

Going back to our roots: a case for the welfarist approach?

Sabina Sanghera^a, Emma Frew^a, Joe Kai^b, Janesh Gupta^a, Tracy Roberts^a

^a *University of Birmingham, UK*

^b *University of Nottingham, UK*

Correspondence to: Sabina Sanghera, Health Economics Unit, Public Health Building, University of Birmingham, B15 2TT, UK. Tel: +44 (0) 121 414 8176. Email: s.sanghera@bham.ac.uk

Paper no: A068

Abstract

The National Institute for Health and Care Excellence (NICE) make recommendations using the extra-welfarist framework with its focus entirely on health and aim to maximise health. It is increasingly recognised that this school of thought is not suitable for certain conditions where benefits can additionally lie beyond health. It is the extent to which extra-welfarist measures do not capture these aspects that is important and with the remit of decision-makers broadening to include health and care excellence, this issue of how to incorporate benefits beyond health must be resolved.

In menorrhagia (or heavy menstrual bleeding), an assessment of quality of life (QoL) is the primary indicator of treatment success so it is particularly important to ensure that a suitable measure that accurately captures women's concerns and experiences is used. Menorrhagia is an interesting condition for two reasons; first the condition is chronic but symptoms occur in episodes and second, the condition is known to greatly impact on non-health aspects as well as health aspects of women's lives. These are two properties that can make valuing outcomes in menorrhagia more problematic than other conditions.

This paper will present the findings from recent research exploring outcomes in menorrhagia from two theoretical frameworks: 'welfarism' and 'extra-welfarism'. We will show how the extra-welfarist measures, EQ-5D and SF-6D produce results that imply a different treatment decision, followed by evidence of the use of willingness-to-pay (WTP) which show it to be both feasible and acceptable to the patient population. Discussion will focus on the implications of these findings particularly with respect to decision-making. We will provide a critical review of both approaches within the context of interventions to treat menorrhagia. Practice is unlikely to change instantaneously but the gradual accumulation of evidence against the use of current extra-welfarist measures for valuing outcomes for women experiencing menorrhagia may encourage decision-makers to consider alternative techniques. This paper will develop a case for recommending that policy-makers should consider alternative approaches and provide recommendations for what future research needs to focus on.

Introduction

Decision-makers such as the National Institute for Health and Care Excellence (NICE) make recommendations for clinical practice based on the extra-welfarist framework which focuses on health related quality of life (QoL) and aims to maximise health. Extra-welfarist measures such as, EQ-5D and SF-6D, are recommended to be used to value outcomes across all types of conditions. These extra-welfarist measures are used to generate the quality-adjusted life year (QALY) outcome, for use in economic evaluations, enabling comparisons to be made across conditions to aid decisions about resource allocation. However, in certain conditions, this extra-welfarist framework along with its currently recommended measures may not be entirely suitable. One such condition is heavy menstrual bleeding, clinically termed menorrhagia. Menorrhagia is defined as “excessive menstrual blood loss which interferes with the woman’s social, emotional, physical and material quality-of-life” (NICE 2007). In menorrhagia an assessment of QoL is the primary clinical and economic indicator of treatment success so it is particularly important to ensure that a suitable measure that accurately captures women’s concerns and experiences is used. Menorrhagia has two properties that can make valuing outcomes more problematic than others; first the condition is chronic but symptoms occur in episodes and second, the condition is known to greatly impact on non-health aspects, such as ‘practical difficulties’, ‘effect on family life/relationships’, as well as health aspects of women’s lives.

In a recent review of the literature, it was found that there is no consensus on the most appropriate measure to use (Sanghera et al, 2013). Extra-welfarist measures may be unsuitable for menorrhagia due to their narrow health-related focus as evidence suggests that women do not consider menorrhagia to be solely a health related condition. Impact on social life, daily activities and family life/relationships are important (Shaw et al., 1998). Therefore the use of these measures with a primary focus on health-related QoL is questionable in menorrhagia. These measures have poor face validity. Further, due to the chronic but episodic nature of the condition the standard recall periods referring to the past month (for SF-36) and health today (for EQ-5D) mean that the results are likely to be affected by the timing of assessment (Sculpher et al 1998). Women report that they are unsure how to complete the measures, whether it should be with reference to the condition or general health, as some consider the condition to be separate to health (Jenkinson et al 1996). Together, this evidence suggests that QALYs derived using these measures could be misleading.

Although not supported by decision-makers, an alternative *welfarist* framework can be used to value outcomes. This framework considers a broader assessment of wellbeing, which may prove to be potentially beneficial in conditions such as menorrhagia. There are several methods to elicit outcomes using the welfarist framework, but the most commonly used is the willingness-to-pay (WTP) contingent valuation approach. Based on the notion of sacrifice, respondent's value (utility), or strength of preference for a treatment is derived by determining how much they would be willing-to-pay for a treatment. The greater the WTP, the greater the utility gain or strength of preference for the treatment.

This paper reports on a 3-year programme of research designed to explore outcome measurement from the two perspectives: extra-welfarist and welfarist, within menorrhagia. The research was conducted within a large HTA-funded trial (ECLIPSE) that addressed the treatment question of which pharmaceutical treatment should be the first line treatment for menorrhagia. The two treatments that were assessed were 'Mirena' versus 'usual medical treatment'. Mirena is an intrauterine device, while usual medical treatment can include one of the following: tranexamic acid, mefenamic acid, norethisterone, depo-provera, or combined oral contraceptives. Hence the alternative theoretical frameworks were explored by valuing outcomes and conducting economic evaluations comparing the pharmaceutical treatments, Mirena and usual medical treatment.

The objective of this paper is to:

- Collate the findings from recent research carried out on this subject;
- Reflect on the welfarist and extra-welfarist approaches to valuing outcomes in menorrhagia and;
- Discuss the implication of our findings for policy-makers.

Exploring the use of extra-welfarist measures in menorrhagia

To address the question of cost-effectiveness using the conventional extra-welfarist approach, outcomes were measured using first, the EQ-5D-3L, and then re-estimated using SF-6D in a cost-utility analysis (CUA). The results showed that the most cost-effective treatment differed depending on the outcome measure used and all sensitivity analyses supported the findings of the base case analyses (Sanghera et al, forthcoming). The results based on EQ-5D showed

Mirena to be the most cost-effective treatment, while in contrast the result based on SF-6D showed usual treatment to be most cost-effective. The recommendation, regarding cost-effectiveness, to decision-makers was therefore shown to differ depending on the choice of extra-welfarist instrument. Therefore, these measures were shown to capture different aspects of QoL and these findings highlighted how sensitive the results were to the choice of measure. As others have reported (Sach et al 2009; Davis et al 2012) the choice of an alternative extra-welfarist measure to value utilities has a considerable impact on the cost-effectiveness results and in turn on the decision recommendation.

These findings coupled with those of the systematic review (Sanghera et al, 2013), illustrate that there are concerns around the use of extra-welfarist measures in measuring outcomes for women suffering with menorrhagia which therefore suggests a case for the exploration of a measure from the alternative welfarist framework.

The use of a welfarist measure - Willingness-to-pay

WTP values for Mirena and usual medical treatment (Sanghera et al, forthcoming) were elicited from the ex-ante perspective. The WTP values therefore were derived prior to the change in outcome occurring, from respondents who are at-risk of the condition, or at-risk of requiring treatment. Hence *expected* utility for the change in outcome was elicited. Further, WTP was also elicited from the alternative ex-post perspective, where WTP was elicited after the change in utility from respondents who had experienced the treatments. Thus, here, WTP for the *actual* utility change was elicited. This was with the intention of drawing comparisons between the two perspectives (ex-ante and ex-post).

A self-complete questionnaire was used to elicit maximum WTP as a monthly out of pocket payment up until menopause using a payment scale format, where respondents were presented with a vertical list of monetary values and asked to circle the maximum amount they would be willing to pay for treatment.

In the ex-ante perspective, since women did not have experience of menorrhagia they were presented with a scenario of menorrhagia and scenarios of the expected outcomes and process of care associated with each treatment, Mirena and usual treatment. The scenarios were derived on the basis of the domains contained within a condition-specific measure (MMAS) which was collected during the ECLIPSE trial. MMAS captures information on both the

health and non-health aspects of QoL that are affected by menorrhagia. The scenario for menorrhagia was based on the baseline MMAS values and the treatment scenarios were based on the 6 month MMAS values. Women were asked to provide a WTP value for both Mirena and usual medical treatment. With the ex-post perspective, WTP was elicited for *current* treatment from women who were enrolled in the ECLIPSE trial. Since these women had experience of menorrhagia and its treatment, either Mirena or usual treatment, it was not necessary to present them with scenarios as they were valuing their *actual* change in utility.

With the ex-ante perspective, the difference between mean WTP values for treatments were very small, but usual treatment had a slightly higher mean WTP value. In contrast, when WTP was elicited from the ex-post perspective, the difference between treatments was greater and Mirena had the greatest WTP value. Hence the treatment with the greatest mean WTP differed between perspectives (Sanghera et al, forthcoming). It was shown that respondents who provided WTP values were completing the questionnaire as is expected and required by analysts. That is, they were shown to be focussing on the effects of treatment and the value of the treatments to themselves, providing support for the use of WTP as a suitable measure of outcome for menorrhagia. WTP was shown to be acceptable to respondents as very few non-responses were observed, specifically for the WTP question, and very few protest responses were observed. However, there was not a strict monotonic relationship between mean WTP and income.

As the ex-post perspective involved women who were part of the ECLIPSE trial, we had available baseline EQ-5D-3L and the disease-specific MMAS scores allowing us to assess the association between the change in these measures from baseline, and ex-post WTP. Our a-priori assumption was that the greater the improvement in QoL (measured by MMAS and EQ-5D-3L), the greater the WTP value. It was found that there is a significant correlation between ex-post WTP and change in disease-specific MMAS (Sanghera et al, forthcoming). However, a significant correlation was not observed between MMAS and EQ-5D, which would suggest that WTP is detecting certain effects, most likely the non-health effects, which EQ-5D does not capture.

To draw a comparison between the welfarist and extra-welfarist approaches to valuing outcomes in menorrhagia, the WTP values were used to conduct a cost-benefit analysis (CBA). The treatment recommendation from the CBA can then be compared to the CUA.

Findings from the Cost-Benefit Analysis

The ex-ante WTP values, described previously, were used in a CBA comparing Mirena and usual medical treatment for menorrhagia (Sanghera et al, forthcoming). To enable comparisons between the treatment recommendations from the welfarist and extra-welfarist perspectives, the CBA was carried out using similar methods to that of the CUA. That is, the CBA had the same time horizon, discount rate as the CUA and also had an NHS perspective for costs. The limitations of using an NHS perspective for costs in a CBA were realised, but this method enabled comparability. The ex-ante WTP values were used, rather than the ex-post values, because it is theoretically preferred and because decision-makers recommend societal values be used in a publicly funded health care system, as all of society contributes to funding healthcare and therefore all views should be considered. For these reasons, the findings from the ex-ante CBA would be likely to be more acceptable to decision-makers than those of the ex-post perspective. In the CBA, with both the base case and the sensitivity analysis, it was found that usual treatment was the most cost-beneficial intervention as it was both less costly and resulted in a greater WTP value, hence produced the greatest welfare gain. The results of the CBA therefore recommended usual treatment as the first line treatment for women with menorrhagia. Table 1 provides a summary of the recommendations for clinical practice from each of the measures.

Table 1 Overall decisions from the welfarist and extra-welfarist measures

	EQ-5D (LNG-IUS v Usual treatment)	SF-6D (LNG-IUS v Usual treatment)	WTP (ex-ante)
Overall	ICER: 1600	Dominates	INB: -110
Cost-effective treatment	LNG-IUS	Usual treatment	Usual treatment

LNG-IUS; levonorgestrel-releasing intrauterine system, ICER; incremental cost-effectiveness ratio; WTP; willingness-to-pay

Reflections on welfarist and extra-welfarist measures in menorrhagia

To summarise, in the context of menorrhagia, it has been shown that;

1. There are concerns around the use of the extra-welfarist measures EQ-5D and SF-6D

- a. The results are likely to be affected by the timing of assessment due to the different recall periods used in these measures and the episodic nature of the symptoms
 - b. Women do not consider menorrhagia to be solely a health related condition, so it must be questioned whether the extra-welfarist health-related QoL measures are entirely suitable
 - c. EQ-5D and SF-6D produce different recommendations, which has significant implications for which treatment should be recommended in clinical practice
2. The findings from the welfarist approach demonstrate that it is feasible to elicit WTP and conduct a CBA within this condition. However, further research must be carried out to validate the findings from this feasibility study.

Previous studies have similarly observed that SF-6D and EQ-5D generate different utilities for the same patient in seven other conditions (Brazier et al., 2004; Whitehurst & Bryan, 2011; Whitehurst et al., 2011). Brazier et al (2004) also observed these differences and found that one of the reasons is likely to be due to the ceiling effects associated with EQ-5D, as SF-6D is able to discriminate between health at the top end of the utility scale, which is anchored by 0 and 1. It has been suggested that when full health is observed on EQ-5D, but not SF-6D, it is likely that the dimensions affected are those that are not comprehensively captured by EQ-5D such as ‘vitality’, ‘mental health’ and ‘physical functioning’ (Brazier et al., 2004). We believe that these additional SF-6D attributes are more likely to be those impacted by menorrhagia. Whilst the differences observed in our research are similar, there are likely to be additional specific reasons associated with the nature of menorrhagia, such as the measures capturing different aspects of QoL, and the different recall periods presented in the questionnaires. SF-6D refers to health ‘during the past month’ which is likely to lead to an underestimation or overestimation of the QALY. Similarly, EQ-5D could also lead to an under or overestimation of the QALY due to the recall period used of ‘health today’. For example, if the woman is experiencing the symptoms on the day she completes EQ-5D, the QALY gain from treatment is likely to be overestimated, and underestimated if she completes the EQ-5D questionnaire when she is not experiencing symptoms. This is due to the calculation of the QALY, i.e. the linear interpolation of the value and the use of area under the curve to generate the QALYs.

Evidence suggests that the combined focus on health-related QoL and the specified time-frame used in extra-welfarist measures poses difficulties for completion and could result in misleading decisions about which treatment is most cost-effective (Sanghera et al, 2013). It has been argued that these measures do not capture what is important to women with menorrhagia and that the questions are presented in a manner that women find difficult to comprehend given the chronic, but episodic nature of the condition (Sanghera et al, 2013). However, the primary advantage of the use of extra-welfarist measures in menorrhagia is that they are currently recommended to generate QALYs. A QALY is argued to be the most suitable outcome due to its focus on health related QoL, which is considered most appropriate for decision-makers to allocate healthcare resources (NICE, 2008). Decision-makers such as NICE endorse the extra-welfarist school of thought with its focus entirely on health and its aim to maximise health. It is increasingly recognised that this school of thought, which focuses entirely on health, is not suitable for certain conditions where benefits can additionally lie beyond health (McIntosh et al., 2010). However, it is important to state here that this paper challenges the use of current extra-welfarist measures EQ-5D and SF-6D and not all extra-welfarist multi-attribute utility instruments. Furthermore, EQ-5D and SF-6D are the only recommended extra-welfarist measures in the UK to have a value set based on the UK adult population.

The advantage of the use of WTP is that it is based on economic theory and captures a broader assessment of wellbeing than extra-welfarist measures (McIntosh et al, 2010). In menorrhagia it is known that benefits beyond health are important to sufferers and that these benefits are unlikely to be captured using the current extra-welfarist measures (Shaw et al., 1998). This would therefore have a negative effect on policy implications for healthcare resource use as treatments may not be shown to be cost-effective and would not be recommended for clinical practice because an inappropriate measure was used to assess the benefits. Evidence for the use of WTP in menorrhagia has been demonstrated and has shown that WTP is capturing aspects of life that are important to women, most likely the non-health benefits, that extra-welfarist measures do not capture. WTP values from both the ex-ante and the ex-post perspective have been elicited and respondents were shown to focus on the value of the treatments, as is required. Very few protest responses were observed and a low non-completion rate provides an indication of the acceptability of the method from both perspectives. The potential for the use of WTP in menorrhagia has been demonstrated but the

reasons behind the WTP values and the respondents understanding of the exercise would need to be verified with the use of in-depth interviews.

The way in which WTP should be used is open to debate. In keeping with theory and decision-makers recommendations for the use of societal values, the ex-ante perspective WTP should be applied (McIntosh et al., 2010; NICE, 2013). A further potential benefit of using the ex-ante perspective WTP could be its possible ability to overcome the issue of timing of assessment. Since these women are not experiencing the condition, the results are unlikely to be affected by the timing of assessment, which is the case when patients complete the extra-welfarist measures.

Despite these positive findings, there are obstacles against the use of WTP. It is clear that the use of WTP does not currently 'fit' within the guidelines provided by decision-makers (NICE, 2013). A wide range of alternative extra-welfarist techniques are recommended prior to the consideration of any other information. Other methods are suggested in the guidelines but can only be presented as additional to information using the extra-welfarist approach. Decision-makers' aversion to the welfarist method is due to the perception that WTP is based on ability to pay alongside on-going methodological issues (NICE, 2004). Practice is unlikely to change instantaneously but the gradual accumulation of evidence against the use of extra-welfarist measures in menorrhagia may encourage decision-makers to consider alternative techniques.

In relation to the findings of the association of measures, it is not surprising that WTP and the change in the disease-specific measure MMAS from baseline are correlated. WTP is elicited specifically for the treatment of the condition so it will always inherently be more sensitive than EQ-5D and SF-6D, as these extra-welfarist measures are designed to capture general aspects of health. However, it is interesting that change in EQ-5D and change in MMAS did not have a significant correlation (Sanghera et al, forthcoming). This lack of correlation suggests that the changes that occur on the disease-specific measure are not comprehensively captured by EQ-5D. This finding could be excused because these extra-welfarist measures are not designed to capture disease-specific health. However, as mentioned previously, the issue with the extra-welfarist measures arises because menorrhagia is not primarily a health related condition, but healthcare resources are used to treat the condition. Therefore, extra-welfarist measures are not comprehensively valuing outcomes in menorrhagia because they do not sufficiently capture the impact of the condition. It is the extent to which these

measures do not capture these aspects that is important and with the remit of decision-makers broadening to include health and care excellence, this issue of how to incorporate benefits beyond health must be resolved.

A second issue is the episodic nature of the condition and the timing of assessment when attempting to value outcomes in such conditions. Sculpher et al (1996) have previously argued that it is difficult to determine when the correct time would be to elicit values. The resolution to this problem is to identify which measure most comprehensively assesses values. Due to its recall period, the EQ-5D will be the most influenced by the timing of assessment. The SF-6D, because it has a wider recall period, is less likely to be affected, and the WTP measure avoids issues of timing. Perhaps the solution is to provide decision makers with information from several measures to ensure that the most comprehensive information is provided.

The EQ-5D produced a greater difference in QALYs between the two interventions, in favour of Mirena, and the SF-6D and WTP did not support this recommendation (Sanghera et al, forthcoming). SF-6D and WTP suggested indifference between the treatments that was slightly more in favour of usual treatment. Despite the EQ-5D producing a result which was in agreement with the disease-specific measure (Mirena more effective), the EQ-5D and the disease-specific measure were not significantly correlated (Sanghera et al, forthcoming). With menorrhagia, there may be scope for preference based treatment allocation as originally discussed by Sculpher (1998) and most recently mentioned by Roberts et al (2011). Sculpher suggests a preference based treatment system in menorrhagia rather than assuming the 'all or nothing' approach to decision-making. Both usual treatment and Mirena are already, and will be, used in clinical practice because Mirena provides contraception, which is not required by all women. Furthermore, the randomised allocation of the ECLIPSE trial, and clinical practice, is based on whether women have a preference for contraception. This coupled with the relative indifference between treatment effectiveness may mean that there is scope for the use of preference-based treatment allocation in menorrhagia.

In summary, information provided by the three measures EQ-5D, SF-6D and WTP are providing decision-makers with different information and therefore our findings suggest that they should not be considered individually but rather altogether. The differences observed in the results are likely to be caused by a number of factors, as all three measures have different elicitation methods, each measure captures different aspects of QoL and each measure refers

to different recall periods. It is difficult to directly compare the measures for these reasons, although comparisons are often drawn between EQ-5D and SF-6D because they are used to generate the same QALY outcome. Therefore, the most suitable method of assessing the measures is by comparing the treatment recommendations and attempting to identify the potential reasons for these differences. Based on the current evidence it is difficult to categorically say which measure is more suitable than the other, therefore the information from the measures should be considered altogether. WTP should be considered as a complement to the extra-welfarist measures, it cannot be recommended as the sole measure in menorrhagia because there is not sufficient evidence to support this. However, a clear case has been presented for the use of WTP in menorrhagia because it more comprehensively overcomes the issues associated with extra-welfarist measures.

Implications for Policy

Current recommendations from the NICE guidelines state that Mirena should be implemented as the first line treatment for menorrhagia (NICE, 2007). The cost-effectiveness results of the economic evaluation carried out alongside the ECLIPSE trial using the decision-maker recommended EQ-5D correspond with the NICE guideline recommendations, as Mirena was shown to be cost-effective. However, the recommendations from the NICE guidelines were not supported when SF-6D was used to generate QALYs (Sanghera et al, forthcoming). Hence the implications for clinical practice are dependent on the measure used, clearly indicating a need to ensure that the most suitable measure is used. It has been shown in the literature that neither measure is considered to be entirely suitable for menorrhagia. In fact there was strong support from the clinicians involved with the ECLIPSE trial to explore the use of WTP in menorrhagia because the shortcomings of the extra-welfarist measures are widely recognised. This evidence suggests that there is a need for policy-makers to consider the findings from measures other than EQ-5D and SF-6D when evaluating interventions for menorrhagia. The potential for the use of WTP in menorrhagia has been demonstrated, though further research is required. NICE recommend that measures other than EQ-5D should not be used without justification (NICE, 2008). Below we present a strong justification for the use of other measures in menorrhagia:

1. Evidence of the limited psychometric properties and use of EQ-5D and SF-6D in menorrhagia have been presented in our systematic review of outcome measures in

menorrhagia, suggesting that neither measure is able to fully capture the concerns and experiences of women with menorrhagia

2. EQ-5D and SF-6D have provided treatment recommendations in an economic evaluation alongside a trial.
3. The use of an alternative generic multi-attribute utility instrument would not be suitable given that a value set using a UK population has not been generated. Furthermore, a condition-specific multi-attribute utility instrument cannot be used as one has not been developed for menorrhagia.
4. Mapping from the condition-specific measure is unsuitable because the recommended measure that the condition-specific measure should be mapped on to is EQ-5D, which has been shown to have limited use in menorrhagia.

There is a need for policy-makers to develop a standard framework for evaluating conditions where the benefits can lie beyond health, as EQ-5D and SF-6D are known to be problematic in these cases. The use of WTP in menorrhagia has been demonstrated. However, whilst our research shows the feasibility of using WTP in menorrhagia, further research does need to be carried out within the area. Until additional evidence in support of WTP is available, it is recommended that a cost-consequence analysis be used as each measure is providing additional information that should be taken into consideration. In this case, decision-makers are presented with all of the relevant information that is currently available, thereby enabling decision-makers to make decisions about resource allocation in menorrhagia with a more informed opinion.

Further Research Recommendations

Further research is required. Whilst the feasibility of using WTP in menorrhagia has been shown, it is recognised that subsequent studies with larger sample sizes will need to be carried out to validate the findings. Additionally, it may be beneficial to carry out some in-depth interviews to determine the respondents understanding of the elicitation exercise. There are also still some methodological issues that need to be resolved before WTP will be

implemented on a wider scale. The research around the methodology of WTP has shown that the WTP values are sensitive to the elicitation format and also to the choice of administrative mode, e.g. interview or questionnaire. Further, the sensitivity of WTP to the size of the good has also been questioned. Whilst sensitivity was shown in our research, the issue needs to be assessed on a wider scale. Even the method by which WTP should be assessed is debated i.e. whether it be through psychometric testing or whether theoretical validation is more important (O'Brien & Gafni, 1996). It is agreed amongst health economists that these differences exist but a framework has yet to be produced that outlines how to move forward within the area. For the use of WTP to progress in menorrhagia, and in other conditions, there first has to be a set of guidelines or recommendations for how contingent valuation techniques should be used in health care, which analysts should follow. Without such guidelines, it will be difficult to convince decision-makers of the merits of WTP due to the methodological differences observed in studies.

An alternative welfarist method, which is still in its infancy in health economics, could also be considered for chronic conditions with episodic symptoms. The stated preference discrete choice experiment is gaining increased credibility for valuing outcomes. It appears to have so far received a more welcoming reception from health economists to the conventional contingent valuation method as WTP values can be elicited indirectly for changes in separate and in groups of attributes (Ryan & Gerrard, 2003). It has been argued to be able to overcome some of the shortfalls of direct elicitation of WTP (Hanley et al., 2001). However, methodological work does need to be carried out to establish this.

Next steps for extra-welfarist measures include research into identifying the drivers behind the differences between the findings of SF-6D and EQ-5D within these types of conditions. In this specific case study more in-depth analyses could be carried out to assess the sensitivity to change of EQ-5D and SF-6D compared to the disease-specific measures MMAS. It would also be interesting to identify which domains on SF-6D are affected by the condition compared to the domains affected by EQ-5D by using the changes in response to changes in MMAS. Similarly, it will be important to assess whether a meaningful change in MMAS is reflected in both the extra-welfarist measures and whether this change is matched by the respondents' report of general health.

Further, as the EQ-5D-5L has now been developed (Herdman et al., 2011) to help overcome the issues with ceiling effects of the EQ-5D-3L, there may be scope for testing this measure

against WTP, though the results are likely to be similar to those found here as the recall period remains the same and the non-health benefits are not captured.

References

Brazier, J., Roberts, J., Tsuchiya, A. et al. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. **Health Economics**, 13 (9): 873-884.

Davis, J.C., Liu-Ambrose, T., Khan, K.M., et al. (2012) SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. **Osteoporosis International**, 23 (7): 1849-1857

Hanley, N., Mourato, S., and Wright, R.E. (2001). Choice modelling approaches: a superior alternative for environmental evaluation? **Journal of Economic Surveys**, 15 (3): 435-462

Herdman, M., Gudex, C., Lloyd, A., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). **Quality of Life Research**, 20 (10); 1727-1736

Jenkinson, C., Peto, V., and Coulter, A. (1996). Making sense of ambiguity: Evaluation of internal reliability and face validity of the SF 36 questionnaire in women presenting with menorrhagia. **Quality in Health Care**, 5 (1): 9-12.

McIntosh, E., Clarke, P.M., Frew, E.J., et al. (2010). **Applied methods of cost-benefit analysis in health care**. Oxford: Oxford University Press

National Institute of Health and Clinical Excellence (2004). **Guide to the methods of technology appraisal**. London: National Institute of Health and Clinical Excellence

National Institute of Health and Clinical Excellence (2008). **Guide to the methods of technology appraisal**. London: National Institute of Health and Clinical Excellence

O'Brien, B., & Gafni, A. (1996). When do the dollars make sense?: Toward a conceptual framework for contingent valuation studies in healthcare. **Medical Decision Making**, 16: 288-299

Roberts, T.E., Tsourapas, A., Middleton, L.J., et al (2011). Hysterectomy, endometrial ablation, and levonorgestrel releasing intrauterine system (Mirena) for treatment of heavy menstrual bleeding: cost effectiveness analysis. **British Medical Journal**; 342:d2202

Ryan, M., & Gerard, K. (2003). Using discrete choice experiments to value healthcare programmes: current practice and future research recommendations. **Applied Health Economics and Health Policy**, 2 (1): 55-64

Sach, T.H., Barton, G.R., Jenkinson, C., et al. (2009). Comparing cost-utility estimates: Does the choice of EQ-5D or SF-6D matter? **Medical Care**, 47 (8): 889-894

Sanghera S, Frew E, Kai J, Gupta J, Roberts TE (2013). An assessment of economic measures used in menorrhagia: a systematic review. *Social Science and Medicine* 98: 149-153

Sanghera S, Roberts TE, Barton P et al (forthcoming). Levonorgestrel-releasing intrauterine system vs. usual medical treatment for menorrhagia: An economic evaluation alongside a randomised controlled trial.

Sanghera S, Frew E, Kai J, Gupta J, Roberts TE (forthcoming). Cost-benefit analysis using willingness-to-pay to compare Mirena and usual medical treatment for menorrhagia.

Sculpher, M.J.; Dwyer, N.; Byford, S.; et al. (1996). Randomised trial comparing hysterectomy and transcervical endometrial resection: effect on health related quality of life and costs two years after surgery. **British Journal of Obstetrics and Gynaecology**, 103 (2): 142-149

Sculpher, M. (1998). A cost-utility analysis of abdominal hysterectomy versus transcervical endometrial resection for the surgical treatment of menorrhagia. **International Journal of Technology Assessment in Health Care**; 14: 302-319

Shaw, R.W., Brickley, M.R., Evans, L., et al. (1998). Perceptions of women on the impact of menorrhagia on their health using multi-attribute utility assessment. **British Journal of Obstetrics and Gynaecology**, 105 (11); 1155-1159

Whitehurst, D.G., Bryan, S., and Lewis, M. (2011) Systematic Review and Empirical Comparison of Contemporaneous EQ-5D and SF-6D Group Mean score. **Medical Decision Making**, 31 (6): E34-44

Whitehurst, D.G., & Bryan, S. (2011). Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? **Value in Health**, 14 (4): 531-538