NICE Clinical Guidelines – what is being maximised and how should it be measured?

Hannah-Rose Douglas PhD
Health Services Research Unit
Department of Public Health and Policy
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT

Tel 020 7580 9521

Email: Hannah-Rose.Douglas@lshtm.ac.uk

Presented at the Health Economics Study Group, June 30th -July 2nd 2004, Glasgow

Introduction

This paper arises out from challenges to health economists working on clinical guidelines and relate to the application of the economic evaluation methodology to clinical guidelines in general. It paper reflects on the NICE guideline development process, and considers how to promote economic ways of thinking alongside clinical priorities. A second set of questions arise from specific issues in maternal and child health (MCH). These address the problem of employing the accepted methods of presenting economic evidence in pregnancy, childbirth, contraception and fertility treatment where maximands other than individual's health gain are also central to decision-making. The paper reflects on the place of choice, individual rights and autonomy in MCH as endpoints for policy, and how to agree the basis of determining what should be offered, how much and to whom in the context of resource scarcity. It suggests that other social welfare maximising decision rules are being adopted in the clinical guideline process in MCH and asks what effect this has on the transparency of the process. This paper represents an early stage of inquiry into these issues and is meant to lead on to more focussed reading and development of ideas, particular in developing more transparent means of considering the trade-offs between health gain and other maximands in the guideline development process.

Background and structure of the paper

The clinical development programmed set up by NICE funds seven National Collaborating Centres each covering a clinical area (e.g. primary care, acute care, mental health). Each centre employs systematic reviewers and health economists to synthesise the best available evidence on a specific topic. It convenes guideline development groups (GDGs) which are made up of representatives from clinical specialties with an input into the care pathway, and patient advocates. The GDG members make recommendations on clinical practice based on the available evidence of clinical and cost-effectiveness, which may include some economic modelling where data is sparse or to explore specific scenarios and clinical pathways. This paper draws on the experience of working as a health economist member of GDGs convened by the National Collaborating Centre for Women and Children's Health. The paper starts by briefly outlining the development of clinical guidelines, and how they differ from the more established NICE technology appraisal process. It reflects on how the economic evaluation evidence has been presented in clinical guidelines in MCH and why CUA has been lacking in the final guideline documents. It draws on literature on social welfare functions and whether these provide helpful insights in particular areas of health care where the health of the individual is not the main endpoint. Finally it considers ways in which the NICE guideline development process could achieve greater transparency in making its recommendations.

Section 1 - The development of NICE clinical guidelines – some common issues

Alongside its programme of technology appraisal, NICE has begun a programme to develop clinical guidelines. These are designed to be authoritative sources of guidance on clinical and cost-effectiveness to the NHS. Clinical guidelines (CGs) examine the best use of existing interventions in order to improve clinical practice or increase access to health care. The process of topic selection is not that different from the technology appraisal (TA) programme. Topics are filtered through Department of Health and the National Assembly for Wales who refer the topics formally to NICE in 'waves'. Topics can be suggested by patient groups, manufactures and professional bodies. The selection teams are made up of health officials who choose topics against a set of selection criteria and taking into account the views of ministers (see table 1).

Table 1 Criteria for consideration for NICE clinical guideline development and technology appraisal

Topics:-
That add significant value by resolving uncertainties
That provide significant positive health benefits – potential to reduce morbidity
That contribute to existing government health policies
Where there is sufficient current evidence
Where there are significant impact on resources
That will solve unacceptable variation in practice

http://www.dh.gov.uk/assetRoot/04/05/51/94/04055194.pdf, published in 2002, accessed 19 May 2004

However, here are differences in approaches and final products between the TA and CG process which play a part in influencing the rôle of health economists in both programmes. Health economics and cost-utility analysis has become central to decision-making in the TA process, but not in clinical guidelines. Alan Williams recently referred to this difference in approach as an implicit separation of spheres of influence between clinicians and health economists in the NICE appraisal process: "The TAs have been captured by health economists, and the CGs by the clinicians, and consequently they manifest two very different cultures" (Williams 2004)

The differential influence of the health economist is reflected in the NICE technical manual for clinical guidelines. It should be made clear here that health economists (I among them) were involved in drafting this guidance. The rôle was seen as educational, encouraging the group to consider resource consequences, rather than explicitly making the health economic analysis central to decision-making. The role of the health economists is clearly different between the TAs and the CGs, as the contrasting text in the technical manuals for

developing TAs and CGs implies[1]. The contrast in type of economic input into the decision-making process is a consequence of the type of decisions that NICE is requiring the CG or TA development group to make. The focus of the CG is the pathway of care, how people access treatment, what should be offered at each stage, level of required training, aftercare. This detailed advice is not part of the TA process.

In the wider context of implementation, the differences between TAs and the CGs are even more marked. The CG makes a wide range of recommendations which, as NICE has clearly stated, "sit alongside, but do not replace, the knowledge and skills of experienced health professionals". http://www.nice.org.uk/ accessed May 19 2004). Furthermore, while health professionals are expected to take the Institute's guidance "fully into account" when exercising their clinical judgement, it is explicitly stated that this guidance does not override individual responsibility for making decisions appropriate to the circumstances of the individual patient. By contrast, the TA usually leads to one over-arching recommendation (e.g. to recommend or not to recommend a new technology) arising from the final appraisal determination document (FAD). The clinician's role and responsibility for decision-making is less emphasised in TA technical guidance. Since January 2002, the NHS has had a statutory obligation in England and Wales to provide funding for treatments and drugs recommended by the TA work programme with local commissioners responsible for making the funds available to support implementation.

Clearly, the TAs and CGs are separate entities with different functions. The existence of a reference case showing the economic evaluation methods (CUA) in TA is a development which has not been emphasised in CGs where no such reference case is available to be referred to. But without cost-utility analysis used uniformly in CGs and across different health care areas, the contribution of the economic analysis to decision-making and priority setting is significantly reduced. One of the reasons for the lack of consideration of QALYs in clinical guidelines may be the deficiency of published evidence (say from EQ-5D scores) on which to base economic models in the guidelines. Since CGs cover a far broader pathway of clinical care than the TA, there can be wide areas of clinical practice where there is published evidence at all. Where there are published studies, they may not include any valuation of health-related quality of life let alone a utility/ preference-based measure such as

1

(NICE Technical manual for clinical guidelines, chapter 8) "The health economist should encourage the GDG to consider the economic consequences of a guideline recommendation as well as the clinical implications….economic modelling may not be warranted where, for example, the clinical evidence is so uncertain that even a ball park figure cannot be estimated.
(NICE Technical manual for technology appraisals) Assessment normally has two mutually dependent components: a systematic review of the evidence and an economic evaluation (pp2)… The evidence requirements for the economic evaluation include the quantification of the effect of the technologies under comparison (pp 11-12). The Institute has developed a reference case that specifies the methods considered to be the most appropriate for TAs consistent with the objective of the NHS of maximising health gain (pp20) … The reason for failure to meet the reference case should be clearly specified and justified (pp21) The appraisal committee will then make a judgement regarding the weight it attaches to the results of such a non-reference case analysis (pp21)

EQ-5D. However, as suggested in his recent OHE lecture, outlining a clinical pathway and setting put the decision / chance nodes for the investigations and treatments covered by a guideline could incorporate at least some attempt to attach QALYs to final endpoints (Williams 2004). The presence in the GDG of clinical experts and patient representatives also provides an opportunity to explore the putative QALY weights attached to the relevant states of health. This could lead to some indicative analysis of the cost-per-QALY value of clinical pathways and decisions facing patients and clinicians. It is a technical problem that could be resolved by better (and more accessible) evidence which is feasible in the medium to longer term. (Although in presenting economic models that quantify QALYs gained for clinical pathways based on little or no research evidence, I am mindful of the maxim that it is better to be imprecisely right than precisely wrong!)

Leaving aside the theoretically 'solvable' problem of the lack of research evidence in some areas of health care, I turn to an additional problem of benefit estimation in a particular class of clinical guidelines whose aims are not dominated by the requirement to maximise human health. Apart from the aim of promoting equity of access to health care and fairness in resource allocation, the NHS was also conceived as having a role in supporting wider economic and social goals by maintaining a healthy workforce and enabling people to reach their full potential as human beings. There are also societal preferences for how NHS resources should be spent that do not relate only to health gain but to having a positive experience at the end of life and peaceful death, as inferred by the Older People's National Service Framework (2001) and stated in the Department of Health Cancer Plan (2001). There are also positive and negative externalities of health care consumption.

For interventions that have an impact on these wider goals, the CUA approach has its limitations due to the fact that these welfare benefits are not incorporated into the QALY paradigm. Since the main aim of the NHS is to maximise health within a given budget, this has not been seen as a serious impediment to progress in priority-setting using the QALY metric as the unit for priority-setting. However, in the context of a CG in MCH, there are particular challenges relating not only to the additional non-health-related benefits of consumption of healthcare, but also because there are important externalities the consumption of health care by a woman, namely the impact on the child[2].

[2] The reflections in the following section relate only the MCH guidelines because this is the area I have experience of. One of the areas for development in this paper would be to assess the common and specific challenges facing health economists working on other clinical guidelines whose maximands are also hard to conceptualise within a QALY framework, for example mental health, nursing and primary care.

Section 2 – QALYs and maternal and child health

Pregnancy and childbirth

In the area of MCH, the first clinical guidelines to be published were on antenatal care, fertility and caesarean section. Each of these areas of health care had specific characteristics that made it difficult to undertake a cost-utility analysis. Some observations about health care affecting pregnancy and childbirth will be considered first before addressing the specific issues related to interventions that affect fertility.

In all interventions that affect the mother and the baby *in utero* and after birth, the healthy mother and baby is the primary outcome. All recommendations for high quality antenatal, intrapartum and post-natal care promote the importance of the mother being able to exert her autonomy and choice as central to good quality care. Since there is no perceived need to show the association between maternal autonomy and consequent health gain, this aspect of good quality care reflects a societal (and historically-specific) preference for the woman to be at the centre of decision-making. This preference for woman-centred care is expressed in terms of her rights and not explicitly her health, or the health of her future child.

The second rather obvious point is that pregnancy and birth is an experience of two symbiotic entities, the mother and the (potential) baby. The consumption of health care involves trade-offs of risks and harms between the two, with a mother potentially choosing to increase risks and harms to herself in order to maximise the chance of a healthy baby. Therefore the additional benefits to the (potential) baby could also be benefits to the mother in terms of her satisfaction knowing she has maximised the chances of a healthy baby, or disbenefits in terms of compromising her own physical health. Whether the benefits to the foetus and life of the future child should be counted as part of a QALY analysis is not an issue that has been satisfactorily resolved. If benefits to the future child do <u>not</u> count (expressed in terms of increasing the chance of a healthy baby), then this implies that the value of antenatal care is only in terms of the mother's own health, which seems counterintuitive. If the benefits to the future infant do count, then this brings with it the hugely contentious moral and ethical debate about the disbenefits to the foetus of decisions taken by the mother, including termination, which NICE and the GDG members would keenly like to avoid having to address.

The third observation is that MCH interventions have an effect on the risks and harms of future pregnancies and the health of subsequent children. These future health benefits could be incorporated into a clinical pathway with QALYs as the unit of outcome. The structure of the pathway would be considerably more complex, and the requirement to populate this pathway with data that is more than mere guesswork from the GDG would be an additional burden on the CG process. These technical problems could be overcome with adequate support. The more serious difficulty is in justifying why the QALYs generated by the consumption of health care by a

pregnant women/ new mother should include benefits to future children (who may never be born) if the QALYs associated with the future life of the current foetus are not justified.

In previous NICE technology appraisal guidance, the measurement of the value of risks and harms to mother and child and future children have been only partially addressed. The NICE technology appraisal on anti-D prophylaxis for rhesus negative pregnant women had to address the problem of how to measure the health outcomes of future children. (http://www.nice.org.uk/pdf/prophylaxisHTAreport.pdf. accessed 19 May 2004) Due to the difficulties of estimating both the parental grief and fetal loss avoided as a result of prophylactic treatment, the economic analysis focused on future disability of the child avoided. Infants born with minor developmental problems were given a QALY weight of 0.8, assumed to last for the first ten years of life; a year of life for those with more severe neurological damage was assigned 0.4 QALYs. The economic analysis expressed the disbenefits due to parental grief and the need for a high level of intervention of another pregnancy as a threshold. It showed that if these disbenefits avoided by treatment were valued at 9 QALYs, then the intervention would be cost-effective at the £30,000 per QALY threshold. This approach avoided having to express a real value for the loss in such an emotive area.

QALYs, fertility treatment and contraception

The challenge of constructing a cost-utility analysis becomes more complex for interventions that control a woman's natural level of fertility, for example contraception and fertility treatment for sub-fertile couples.

As for pregnancy and childbirth, the question of whose QALYs should count needs to be considered. In the case of fertility treatment, the recipient of the outcome is not the mother or the child, but a couple who wish to be parents. This shifts the emphasis away from the future child and towards the woman and her current partner as opposed to her future child. The net value to the couple of a baby is extremely high even where there is physical and psychological stress involved in treatment (up to 40% of couples do not continue with IVF treatment after every cycle of IVF (NICE Fertility Guideline 2004). In the case of contraception, the purpose is to prevent the birth of unwanted children. The net value of a baby is perceived by the woman and by society as low. One consequence of this basic analysis is that the value of a child appears to depend on whether that child is "wanted" by the parents bringing them into the world.

These arguments might not be relevant since the fertility and contraception guidelines consider two different situations and are not connected with one another. For one guideline, the maximand can be live births gained; for the other, it can be live births prevented. A priority-setting exercise of MCH services would not set 'live births' as the proxy for health gain to be maximised. The unit of outcome in a fertility or contraception

guideline should not be the live birth itself but the impact of that life on people who already exist (parents, other family members and society). This would fit in with the extra-welfarist paradigm of QALYs and avoid difficult questions about the value of an unborn child. The question is whether this 'value' can be adequately reflected by improvements in health-related quality of life as measured by EQ-5D.

QALY measurement and fertility treatment

Sub-fertility clearly has an impact on health-related quality of life in terms of anxiety and/ or depression associated with the unfulfilled desire to experience parenthood. A positive outcome of fertility treatment is a healthy baby, or, failing this, the comfort in knowing that all that could be done was done even if a healthy infant does not result from treatment. However, the impact of IVF on anxiety and depression is not straightforward – some people could move on and lead full and healthy lives after IVF fails to result in a child; for some parents their health-related quality of life might be severely reduced by the same failure. Where IVF succeeds in achieving a live birth, some parents may experience reduced quality of life as a result of the experience of parenthood. The impact of sub-fertility and parenthood on mental health is highly subjective, context specific and not related to the severity of the clinical condition causing sub-fertility. The challenge in producing QALYs is to determine what state of health (represented by an QALY weighting, derived say from an EQ-5D score) sub-fertility would represent for different groups in the population. Furthermore, the time over which additional quality of life associated with successful and unsuccessful fertility treatment would last (which could be potentially over a life-time) would need to be determined. In the context of a clinical guideline where no such evidence of quality of life was searched for in the systematic reviewing process, this presents a technical challenge.

A theoretical exercise to do just this is presented here, using the cost-effectiveness data published in the Fertility guideline. For illustrative purposes only, it considers a best case, medium level and poor scenario. The QALY weightings derived from EQ-5D schedule are used in this exercise[3]. It is assumed that treatments for sub-fertility have an impact in the domain of anxiety and depression only. In the best case scenario, all couples achieving a live birth from IVF move from anxiety/ depression level 3 (acute problems) to level 1 (no problems). In this scenario, for those couples who do not achieve a live birth from IVF, it is assumed they still derive some benefit from having the opportunity to try for a baby, and they move from level 3 to level 2 (some problems). In the medium case scenario, couples who achieve a live birth move from level 3 to 1 and the couples who do not achieve a live birth remain do not have any benefit. In the worst case scenario explored in the data, couples who achieve a live birth did not have acute problems with anxiety and depression due to their sub-fertility, and so move only from level 2 to level 1.

[3] I am grateful to Paul Kind at CHE for making these formula available.

The calculations consider only the additional benefits accruing in the first year of treatment for want of any clear indication of how long benefit would last (a life time, until the child becomes an adult, or end after a tail off in the first few months) which could result in a gross distortion of the QALY benefits from IVF. It also considers the worst case scenario where …. The example is worked out in Appendix A.

The cost-per QALY analysis indicated that the cost per QALY for fertility treatment lies somewhere between £5,000 (best case) and £75,000 (worst case) with the medium case reporting a cost per QALY of around £23,000. This is a very wide margin that crosses the cost-effectiveness of around £30,000 (Devlin and Parkin 2004). This analysis did not take into account any future benefits (and disbenefits of parenthood) beyond the first year of birth associated with successful treatment. Without empirical evidence this is purely guesswork, without any merit in terms of aiding decision-making.

A more interesting question is whether this £5-£75K per QALY actually reflects societal preferences for fertility treatment. Is the change in anxiety and depression level a proxy for the satisfaction from achieving a pregnancy and live birth when faced with the possibility of infertility? Certainly, the additional 'satisfaction' achieved from preventing premature death or increasing quality of life is not taken into account in the measure of the QALY associated with other forms of health care so there is an argument that they should not be considered in fertility treatment. Representatives of those who face problems of sub-fertility have presented their case in terms of the 'rights' of sub-fertile couples to have a chance of experiencing parenthood, or in terms of "individuals' wish to exercise newly available choices in the begetting of children" (Evans 1996).

The NHS has now accepted this argument and the role of the clinical guideline was to recommend how much treatment to offer per couple and the age limit for treatment. The GDG required guidance as to where and how to set this limit. The published and unpublished data presented to the GDG was presented in terms of the live births achieved through fertility treatment as the main outcome (alongside the minimisation of ectopic pregnancies, miscarriages, multiple pregnancies and morbidity with treatment). These were the data the GDG weighed up in considering the evidence. A cost-utility analysis was not presented to the GDG. This was due to the problems outlined above about how to conceptualise and measure the QALY value of fertility treatment, and the lack of robust quality of life data obtained through the systematic review. Also the group had to make certain trade-offs between health gain and equity. They had to balance the number of live births achieved (i.e. offering treatment to younger women in sub-fertile couples) with the equity aim of offering a wide age-range of women the chance of treatment to satisfy the need for couples to feel that at least something was offered to them even if it did not achieve a live birth. These maximands do not fit easily into a single-outcome metric of analysis such as the QALY. At the risk of having no economic evidence in the guideline at all, a cost-

effectiveness analysis based only on live births (which was the only data available) was presented. This was *de facto* an analysis of the recommendation once it had been made, rather than an aid to the decision itself[4].

QALYs and contraception

NICE have embarked on a new guideline to consider the benefits, risks and cost-effectiveness of long-acting reversible contraceptives (LARCs), such as intrauterine devices and sub-dermal implants. The issues here are somewhat different from the challenges of the fertility guideline. The usual outcome of focus has been prevention of unwanted births. No contraceptive method is 100% reliable and, under conditions of typical rather than perfect use, unwanted pregnancies result[5]. Unwanted pregnancies can lead to spontaneous miscarriage, to termination of the pregnancy or to live birth. Live birth can be to a healthy or unhealthy infant. If an unwanted pregnancy leads to a live birth, there are associated costs and benefits to that live birth (as to any infant who is born). A high birth rate leads to welfare loss to society as a whole, to women in particular, and to the 'unwanted' children born as a result of no contraception or failed contraception (Pezzini et al 2004). Furthermore, contraceptive uptake leads to overall cost savings due to the prevention of terminations (of unwanted pregnancies) and prevention of the cost of pregnancy and birth (including the cost of caring for unhealthy infants). An increased uptake of contraception (of any kind) leads to cost savings to the public purse as well as more births prevented. Therefore all interventions are in the south-eastern quadrants of the cost-effectiveness plane compared with no contraception. However, the guideline considers the relative cost-effectiveness of a particular group of contraceptive methods, some of which may be more costly as well as more effective than other more common methods (such as the Pill) and therefore some decision rule is still needed to determine the threshold value of preventing an additional unwanted pregnancy.

The evidence is clear that the availability of contraception has had an impact on social welfare in terms of the economic and social burden of unwanted children and the (health) disbenefits of multiple childbirth and large family size. (Pezzini et al 2004). However, the marginal value of each additional pregnancy prevented is not as easily quantified. There is also difference between the costs and benefits of unwanted pregnancy and a mistimed pregnancy[6]. Hughes and McGuire conducted an analysis of the cost-impact of unwanted pregnancy from a societal perspective (in terms of overall savings to the public purse) but did not express the benefits of contraception in terms of individuals' health or welfare gain (Hughes and McGuire 1996). Considering that

[4] This is not, I believe, an unusual state of affairs, either for clinical guidelines or for technology appraisals

[5] Perfect use implies using a contraceptive method as described by the manufactures or guided by a clinician. Typical use usually implies a higher pregnancy rate as a result of not following manufacturer's advice exactly, e.g. not taking the Pill every day. LARCs are supposed to decrease the gap between perfect and typical use by not relying so much on the decisions and actions of the user.

[6] Poorly timed pregnancies do not result in additional costs if the parents are only changing the timing and/ or spacing of children unless the age of the mother has an effect on the costs of pregnancy and childbirth (if anything these would be lower for younger , healthier mothers), or on her ability to invest in her education and career prospects, or the likelihood of the father maintaining contact, and / or the emotional maturity to care for a child as a result of an earlier pregnancy – see Pezzini 2004

contraception is consumed by healthy adults who will remain physically healthy (except for in rare cases) throughout use, even if they become inadvertently pregnant (except for in rare cases), the impact on quality of life must again be explained in terms of anxiety and depression associated with the adverse outcome (pregnancy) if a CUA were undertaken.

However, since most economic studies on contraception do not express the benefits of contraception in terms of health gain (since it is one occasion when a cost-benefit analysis could be presented in a meaningful way in a health care context), it could seem unduly dogmatic to adhere to the principle of CUA for a contraception guideline. If an evaluation were to be undertaken, it would have to consider what the health impact to a couple of an unwanted child was, based on the fact that a pregnancy may start out as inadvertent, but later on in the pregnancy or in the life become wanted (that is, a net benefit to the parents). It would have to consider the side-effects of different forms of contraception which are valued differently by different women. For example, some of the sub-dermal implant contraceptives can lead to amenorrhea (absence of menstruation) which for some women is seen as a positive side-effect and to others a negative side-effect.

On the outcome side, women may have strong preferences for different forms of contraception based on the mode of delivery (e.g. oral contraception or sub-dermal implant or intrauterine device) which affect uptake and the rates of unwanted pregnancy following discontinuation where they are dissatisfied with their contraceptive method. Some methods are more suitable to different stages of life, to lifestyle choices or to specific sub-groups of the population (e.g. based on cultural acceptability). Therefore clinicians and policy-makers put a high value of increasing access and choice to contraception. This demand for a wide choice of contraception is not based on any systematic analysis of the cost of maintaining a wide range of choice (in terms of pregnancies prevented by providing more/less choice), but is perceived by policy-makers to reflect reality, and is a key part of the WHO strategy in reproductive health (WHO 1994)

If an economic analysis (CUA or CEA) were undertaken and found that contraceptive A was more cost-effective that contraceptive B, would that undermine the need for a wide choice to be available to maximise uptake? Does this go against the aim of the clinical guideline? How could the perceived need for wide choice be explored empirically in an ethical way? Perhaps it could be if a natural experiment were to present itself (i.e. by comparing geographical areas with different levels of choice of contraception available). However, the LARC clinical guideline is likely to reduce the inequity of access to contraception and increase choice for women, so such an experiment is likely to become less feasible in the future.

The need to provide wider access and increased choice in contraceptive services is expressed, like fertility treatment, and pregnancy and childbirth, in terms of the 'rights' or 'claims' of women to exercise control over their fertility and their choices. This is the theme running through all the MCH guidelines. It is relevant therefore to turn to some further discussion of the nature of these rights and claims and to consider whether social welfare maximising functions other than health gain/ equity are guiding the priorities in the MCH guideline development process. If so, this might help to explain the reason why the technical manual presents the value of health economic analysis (as conceptualised as CEA/CUA only) in circumspect and guarded terms.

Section 3. Which social welfare function is being maximised?

Recent health economic literature based on Sen's "capabilities" approach may provide a point of departure to think about what is being maximised in the clinical guidelines on fertility contraception. Sen argued that the appropriate focus for understanding well-being is on what people can do or can be (what they are capable of) rather than on what they have (what they actually do in terms of their functionings (Sen 1990). People's assets (endowments) have consequences for the claims they can make (entitlements) on goods and services. Entitlements are claims that are accepted by society as legitimate (not necessary moral, or formally guaranteed in law).

This leads to a framework for decision-making about public policy rooted in social justice, based on human need rather than the maximisation of utility (that is, desire or preference fulfilment as the primary maximand). This philosophical approach has been at the forefront of policy-making in international development; for example, the United Nations conceptualises human development as both "widening people's choices and the level of well-being they achieve, and has developed a Human Development Index constructed using the capabilities approach (UNDP 1997: 13-14).

In the international reproductive health literature, "legitimate" claims to resources are understood to be historically and culturally defined within multiple rule orders or plurality of norms available for claim-pressers to draw on to legitimate or redefine entitlements (Jackson 1998, in Locke 2001). While much of the reproductive health literature in the clinical domain has concentrated on health outcomes (maternal and infant morbidity and mortality), the reproductive rights/ international development literature has adopted the approach that individuals do not hold distinctive reproductive goal that are individualised and inherent, but that social norms and practices play a key part in shaping these decisions over time. Sen's capabilities approach has drawn attention to the extent to which people can make choices and achieve reproductive aspirations in relation to their reproductive endowments (Sen and Snow 1994). The capabilities framework draws attention to rights and claims, and moves the focus away from a narrow definition of health gain.

The capabilities approach is extra–welfarist rejects utilitarianism, but differs from the QALY approach which is based on consequentialism (measuring health gain as the outcome of policy), since it focuses on claims and rights which incorporates both deontological considerations (rights and duties) and procedural ones (justice), (Arnand 2002). Health gain *per se* has not been considered to be a useful maximand as the social choice rule of preference for health care as the overarching goal of rationing (Arnand ibid.). There has been much debate about whether other considerations such as equity can be incorporated into a QALY maximising framework. A sub-discipline of health economics has grown up to address this goal and to ensure that the decision-making is transparent. At the extreme of this thinking is the all-inclusive framework based on the Willingness-to-Pay of society per Incremental Expected QALY (WTPperIEQ) incorporating both socio-political decisions on the willingness to pay for the maximand, and the clinical decision incorporating individual's utilities and patients' specific probabilities of outcome (Dowie 2002). This literature is based on a consequentialist, decision-analytical framework which is committed to "pursuing the Holy Grail of the equitably efficient and transparently decidable use of public funds for health care (Dowie, ibid:93).

Is it possible, in the context of reproductive rights, to construct a transparent decision-making process, that can incorporate Sen's rights, duties, and procedural considerations into a coherent social welfare maximising function? Paul Arnand has used this approach and incorporated consequentialist, deontological and procedural "claims" into nine social welfare functions applied to the problem of rationing in 7 treatment areas that included in vitro fertilisation of sub-fertility (Anand, 2002). He showed that at least some IVF would be funded in all but one of the social choice rules (Surprisingly, this rule reflected the desire to maximise of treatment for active voters, reflecting the political influence of active interest groups). He concluded that "a normative rationale can be found for rationing most types of health care" (Anand, ibid:12)

The argument for the ubiquitous adoption of the QALY approach for NICE TAs and CGs is that QALYs provide a comparative unit of analysis across clinical specialties and interventions. It has also been argued recently that QALYs can be grounded in Sen's capability approach as measure of health related functioning and that since NICE aims to improve health rather than freedom to achieve health, this is an appropriate maximand. While this is true for a wide range of interventions and clinical areas examined by NICE, it is not true for reproductive health, and faces challenges (as outlined above) in determining whose health is to be maximised in MCH

The argument against the use of other social choice rules other that QALYs is the lack of transparency in decision-making that would result. The fertility guideline recommended that three cycles of IVF be offered on the NHS to sub-fertile couples. This was not based on cost-effectiveness analysis since there was no decision-

rule agreed by the group for deciding the maximum value or willingness to pay of a successful outcome (an infant, or reassurance). Further guidance issued by the Secretary of State for Health that a minimum of one cycle of IVF should be offered by April 2005 (with the aim of increasing this to three cycles in the future) was also not made on the basis of any transparent decision rule.

Alan Williams has suggested that NICE needs to consider the trade-off that would be acceptable between conflicting objectives and that "a compromise will have to be struck" (Williams 2004). A compromise is pareto sub-optimal, but since we have all but abandoned the pareto principle in rejecting utilitarianism in health care decision-making, perhaps a compromise that is transparent and therefore can be challenged by others, is the social choice rule that fits Sen's framework where social justice is done and seen to be done.

If the chance to experience parenthood and to experience reproductive freedom is perceived as a right due to the citizens of England and Wales, then the economic problem is to identify the maximum willingness to sacrifice health gain (assuming a finite NHS resource) to allow each additional individual to make their claim on reproductive heath services, or estimate the marginal rate of substitution between these maximands. This will almost certainly not be decided empirically, since it is a product of preferences formed by cultural and social norms that change over time since "entitlements must be understood as the outcome of an earlier period of claim-pressing" (Dasgupta 1990). The trade-off will have to be estimated for each TA or CG by a process of under and over estimating the MRS with each guideline published, getting closer to an equilibrium value. If this is the trade-off to be considered, the question is whether the clinical guideline group, as representatives of reproductive health clinicians and advocates, should be making these decisions on behalf of the general population. This is what they are doing implicitly, without other guidance from NICE as to what to do otherwise (where QALYs are not the maximand of the guideline process). Some health economists who have experience of clinical guideline development have suggested that the economic evidence should be considered at a different stage to the clinical evidence since it has a different remit (Wailoo et al 2004). But the use of economic modelling does not inspire confidence in the non-economic audience (Hutton and Maynard 2000, Smith 2002). Therefore there is a danger that the clinicians who are the erstwhile change leaders in the implementation of the guideline will not sign up to its recommendations (the GDG is made up of the 'change leaders' in each clinical area) which could threaten the legitimacy of the guideline and prevent it achieving influence.

Summary and tentative conclusions

This paper has set out some of the questions that have arisen from the experience of working in the NICE clinical guideline development programme, and considering why the relatively straightforward approach to presenting the economic evidence set out in the technical guidance for the health economists has presented particular headaches in MCH. I have attempted to set down what I think are the challenges in MCH, one of

which is deciding whose health is to be maximised (e.g. the mother, the future infant, future siblings). The impact of pregnancy, and consequent termination, miscarriage or parenthood has not been explored using a QALY approach empirically in the published literature, and attaching QALY estimates to the outcomes of antenatal care or caesarean section is not something that I felt we could do without some evidence from the published literature, if the maxim holds that it is better to be imprecisely right that precisely wrong.

In the case of contraception and infertility which are more recent guidelines under development, the comparison between the two outcomes - the birth of a child in one, the prevention of a child in the other - presented an interesting case study to examine what was being maximised, and how this outcome is valued differently depending on whether the child that would result from the pregnancy if it is not terminated was wanted or not by the couple. (This lead to some naval-gazing philosophical debates between us about whether the value of a person was dependent on whether they were 'wanted' by their parents or not at conception, and since 30-40% of pregnancies are unplanned in the UK, how many people now existed who were originally not wanted, etc).

The guideline development groups for fertility and LARC contraceptives appear to be making decisions based on maximands other that overall health gain, and use the language of rights and claims to resources in making their arguments. This led me to consider whether Sen's capability approach which incorporates social justice and procedural rules as part of the decision rules help to explain what other social welfare maximising functions are influencing decisions. Some developmental work has been undertaken to shows that under social decision rules that are based on Sen's capabilities approach, at least some IVF would be funded.

Where does that leave the NICE guideline development process? I suggest that the trade-off between health maximisation and other claims are historically determined (based on changes in values, in culture and on health care technology), and will change over time making it almost impossible to perfectly express the preferences of the citizens of England and Wales in terms of who should have preference in their claims over health care resources. The approaches that have been suggested from muddling through to the use of clinical decision trees incorporating the rather than sub-optimal statements are themselves a trade-off between transparency and the dynamic nature of reality. NICE could make it a prerequisite that GDGs consider these trade-offs explicitly in their decision-making – and to ask the group to include in the guideline and in their discussion the criteria by which they make their recommendations. These might be, for example that equity of access is a constraint on health maximisation (say in the case of who should receive IVF treatment). This is not currently a requirement of either the TA or the CG process. It would make the trade-offs that the panels have to address more explicit in their deliberations as a group, and in their recommendations. If these were more explicitly put into the guideline, the stakeholders and general public could challenge these assumptions. At present, the implicit nature

of these values makes the final recommendations obscure and consequently less likely to be implemented by those who disagree with them. The requirement should be to set aside a specific meeting of the GDG to consider the groups agreed values and preferences - what they are trying to maximise, and the trade-offs they are willing to make- and to consider the economic analysis presented in the guideline and how it affects their decision-making. This would be a positive starting point along the road towards transparency.

References

Anand P. (2003). The integration of claims to health-care: a programming approach. *Journal of Health Economics*, **22,** 731-45.

Dasgupta P. (1999) Well-being and the extent of its realisation in poor countries. *The Economic Journal* 100, Supplement 1-32

Devlin N. & Parkin D. (2004). Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics*, **13,** 437-52.

Dowie J. (1998). Towards the equitably efficient and transparently decidable use of public funds in the deep blue millennium. *Health Economics*, **7,** 93-103.

Evans D. (1996) Genuine clinical need: The clinical classification of infertility, in (Eds.) Evans D. *Creating the child: the ethics, law and practice of assisted procreation*. The Hague, Martinus Nijhoff Publishing

Hughes D. & McGuire A. (1996). The cost-effectiveness of family planning service provision. *J Public Health Medicine*, **18,** 189-96.

Hutton J. & Maynard A. (2000). A nice challenge for health economics. *Health Economics*, **9,** 89-93.

Locke K. (2001) Research Paper III. *Wellbeing and Reproductive Freedoms: Assessing Progress, Setting Agendas.* Research paper III. Overseas Development Group Research Series. School of Development Studies, University of East Anglia

NHS Executive (2000) Cancer Plan: A plan for investment, a plan for reform. London, Department of Health

NHS Executive (2001) The National Service Framework for Older People, London. Department of Health

Pezzini S. (2004) *The effect of women's rights on women's welfare: evidence from a natural experiment.* Paper presented to the Royal Economic Society's Annual Conference, April 5-7 2004

Sen A. (1985). *Commodities and capabilities*. Oxford University Press: Dehli, India.

Sen A & Snow RC (1994) *Power and decision: the social control of reproduction. Harvard series on population and international health*. Boston Massachusetts, Harvard University Press.

Smith R. New BMJ Policy on economic evaluations *British Medical Journal* 2002;325:1124

UNDP (United National Development Programme). *Human Development Report.* New York, Oxford University Press

Wailoo A, Roberts J, Brazier J. & McCabe C. (2004). Efficiency, equity, and NICE clinical guidelines. *British Medical Journal*, **328,** 536-7.

Williams A (2004a) *Making NICE even NICER.* 11th Annual Lecture of the Office of Health economics. Royal College of Physicians, 13th May 2004

WHO (1994) *Contraceptive introduction reconsidered: a review and conceptual framework*. World Health Organisation, Geneva.

Appendix A: Cost per QALY calculations for IVF treatment, based data published in the NICE Fertility Guideline (2004)

EQ-5D dimension	**Level 2**	**Level 3**
Full health	*1*	*1*
Mobility	-0.069	-0.314
Self care	-0.104	-0.214
Usual activity	-0.036	-0.094
Pain/ discomfort	-0.123	-0.386
Anxiety/ depression	-0.071	-0.236
	Constant for any level two or higher, subtract 0.081	Any level 3, subtract an additional 0.269

Full health	Level 1	Level 2	Level 3
Anxiety/ depression	0	0.071	0.236
Total QALYs per year		0.848	0.495
Change in QALYs by achieving level 1		0.152	0.505
Change in QALYs : level 3 to level 2			0.353

Best case scenario		
Couples		100
Live birth per IVF cycle		17.50%
Cost per cycle	£	2,000
Live birth per cycle		17.50
Live birth not achieved		82.50
QALYs/ couple with live birth		8.84
(assuming transfer from level 3 to level 1)		
QALYs/ couple with no live birth achieved		29.12
(assuming transfer from level 3 to level 2)		
Total QALYs		37.96
Cost per cycle	£	2,000
Total cost	£	200,000
Cost/QALY		**£ 5,269**

Medium scenario	
Couples	100
Live birth per cycle	17.50%
Cost per cycle	2,000.00
Live birth per cycle	17.5
No live birth per cycle	82.5
QALYs/ couple with live birth	8.8375
(assuming level 3 to level 1)	
QALYs/ couple with live birth	0
(assuming no change in level)	
Total QALYs	8.8375
Total cost	200,000
Cost/QALY	**£ 22,631**

Poor scenario	
Couples	100
Live birth per cycle	17.50%
Cost per cycle	2,000.00
Live birth per cycle	17.5
No live birth per cycle	82.5
QALYs/ couple with live birth	2.66
(assuming level 2 to level 1)	
QALYs/ couple with live birth	
(assuming no change in level)	
Total QALYs	2.66
Total cost	200,000
Cost/QALY	**£ 75,188**