issue of the estimation of productivity losses due to illness for instance. These costs have two distinct components: an internal one that relates to the loss of income an ill person may suffer and an external one since part of the productive output of the ill person who does not work anymore is lost to the rest of society (Weinstein et al., 1997). Currie et al. (2002) pointed out that when productivity losses are estimated as a separate item in economic evaluations the WTP for health care benefits has to be assessed net of the value of individual income changes. Otherwise, the internal part of productivity costs is counted twice since it will also be included in WTP amounts. Hence, even though WTP is a broadly defined measure of individual preferences, it can be necessary to assess it in a "restricted" form so as to avoid risks of double counting (Currie et al., 2002; Drummond et al., 2005). Tessier et al. (2011) also identified a specific situation, yet generic enough, when the inclusiveness of WTP proves to be even more problematic.

The traditional approach to the estimation of productivity losses due to illness uses the human capital approach. This consists in multiplying the potential time of work lost by the individual's gross wage so as to estimate the expected present value of forgone consumption streams (Rice, 1967). Koopmanschaap et al. (1995) argue that this approach overestimates the "real" losses incurred by society when the economy is not at full employment. They suggest taking into account the possibility to replace an ill worker by a previously unemployed person. This leads to reduce productivity losses to a friction period that corresponds to the time necessary for production to return at its initial level (Koopmanschaap et al., 1995). Thus, if the friction costs approach is employed the measurement of health care benefits has to be net of the value of individual income changes. As suggested above, this can be done by asking people to state their WTP for health care whilst imagining that if they become ill their income will be held constant (Drummond et al., 2005). However, this strategy is inapplicable to the monetization of the consequences of life saving programs. In this case, the most commonly used approach consists in estimating the value of a statistical life (VSL) which is derived from questions that ask the individuals to state their WTP for small reductions in the risk of death. Answers to such questions are likely to be influenced by income (Dolan et al. 2008; Viscusi and Aldy 2003). Furthermore, due to the various transfers that replacement at work by an unemployed person involves the death of a working person does not change the total monetary value of society's global consumption[7] (Tessier et al., 2011). This implies that using VSL estimates alongside the friction costs approach will lead to double count the internal part of productivity costs. Tessier et al. (2011) suggested that a possibility to avoid this problem would lie in the association of WTP and QALYs. The idea is that through the estimation of the WTP for a QALY one may obtain a monetary estimation of the non-material part of the value of life (see Mason et al., 2009). Otherwise stated, under the specific situation just described WTP and QALYs seem to be better seen as complementary rather than substitutable. In this specific case QALYs could be invoked to solve a problem encountered when using WTP and CBA not the reverse. This illustrates that the inclusiveness property of WTP estimates should not be systematically regarded as an advantage of this method.

5 Conclusion

In this paper, we have proposed and reviewed several reasons to qualify the theoretical advantages ascribed to WTP as compared to QALYs to represent the benefits of health care interventions. Our findings are twofold. First, we showed that the measurement of WTP requires restrictive conditions on the individual preferences structure when one interprets CBA as an ethically relevant alternative to CUA. Hence, while we do not endorse Adler's (2006) view that social welfare should be defined according to Harsanyi's utilitarianism, we share his conclusion about the non-theoretical superiority of WTP over QALYs. Interestingly, this result also suggests that differences between CUA and CBA might be less dramatic in terms of programs ranking than has sometimes be claimed. Second, we further pointed out that the benefits from the greater inclusiveness of WTP, i.e. its ability to include non-health factors, might be less advantageous than expected. This stems from the possibility to consistently broaden

[7] This does not mean that the death of a person comes at no cost for the society but only that this cost corresponds to the premature loss of years of life and of health-related quality of life.

the definition of preference-based quality weights in QALYs measurement and from the possibility of double counting in economic evaluations when all-inclusive WTP are assessed. In sum, our analyses suggest that under a welfarist perspective there might be no conclusive argument in favour of the theoretical superiority of WTP in comparison with QALYs to represent individual preferences. Consequently, the advantages of a move from CUA to CBA could be rather limited.

To be clear, we do not claim that CUA has to be preferred to CBA or that the latter has no advantages over the former. Indeed, our discussions were mainly concerned by the comparison of some of the actual and potential properties of WTP and QALYs. There are obviously other aspects to consider in forming a choice between CUA and CBA that we did not discuss. For instance, one can ask whether employing an incremental cost-utility ratio as an evaluation criterion can lead to an efficient allocation of resources within a fixed budget constraint (see Gafni, 2006). In addition, part of our analysis relies on a central normative argument that given the predominance of CUA in the health domain CBA should endorse some of its underlying value judgements to be seen as an ethically acceptable substitute. Even if we believe that this assertion might be less controversial than reliance on Harsanyi's utilitarianism as proposed by Adler (2006), it is still a value judgement to which one may not necessarily adhere. If CBA is justified by reference to the Kaldor-Hicks criterion as is still common in the literature then WTP is a less restrictive measure of preferences than QALYs. Yet, in this case we think it is doubtful that CBA and CUA are normatively comparable to guide resources allocation decisions in health care.

The main message of our paper is thus to warn against the use of WTP estimates as a remedy to some of the well-known limitations of QALYs and hence against the possible overestimation of the benefits of substituting CBA to CUA. Consequently, we believe that the search for improvement in the methods of WTP/CBA and QALYs/CUA on the theoretical and on the empirical sides offer more promising avenues for progress than trying to establish the theoretical superiority of one approach over the other. Finally, it has to be mentioned that our discussion has been confined to the welfarist interpretation of QALYs and CUA. Consideration of extra-welfarist versions of them may also offer invaluable possibilities for the development of ethically acceptable methods of economic evaluation to inform decision making in the health sector.

REFERENCES

Adler MD. 2006. QALYs and policy evaluation: a new perspective. *Yale Journal of Health Policy, Law and Ethics* VI: 1–92.

Arrow KJ. 1951. *Social Choice and Individual Values*. Wiley: New York.

Baker R, Chilton S, Jones-Lee M, Metcalf H. 2008. Valuing lives equally: Defensible premise or unwarranted compromise? *Journal of risk and uncertainty* 36: 125-138.

Baker R, Currie GR, Donaldson C. 2010. What needs to be done in contingent valuation: have Smith and Sach missed the boat? *Health Economics, Policy and Law* 5:113-121.

Bayoumi AM. 2004. The Measurement of Contingent Valuation for Health Economics. *Phamracoeconomics*. 22:691-700.

Blackorby C, Donaldson D. 1990. The Case against the Use of the Sum of Compensating Variations in Cost-Benefit Analysis. *Canadian Journal of Economics* 23:471–94.

Bleichrodt H. 1995. QALYs and HYEs: Under What Conditions Are They Equivalent? *Journal of Health Economics*. 14:17-37.

Bleichrodt H, Quiggin J. 1999. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis? *Journal of Health Economics* 18: 681-708.

Borghi J. 2008. Aggregation rules for cost-benefit analysis: a health economics perspective. *Health Economics*17:863-75.

Brekke KA. 1997. The numéraire matters in cost-benefit analysis. *Journal of Public Economics*. 64: 117–123.

Broome J. 1993. Qalys. *Journal of Public Economics* 50: 149-167.

Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH. 2008. Welfarism vs. extra-welfarism. *Journal of Health Economics* 27:325-338.

Boadway RW. 1974. The Welfare Foundations of Cost-Benefit Analysis. *Economic Journal*. 84:926-939.

Boadway RW, Bruce N. 1984. *Welfare Economics*. Basil Blackwell.

Cohen DR, Patel N. 2009. The Potential to Forgo Social Welfare Gains through Overreliance on Cost Effectiveness/Cost Utility Analyses in the Evidence Base for Public Health. *Journal of Environmental and Public Health*

Cookson R. 2003. Willingness to pay methods in health care: a sceptical view. *Health Economics* 12: 891–894.

Culyer AJ. 1989. The normative economics of health care finance and provision. *Oxford Review of Economic Policy* 5: 34–56.

Currie GR, Donaldson C, O'Brien BJ, Stoddart GL, Torrance GW, Drummond MF. 2002. Willingness to pay for what? A note on alternative definitions of health care program benefits for contingent valuation studies. *Medical Decision Making* 22(6): 493–497.

Dolan P, Metcalfe R, Munro V, Christensen MC. 2008. Valuing lives and life years: anomalies, implications, and an alternative. *Health Economics Policy and Law*. 3: 277–300.

Dolan P, Edlin R. 2002. Is it Really Possible to Build a Bridge between Cost–benefit Analysis and cost–effectiveness Analysis? *Journal of Health Economics* 21:827–843.

Donaldson C. 1999. Valuing Publicly Provided health-care: Does 'Ability to Pay' Preclude the Use of 'Willingness to Pay'? *Social Science and Medicine* 49:551–563.

Donaldson C, Birch S, Gafni A. 2002. The Distribution Problem in Economic Evaluation: Income and the Valuation of Costs and Consequences of health-care Programs. *Health Economics* 11:55–70.

Donaldson C, Farrar S, Mapp T, Walker A, Macphee S. 1997. Assessing community values in health care: is the 'willingness to pay' method feasible? *Health Care Analysis* 1:7-29.

Drèze J. 1998. Distribution matters in cost-benefit analysis: Comment on K.A. Brekke. *Journal of Public Economics*. 70:485–488.

Drummond M, Schulper MJ, Torrance GW, O'Brien B, Stoddart GL. 2005. *Methods for the evaluation of health care programmes*. Oxford: Medical Publications.

Dyer JS, Sarin RK. 1979. Measurable multiattribute value functions. *Operations Research*. 27: 810–822.

Farquhar PH, Keller LR. 1989. Preference Intensity Measurement. *Annals of Operations Research* 19:205-217.

Feldman D. 1998. *Kaldor-Hicks compensation*. in Newman P. (ed). 1998. The New Palgrave Dictionary of Economics and the Law, Macmillan Reference Ltd., London, vol.2:417-421.

Fleurbaey M, Hammond PJ. 2004. Interpersonally comparable utility. in Barbera S, Hammond PJ, Seidl C. Eds., *Handbook of Utility Theory*, vol. 2, Dordrecht: Kluwer.

Fuchs VR. 1980. What is CBA/CEA, and Why Are They Doing This to Us ? *The New England Journal of Medicine*. 303:937-938.

Gafni A. 1997. Willingness-to-pay (WTP) in the Context of an Economic Evaluation of health-care Programs: Theory and Practice. *The American Journal of Managed Care* 3:521–532.

Gafni A. 2006. Economic evaluation of health-care programmes: is CEA better than CBA? *Environmental and Resource Economics* 34: 407–418.

Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). 1996. *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press.

Garber AM, Weinstein MC, Torrance GW, Kamlet MS. 1996. Theoretical foundations of cost-effectiveness analysis. In Gold MR, Siegel JE, Russell LB, and Weinstein MC, edi-

tors, *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, pp. 25-53.

Harsanyi JC. 1955. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*. 63:309-321.

Johansson P.-O. 1998. Does the choice of numeraire matter in cost–benefit analysis. *Journal of Public Economics*. 70: 489–493.

Johansson-Stenman O. 2000. On the Value of Life in Rich and Poor Countries and Distributional Weights Beyond Utilitarianism. *Environmental and Resource Economics* 17:299-310

Just R, Hueth D, Schmitz A. 2004. *The Welfare Economics of Public Policy*. Edward Elgar Ltd.

Keeney R., Raiffa H. 1976. Decisions with multiple objectives: Preferences and value tradeoffs. Wiley: New York.

Kenkel D. 1997. On Valuing Morbidity, cost–effectiveness Analysis, and being Rude. *Journal of Health Economics* 16:121–128.

Klöse T. 1999. The Contingent Valuation Method in health-care. *Health Policy* 47:97–123.

Klöse T. 2003. A Utility-theoretic Model for QALYs and Willingness-to-pay. *Health Economics* 12:17–32.

Kolm SC. 1995. Economic justice: The central question. European Economic Review. 39: 661-673.

Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. 1995. The friction cost method for measuring indirect costs of disease. *Journal of Health Economics*. 14:171–189.

Krol M, Brouwer W, Sendi P. 2006. Productivity cost in health state valuations: does explicit instruction matter. Pharmacoeconomics. 24(4): 401-.

Krupnick AK. 2004. *Valuing Health Outcomes: Policy Choices and Technical Issues*. Washington, DC: Resources for the Future.

Liljas B. 2010. On the welfare theoretic foundation of cost-effectiveness analysis - the case when survival is not affected. *European Journal of Health Economics* 11:5–13.

Liljas B, Lindgren B. 2001. On individual preferences and aggregation in economic evaluation in health care. *Pharmacoeconomics*. 19: 323-335.

Liu L, Rettenmaier AJ, Saving TR. 2008. Longevity bias in cost-effectiveness analysis. *Health Economics* 17:523–534.

Mason H, Jones-Lee M, Donaldson C. 2009. Modelling the monetary value of a QALY: a new approach based on UK data. *Health Economics*.18(8): 933–950.

Morey ER. 1984. Confuser Surplus. *American Economic Review*. 74:163-173.

Mueller DC. 2003. *Public Choice III*. Cambridge: Cambridge University Press.

O'Brien B, Gafni A. 1996. When do the ''Dollars'' Make Sense? Toward Conceptual Framework for Contingent Valuation Studies in health-care. *Medical Decision Making* 16:288–299.

Olsen JA, Smith RD. 2001. Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Economics* 10: 39–52.

Pauly M. 1995. *Valuing health care benefits in money terms*. in Sloan F (ed.), Valuing Health Care: Costs, Benefits and Effectiveness of Pharmaceuticals and Other Medical Technologies, Cambridge: Cambridge University Press.

Rice D.P. 1967. Estimating the costs of illness. *American Journal of Public Health* 57:424-440.

Richardson J, Peacock SJ, Iezzi A. 2009. Do quality-adjusted life years take account of lost income? Evidence from an Australian survey. *European Journal of Health Economics* 10:103-109.

Sarin RK. 1982. Strength of Preference and Risky Choice. *Operations Research* 30:982-997.

Sendi P, Brouwer WBF. 2005. Is silence golden? A test of the incorporation of the effects of ill-health on income and leisure in health state valuations. Health Economics. 14:643–647.

Smith RD, Sachs TH. 2010. Contingent valuation: what needs to be done? *Health Economics, Policy and Law* 5:

Somanathan E. 2006. Valuing Lives Equally: Distributional Weights for Welfare Analysis. *Economics Letters* 90:122–125.

Tessier P. 2011. The conditional QALY model: a proposal to include non-health related dimensions of quality of life in cost-utility analysis. *Working paper*.

Tessier P, Sultan-Taïeb H, Barnay T. 2011. Productivity costs, worker replacement and cost-benefit analysis of life-saving healthcare programs - a precautionary note, *submitted paper*.

Varian H.R. 1993. Intermediate Microeconomics. New York, Norton.

Viscusi WK, Aldy J. 2003. The value of a statistical life: a critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*. 27:5–76

Wagstaff A. 1991. QALYs and the Equity–Efficiency Trade-off. *Journal of Health Economics* 10:21–41.

Weinstein MC. 1995. From Cost-Effectiveness Ratios to Resource Allocation: Where to Draw the Line ?, in Sloan FA (ed.). Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies, Cambridge University Press, New York: 77-98.

Weinstein MC, Manning W. 1997. Theoretical Issues in Cost–effectiveness Analysis. *Journal of Health Economics* 16:121–128.

Weinstein MC, Stason WB. 1977. Foundations of Cost-Effectiveness Analysis for Health and Medical Pratices. *New England Journal of Medicine* 296:716-721.

Wriglesworth JL, Gravelle HSE. 1987. The Three Consumer Surpluses as Individual Welfare Measures. *Scottish Journal of Political Economy* 34:230-48.

Zerbe RO, Bellas AS. 2006. *A Primer for Benefit-Cost Analysis*, Edward Elgar, Cheltenham, UK, and Northampton, MA.