

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

What counts and how to count it: Physicians' constructions of evidence in a disinvestment context

Katherine Hodgetts^a, Adam G. Elshaug^{a,b,c}, Janet E. Hiller^{a,d,*}

^a School of Population Health and Clinical Practice, The University of Adelaide, North Terrace, Adelaide 5005, SA, Australia

^b Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115, USA

^c The Commonwealth Fund, One East 75th St, New York, NY 10021, USA

^d Faculty of Health Sciences, Australian Catholic University, Room 2.78b, Level 2, 115 Victoria Parade, Fitzroy, VIC 3065, Australia

ARTICLE INFO

Article history:

Available online 27 August 2012

Keywords:

Australia
Evidence-based health policy
Deliberative methods
Discourse analysis
Disinvestment

ABSTRACT

Internationally, there is an increasing focus on quality and sustainability measures oriented to reducing inefficiencies in health provision. The use of assisted reproductive technologies (ART) for older women represents a case study in this area. This paper analyses the constructions of evidence brought to bear by ART physicians in the context of deliberative stakeholder engagements (held 2010) around options for restricting public subsidy of ART in Australia. Physicians participated in two deliberative engagements during which they were presented with results of a systematic review of ART effectiveness, as well as ethical and cost analyses. These sessions were part of a broader research program of engagements held with policymakers, community members and consumers. Physicians deliberated around the data presented with a view to formulating an informed contribution to policy. The ensuing discussions were transcribed and subject to discourse analysis. Physicians questioned the evidence presented on the grounds of 'currency', 'proximity', 'selectivity' and 'bias'. We outline physicians' accounts of *what should count* as evidence informing ART policy, and *how* this evidence should be counted. These accounts reflect implicit decisions around both the inclusion of evidence (selection) and the status it is accorded (evaluation). Our analysis suggests that participatory policy processes do not represent the simple task of assessing the quality/effectiveness of a given technology against self-evident criteria. Rather, these processes involve the negotiation of different orders of evidence (empirical, contextual and anecdotal), indicating a need for higher-level discussion around 'what counts and how to count it' when making disinvestment decisions.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

In recent years, health reform has increasingly emphasised measures oriented to supporting safe, effective and cost-effective care within sustainable health systems. Internationally, a growing demand for services, coupled with the development of costly new technologies, has seen healthcare priority-setting become a complex part of this landscape (Williams & Bryan, 2007). Given limited resources, it is inevitable that choices be made regarding 'coverage' of new technologies (deciding whether their cost will be subsidised via insurance), and 'disinvestment' from established

services deemed to deliver limited health gain for their cost (Mitton, Patten, Waldner, & Donaldson, 2003).

Mirroring the developing emphasis on 'evidence-based medicine', priority-setting processes have been largely informed by the guiding principles of 'evidence-based health policy' (Dobrow, Goel, & Upshur, 2004). As such they have foregrounded the use of systematic reviews and health technology assessments, argued to enable "clinical evidence to be presented as potential solutions to policy problems" (Williams & Bryan, 2007: p. 2117).

A shift towards stakeholder involvement and deliberation in decision-making has also been identified as a feature of modern health policymaking (Abelson, Giacomini, Lehoux, & Gauvin, 2007), and has served to inform priority-setting decisions (Russell & Greenhalgh, 2009). In line with the contemporary evidence focus, incorporating community and stakeholder perspectives in decision-making processes has been seen as a means of improving and legitimating policy directives (Davies, Wetherell, & Barnett, 2006).

* Corresponding author. Faculty of Health Sciences, Australian Catholic University, Room 2.78b, Level 2, 115 Victoria Parade, Fitzroy, VIC 3065, Australia. Tel.: +61 3 9953 3566; fax: +61 3 9953 3385.

E-mail address: Janet.Hiller@acu.edu.au (J.E. Hiller).

A move from representative to participatory forms of health policymaking is argued to increase public ownership of policy (Scutchfield, Ireson, & Hall, 2004), to address the 'democratic deficits' inherent in traditionally 'opaque' processes, and to render the provision of health services accountable to users as taxpayers, voters and consumers (Tritter & McCallum, 2006). It is also deemed to represent a particularly important means of addressing the inherently complex, ethical nature of priority-setting decisions (Lomas, Culyer, McCutcheon, McAuley, & Law, 2005).

Participatory and deliberative policy processes may serve to rebut a central criticism of the trend to evidence-based policy, namely that this movement depicts policymaking as a "science free from the inherent messiness of human language, of human interpretation, of human values" (Murray, Holmes, & Rail, 2008: p. 273). Given that priority-setting processes have been shown to require far more than cost-effectiveness data to result in locally implementable decisions (Jenkins & Barber, 2004), deliberative approaches have emerged as a means of incorporating a range of data sources (comparative effectiveness research, professional opinion, the colloquial evidence of stakeholders) into contextualised policy guidance (Lomas et al., 2005).

Exactly *how* clinical evidence and stakeholder perspectives come together to inform policy directives, however, has increasingly been the focus of qualitative research. Informed by an understanding of 'policy as discourse', these studies have shed light on the social processes by which health policy 'problems' and 'evidence-based solutions' are constituted in deliberative contexts (Russell & Greenhalgh, 2009). Rather than seeing deliberative engagement as a means of enhancing a pre-existing (usually clinical) evidence base with additional 'contextual data', studies within this field treat the notion of evidence *per se* as a contestable domain. In deliberative settings where evidence is understood to be 'talked into practice', analytic attention turns to the social process of evidence construction and its consequences for health policy.

The present study connects with this critical literature, offering an examination of the construction of evidence within a policy-oriented priority-setting deliberation. Specifically, it provides an analysis of stakeholders' (in this case, physicians') negotiation of 'what counts' as evidence within a disinvestment debate. In the context of deliberative discussions with medical physicians on the topic of (potential disinvestment from) public subsidy of Assisted Reproductive Technologies (ART) for older women, we analyse the strategies by which different orders of evidence were negotiated, appraised and incorporated into an informed contribution to policy. In addition, we illustrate the processes by which hierarchies of evidence, rather than being self-evident, were actively constructed in the course of this debate.

Disinvestment and the case of ART

The selection of ART as a case study for an analysis of disinvestment deliberations was made on a number of grounds. ARTs possess a number of defining characteristics that meet criteria for review under proposed disinvestment frameworks (Watt, Elshaug, Willis, & Hiller, 2011). For example, there exists substantial temporal and geographic variation in ART service provision, and differential effectiveness of treatment has been noted across patient subgroups. In particular, evidence suggests that increasing maternal age is associated with a decline in positive treatment outcomes, to the extent that a woman's age has been described as the "single most important factor in determining the success of ART" (Assisted Reproductive Technologies Review Committee, 2006: p. 67).

Australia, via its universal health insurance scheme (Medicare) provides one of the world's most generous ART subsidy

programmes. Yet its funding history suggests that a range of factors beyond evidence of treatment effectiveness has influenced the direction of relevant policy. For example, the "electorally significant" (Dill, 2006) force of pro-IVF lobby group opinion has been identified as a key factor in the failure of government-mooted subsidy restrictions, including those seeking to limit subsidies on the basis of maternal age (Watt et al., 2011). The complexity of the argumentative terrain around ART subsidy therefore suggested that deliberation around this topic would offer insights into the social process of evidence negotiation involved in the construction of disinvestment policy.

Physicians and 'what counts'

The analysis presented here is drawn from a broader study that engaged a range of stakeholders in deliberative discussions around the public subsidy of ART, focussing on maternal age and number of treatment cycles. Informed by the theory of deliberative democracy (Dryzek, 2000; Fishkin, Luskin, & Jowell, 2000), this research program aimed to generate an informed and representative contribution to policy directions around potential refinements to ART treatment subsidies. This paper analyses transcripts of the engagements with one particular stakeholder group: physicians who practise in the area of ART or in the related areas of general practice, perinatal care or maternal/foetal medicine.

Previous qualitative analyses of priority-setting deliberations suggest that clinical stakeholders play an important role in the construction and mobilisation of 'evidence' within coverage debates. As drivers of technology and treatment utilisation, physicians' involvement in priority-setting activities has been argued to be imperative (Mitton et al., 2003) and their perspectives afforded particular persuasive authority within decision-making processes (Jenkins & Barber, 2004). The finding that experiential evidence and personalised clinical anecdotes often serve to 'trump' published data in informing complex or controversial coverage decisions (Green, 2000) suggests that physicians' 'category entitlement' (Potter, 1996) to knowledge as medical experts renders their constructions of proof and evidence particularly influential. For example, in qualitative studies of healthcare rationing deliberations involving a range of stakeholders, Jenkins and Barber (2004) found that decisions were frequently guided by 'physician excitement', while Duthie, Trueman, Chancellor, and Liez (1999) found a repeated deference to doctors for a 'final judgement' on rationing directives.

In this study, we analyse the specific argumentative and rhetorical strategies mobilised by physicians as they negotiated constructions of evidence. In particular, we examine how physicians undermined systematic review evidence (presented by the research team as a means of informing disinvestment deliberation around the subsidy of ART) while bolstering the objectivity and facticity of the version they supplied in response.

Methods

Deliberative process

The first phase of the project entailed conducting a systematic review of assisted reproductive technologies (Watt et al., 2011). This mirrored elements of existing (Australian) Health Technology Assessment frameworks and policy processes, and was led by a member of the research team regarded as a pre-eminent expert in Australian HTA. The review was guided by a protocol appraised by content experts, and incorporated a range of evidence including effectiveness data as well as ethical and cost analyses (Carter & Braunack-Mayer, 2011; Griffiths et al., 2010).

In the second phase, the review findings were presented at a series of deliberative stakeholder engagements held in South Australia with groups of ART physicians, non-partisan citizens, ART consumers and State and Federal health policy advisors respectively. Each stakeholder group attended two engagement sessions (two 'rounds' of engagements spaced a number of weeks apart) in which they were encouraged to deliberate around the evidence, and formulate a perspective on their preferred approach to the provision and subsidy of ART in Australia.

A one-page outline of the data described above was circulated to all participants prior to the first engagement, and PowerPoint summaries of each data source presented by the research team at the beginning of the session. Deliberation around this evidence was then professionally facilitated, while members of the research team remained in attendance to clarify technical matters as they arose (rather than acting as participants). In the second round, outputs from all first-round engagements (i.e. the perspectives of each stakeholder group) were reported to participants and informed their subsequent deliberations.

The choice to separate stakeholder groups supported participants to voice partisan opinions freely, without fear of the interactional impact of relative in/expertise (Hendriks, Dryzek & Hunold, 2007), and allowed an analytic focus on within-group concerns and constructions. At the same time, the 'feeding back' of other groups' outcomes served the key deliberative goal of bringing together a range of (potentially contradictory) perspectives (Shapiro, 2003).

While ultimately a research exercise, the process described above sought to mimic, and enhance with deliberation, existing policy processes in which the research team has a history of participation. As such, and given the participation of policy advisors with health system authority, it follows that physicians, in particular, treated their involvement as having the potential for material impact.

Recruitment of ART physicians

Recruitment of the physicians began in a purposive manner. An introductory email detailing the aims and scope of the project was sent to the Fertility Society of Australia, the Royal Australian and New Zealand College of Obstetricians and Gynaecologists, the Royal Australian College of General Practitioners, the Royal Australasian College of Physicians and the Australian Medical Association. Each organisation was invited to send a representative to attend the engagements, and was offered the opportunity to suggest other potentially relevant participants. Recruitment then snowballed: relevant physicians suggested by the original invitees were subsequently approached. This method yielded seven participants (Table 1).

The project received ethics approval from The University of Adelaide. Participants were assured that their discussions would remain anonymous when published, and that they were free to withdraw their participation at any time. Two first-round participants were unable to return in the second round; both conducted during 2010. Consent to record and transcribe deliberations for later analysis was obtained from all in attendance.

Table 1
Deliberative engagement participant characteristics.

Gender	Male	3
	Female	4
Clinical area	Reproductive medicine specialist	3
	Other (e.g. obstetrician, gynaecologist, GP, neonatologist)	4

Theoretical perspective and approach to analysis

This paper takes the perspective that policymaking does not simply represent a process in which self-evident problems are addressed via the neutral appraisal of technical evidence. Rather, our approach is to understand policymaking as a process in which the constitution of policy 'problems' and 'solutions' is actively negotiated within a social context. Underpinning this perspective is a broadly social constructionist epistemology, and a focus on language as the medium in which policy 'issues' and 'evidence' are constituted.

In line with this framework, we take a synthetic approach to the analysis of discourse (Wetherell, 1998), attending to the *structure* of physicians' accounts (how they are put together rhetorically so as to appear factual, credible, etc) as well as their *function* (how they serve to position certain kinds or orders of evidence as more or less compelling). This rhetorical, discursive perspective allows an analytic focus on the "language, arguments and discourses through which policy is constructed and enacted, and provides a conceptual framework for linking decision-making to processes of practical reasoning" (Russell & Greenhalgh, 2009).

In the first analytic phase, transcripts of engagements with ART physicians were worked through to build up a file of instances referencing 'evidence' of ART (treatment and cost) effectiveness. Analysis proceeded to identify the main repetitive patterns evident across the corpus, attending to:

1. The strategies by which accounts of evidence were constructed as more or less 'accurate'.
2. The 'action orientation' of these accounts: how they may be seen to function in supporting a particular construction of 'evidence', and the ideological and material consequences of this construction.

Analysis

In this section we map the strategies by which physicians contested research evidence around the effectiveness of ARTs on the grounds of 'currency', 'proximity', 'selectivity' and 'bias'. In turn, we outline the physicians' constructions of *what should count* as evidence informing ART policy, and *how* this evidence should be counted.

A question of evidence

Statistics drawn from the Australian and New Zealand Assisted Reproduction Database (ANZARD, a registry of the outcomes of all initiated cycles of ART) formed the basis of the evidence presented to physicians for their deliberation (see Watt et al., 2011). One headline statistic from the ANZARD registry reports that, based upon 2007 data collection (the most recent year available), an average of 18.8% of initiated cycles of ART culminated in a live birth, with a maximum 30.1% of initiated cycles generating a live birth for women aged at or under 30 years and a minimum of 1.0% of cycles generating a live birth at ages ≥ 45 (Wang, Chambers, Dieng, & Sullivan, 2009). However, the accuracy of this representation of 'IVF success' (and its role in informing analysis of the cost-effectiveness of ART) was questioned by physicians on a number of related grounds.

Currency

As illustrated in the extracts below, physicians' accounts of significant recent improvements in IVF outcomes served both to undermine claims of 'ineffectiveness' based upon the 'maximum 30.1%' figure presented by the research team, and to question the

legitimacy of identifying ART as even a candidate for disinvestment discussions.

Extract 1

Physician 1: What year were those stats? I think you'd probably find those sorts of stats have changed a fair bit in the last couple of years.

Research team member: They are from the early 2000s. ...

Physician 1: The pregnancy rates in that group would have doubled during that time.

Extract 2

Research team member 1: The ANZARD data, it's (...) reporting still in the 30% not the 50%, the pregnancy rate.

Physician 1: That is 30% overall for all cycles. We are talking about 50% under 38.

Research team member 2: [Referring back to the data she is quoting] Less than 35 years pregnancy ANZARD 2004 to 2007.

Physician 1: That is really old.

In Extract 1 above, Physician 1's 'insider knowledge' of recent improvements in IVF outcomes serves to question the legitimacy of debate based upon outcomes that have since "doubled" – a relative quantifier that bolsters her claim of 'significant change'.

Similarly, in Extract 2, Physician 1 questions the validity of using ANZARD data on the grounds that it is "really old". In turn, she privileges her own representation of IVF outcomes ("50%" for "women under 38") over those derived from the database.

In the context of disinvestment debates, such emphasis on 'current' data – while understandable given physicians' professional interests and day-to-day vantage point – represents a potential barrier to doctors' engagement in the process of policy development. Indeed, engagement in policy debates requires tolerance of some level of 'data lag'. Acknowledgement of the provisional, defeasible nature of medical evidence underpins disinvestment discussions across the board, and represents an important step in moving debate from issues of 'data' to more abstract notions of allocative efficiency. At the same time, however, this discussion suggests that physicians' input represents a means of ensuring that policy is grounded in 'meaningful evidence' – a particularly important contribution in areas of rapid technological change.

Proximity

Linked to questions of currency were questions of 'proximity' – the notion that physicians are privy to more recent, and richer, outcomes data by dint of their insider knowledge.

Extract 3

Physician 2: The total data is published late. We do have month-to-month data.

Extract 4

Physician 2: We know [the data on effectiveness]. I'm sure [Dr X] has her numbers in her head. I know the most recent [Clinic X] numbers are higher than they were a year ago.

In the extracts above, Physician 2 presents an account of physicians' nuanced and up-to-date awareness of ART outcomes. While Physician 2's account of improved outcomes is imprecise (the numbers are simply described as "higher than they were a year ago"), it provides enough information to sustain an argument of growth that casts doubt on the 'official' statistics.

The accounts presented here have implications for disinvestment debates in that they serve to question the validity of

policymakers' use of the ANZARD dataset on the grounds of both currency and status. In contrast to the perspective presented elsewhere in the engagements that ANZARD reporting requirements make ART a disinvestment "target ... because there's evidence to use" (Physician 2), accounts prioritising doctors' 'data-proximity' reclaim physicians' prime position as evidence arbiters. Indeed, in the extracts above, claims of absolute understanding and consensus ("I know", "We know"), alongside indications of intimate statistical knowledge ("I'm sure Dr X has her numbers in her head"), build a possessive account of the data. Such accounts are in contrast to those identified in *Jenkins and Barber's (2004)* study of evidence negotiation in hospital new drug committees, in which physicians (and others) took pains to quote 'official statistics' in order to present themselves as 'objective' in the context of data appraisal. Indeed, here, it is physicians' *subjectivity* that is taken to bolster the credibility of their version of appropriate evidence.

In a policy context necessarily reliant on population statistics, an emphasis on 'data proximity' may preclude physicians' engagement with bigger-picture disinvestment discussions that may take official data records 'as read'. However, it could equally be seen that physicians' access to 'data in context' sees them uniquely positioned to add important nuance to policymakers' interpretations of 'official' evidence repositories.

Selectivity

Physicians also contested the evidence provided by the research team (particularly that relating to the analysis of ART's cost-effectiveness) on the grounds that analytic selectivity could be seen to have coloured the data presented.

In particular, physicians criticised what they saw as the omission of reference to contextual factors impacting upon ART outcomes. These factors, deemed to complicate outcomes data and subsequent cost analyses, included changes to Medicare rebates (the proportion of the treatment cost covered by Medicare, Australia's universal public health insurance) and the lack of funding for egg donation.

Extract 5

Physician 2: You bring in the confounding factors of minimal funding being applied to egg donation which pushes those older women to keep trying relentlessly for their own eggs.

Extract 6

Physician 1: The evidence presented is using Medicare data based on the Medicare funding from last year which is already significantly different this year because of, effectively, the reduction of public funding that occurred with the changes to the Extended Medicare Safety Net*.

(*The EMSN represents a ceiling on the out-of-pocket costs an individual or family can incur through their use of Medicare out-of-hospital services. Once the threshold is reached, the EMSN pays 80% of the out-of-pocket costs for Medicare-related services for the rest of the calendar year.)

In each of the extracts above, important social and economic factors are depicted as "confounding" the ANZARD-derived measurement of treatment outcomes. As these accounts suggest, ART success rates are potentially impeded by the prohibitive cost of egg donation (without it, older women's numerous ART attempts may reduce per-cycle rates of success), and by subsidy restrictions (perhaps limiting the access of couples who might otherwise have found success in subsequent cycle attempts).

While generated in the context of policy-oriented debate, considering ART outcomes within their social and economic location, these accounts appear to privilege evidence of ART's *potential* rates of success in a 'perfect world'. Although advocating the inclusion of contextual factors in success rate calculations (commensurate with an effectiveness model oriented to 'outcomes under real-world conditions'), the accounts present such factors as obscuring the 'real' success of ART treatment programs that would be enjoyed 'if conditions were ideal'.

It may be appropriate that doctors attend to the specific issue of potential treatment outcomes rather than taking a broader view: their vantage point naturally foregrounds treatment 'possibilities' for individual patients rather than population-level outcomes. However, such an evidentiary lens is at odds with a perspective appraising ART's effectiveness-in-context (the implicit basis of a policy-oriented discussion). Ultimately, in the extracts above, what appears to be an empirical debate about the accuracy of evidence (what are the 'real' outcomes of ART?) may instead be seen as an ontological debate about the status of *different kinds* of evidence (What counts as evidence of ART effectiveness: what we see, or what we *could* see under optimum conditions? What are our grounds for deciding where 'accuracy' and 'relevant context' meet?).

As people who see firsthand the important benefits of their treatments, and the potential for improved outcomes if external factors are conducive, it is to be expected that physicians will present the best possible account of ART success. Moreover, in a disinvestment context, it seems natural that physicians will defend against potential subsidy restrictions that will impact upon them both professionally and materially. What is important, however, is that these allegiances, and framings of appropriate evidence (foregrounding ideal/potential treatment outcomes, or effectiveness-in-context), are made explicit within a participatory policy process. Likewise, as part of an enhanced disinvestment evidence base incorporating stakeholder perspectives, openness about physicians' 'stake' (Wetherell, 1998) could serve to contextualise their offering as a vital means of adding nuance to the received statistical wisdom, and to position their contribution as generative as opposed to defensive.

Bias

Linked to questions of selectivity was the more explicit claim of researcher 'bias'.

Extract 7

Physician 3: I think the evidence that was presented was very biased ... I think it was a particular idea that they wanted to push ... We have the idea of what actually happens in ART.

Research team leader: Could I backtrack on that evidence question?

... In what way was that evidence biased, given we followed a systematic review protocol that we had some external input into? That's the steps. ... Did we ask the wrong questions?

Physician 1: Yes.

In this section of talk, the evidence presented by the research team was questioned on the grounds that the outcomes were perceived to reflect a preconceived agenda rather than an impartial appraisal of evidence. The exchange outlined here illustrates two levels of complexity when it comes to policy processes in which stakeholders are engaged in evidence appraisal.

Stakeholders' perspectives may prime them to treat certain realms of evidence as particularly salient or, indeed, as 'obvious'. In turn, policy discussions grounded in a presentation of evidence that minimises or excludes these concerns may be critiqued not only as

inaccurate but, potentially, as 'wilfully' so. The validity of evidence presented may therefore be questioned on the grounds of data accuracy *and/or* the credibility (absence of vested interest) of the sources by whom the data are produced – claims that may be elided into the generic critique that the data are simply "wrong".

The argument outlined here raises a criticism often levelled at deliberative processes: that the presentation of 'evidence' necessarily involves selections that work to frame resultant discussions (Walmsley, 2009). In response, it could be argued that this situation simply mimics real-world policy processes in which (often) non-expert political advisors are presented with a specific range of data to inform their decisions. Through this lens, the methodology of the current project – seeking to enhance a traditional evidence base for policy through the incorporation of stakeholder perspectives – goes some way towards counterbalancing issues of evidence selectivity.

Issues of data selection and evaluation may apply equally to the contributions of stakeholders themselves. In each of the extracts addressed thus far, physicians' critiques of the research team's data also comprised selections (recent data was privileged over retrospective figures; excluding contextual factors was constructed as undermining the validity of calculations of ART success) and evaluations (the credibility of data sources was held to be diminished by 'vested interest' and bolstered by 'data proximity').

We can also see from the extracts that physicians did not simply critique the research team's evidence on the grounds of *what* data was counted, but also *how* it was counted. For example, in Extract 7, the implicit contention that accuracy calculations should foreground the *potential* of the treatment – perceived efficacy under optimal policy and clinical conditions – rather than simply attending to the real-world effectiveness-in-practice as had been the researchers' perspective.

In the context of disinvestment decisions, different conceptualisations of evidence – questions of *what* and *how* to count – are not without effect. We now demonstrate this through an analysis of the alternative accounts of ART's (cost-) effectiveness proffered by the physicians with whom we engaged.

What to count: physicians' perspectives

Throughout the engagements, physicians argued that a range of factors not considered by the research team must be included in any valid appraisal of the value of ART.

Alternative treatment costs

Physicians noted the potential for alternative costs, or unintended consequences, to replace those incurred throughout the course of ART if access to such treatment was restricted.

Extract 8

Physician 4: If IVF was not as readily available and in particular if it was more poorly funded, a lot more couples would have laparoscopies ... which unfortunately can be futile. ... [T]he number of laparoscopies that I personally do has dropped off a lot since IVF has become more successful and more accessible, and laparoscopic surgery is very expensive.

In this extract, the cost of laparoscopies – presented as a consequence of IVF subsidy restriction – is depicted as an essential component of any valid calculation of the sustainability of ART when such technologies are viewed 'in context'. Implicit in this account is a conceptualisation of the evidence required for disinvestment discussions as only appropriate/useful if it measures the

(potential) 'hidden costs' of restrictions – a perspective orienting to 'the entire system-borne cost of intervention/s in this area', rather than that of one specific technology. As evident in the extract above, this perspective is reliant upon physicians' anecdotal data around patient behaviour – a different order of evidence than that traditionally informing cost-effectiveness analyses.

Alternative costs borne by *individuals* were also raised as a requisite component of ART cost analyses, once again expanding the terrain of 'relevant evidence' required to inform an analysis of cost-effectiveness.

Extract 9

Physician 4: You need to compare the cost of a couple conceiving with IVF with the cost of other sorts of treatments they would be looking at if IVF wasn't funded, not compared to a home conception which is zero.

Physician 2: The costs would never get measured: naturopathic costs, acupuncture costs, Chinese medicine costs, herbs, psychological interventions.

Here, comparisons between the cost of home conception and that of IVF are constructed as both unfair and misleading. In response, Physician 2 widens the remit of the cost-effectiveness analysis to include the patient-borne cost of alternative and psychological treatments deemed likely to result from ART subsidy restrictions.

Cost of restrictions: health of mother and baby

Physicians also raised concerns about the hidden costs of ART subsidy restrictions in terms of the possible adverse health outcomes for mother and baby if 'alternative measures' are taken. That is, they argued that subsidy restrictions would reduce the affordability of multiple cycles of IVF, thereby increasing the pressure on physicians to extract more eggs, and replace multiple embryos, in a bid for early-cycle success. The outcome of this patient demand, as the physicians saw it, would be increased rates of both ovarian hyperstimulation and multiple births (resulting from pressure to transfer multiple embryos).

Extract 10

Physician 4: I know of the costs of the alternative often futile surgeries that will be done ... but nothing [is said] about all the costs that we all know that would suddenly occur because of multiple births and increase in ovarian hyperstimulation.

Extract 11

Physician 5: If you are replacing multiple embryos and have twins or higher multiple pregnancies then they automatically become far more complicated, you have a far greater risk.

Within these accounts, ovarian hyperstimulation and risky multiple births are constructed as inevitable and costly consequences of IVF subsidy restrictions. This narrative suggests a causal relationship between subsidy restrictions and the potential for these adverse outcomes. Yet the notion that physicians have considerable professional agency in guiding and regulating these circumstances (i.e. practising 'in the best interests of the patient') is absented in an account that implicitly depicts patient persuasion as an 'insurmountable' pressure.

Interestingly, in the context of a policy-oriented discussion, the construction of these inevitable adverse outcomes serves to bolster the argument in favour of retaining IVF subsidies, rather than to support calls for increased regulation of IVF treatment processes.

Emotional costs

Extract 12

Physician 2: [T]here's very little emphasis given to looking at the evidence of the psychological burden of childlessness, the psychological burden of shame and blame and secrecy around infertility, and perhaps a measure of success with an unsuccessful IVF journey for a woman may be [that] she can move on to a life of childlessness feeling as if they have made as much effort as they can.

Extract 13

Physician 2: I'm talking about childless women who would have otherwise had a chance and didn't because of technology not being available, just because they don't talk about it doesn't mean there isn't this big elephant in the room of hurt and sadness and tragedy. I don't know it's possible to measure that.

The notion that emotional costs must be factored in to analyses was a final critique raised by the physicians. In these extracts, Physician 2 once again expands traditional components of cost-effectiveness calculations to encompass the psychological burden of childlessness potentially alleviated by access to IVF. In these accounts, the cost of being denied access to reproductive technology is simultaneously constructed as measurable (closure in an unsuccessful IVF journey is depicted as a 'measure of success') and *immeasurable* (Physician 2 doesn't "know if it's possible to measure" the detrimental effects of not being given "a chance"). Thus the intangible benefits of IVF are rendered specific enough to sustain a moral argument (providing closure is surely a significant measure of success), yet vague enough to avoid critical scrutiny (it is difficult to imagine how closure may be defined, let alone measured and subject to systematic review). Interestingly, the incorporation of psychological health outcomes in these accounts focuses solely on the positive role that may be played by IVF: no mention is made of the potential for psychological harm (and options for its measurement) that could result from multiple failed attempts.

As these examples show, different conceptualisations of *what* should be counted have significant differential effects on the calculation of the value of IVF subsidies. Likewise, alternative conceptualisations of *how* such data should be counted impact upon the sustainability verdict.

How to count: physicians' perspectives

A range of implicit decisions around how data should be evaluated underpinned physicians' accounts of the sustainability of ART. For example, physicians suggested that evidence of ART's cost-effectiveness should be assessed in relative (rather than absolute) terms: that is, assessed in relation to the cost-effectiveness of other treatments, the notion of 'waste' constructed as only possible to evaluate in relation to other areas of health provision.

Extract 14

Physician 6: If you ask them to cut medical services that are utterly useless, for example putting people 75 years or more on ICU, they will not even dare to go there. We all know that's a total waste of money, that costs far exceed any fertility treatment but they will never touch that because they are not brave, they are politicians.

Extract 15

Physician 3: It seems like there's a universal agreement that \$30,000 approximately for a live birth per cycle, that relative to other modalities of medicine that is good value for money.

Physician 6: It's surprisingly cheap compared to the waste of money we have in healthcare.

Similarly doctors argued that any calculation of the sustainability of IVF should interpret cost-effectiveness data in terms of the 'return on investment' of such technology when compared to other areas of medicine.

Extract 16

Physician 4: You need to look at return on investment. They can pay and look after us in the future. If it's [\$]8000 to create a life, as long as they are healthy which most of are, unless we have to start putting two embryos back, we have a tax payer for 40 years.

In Extract 16, physicians may be seen to apply a cost-utility frame, inferring value for money via what is arguably a version of the 'cost per quality adjusted life year (QALY) gained' from the live births generated by ARTs. While there is considerable contention around the applicability of QALYs in the context of ART (for example, Devlin & Parkin, 2003 argue that additional lives are not 'improvements in health'), it is significant to note that the physician here appears to borrow from, while subverting, the language of 'cost-effectiveness'. Notions of 'efficiency' and 'return' serve in this extract to assess the relative value of technologies in terms of 'taxpayer payoff', a definitional choice that has significant implications in building an alternative account of 'value'.

We conclude with a data extract illustrating the way in which contest over evidence construction was played out in context, with ideological (and, potentially, material) effect.

Extract 17

Physician 4: [T]he poor prognosis and futile [ones] they are such a tiny little proportion of the people we treat where we're generally looking at a 50% pregnancy rate per cycle for the majority, the large group of women we treat. The return on investment for that is amazing. You generally do get a healthy child if you put only one embryo back in

Physician 2: You can't assess if someone is in that group that has a 50% chance of success until they've done three or four cycles.

Physician 4: For someone who gets pregnant on their fourth cycle, may still have a good 50% chance of pregnancy. It randomly takes them – you know, a throw of the dice is four cycles for them.

Physician 2: If you're not pregnant by four then maybe you fall into the group that doesn't have the 50% chance. This is disregarding the ones we know from the starting point they are not 50[%]. The vast majority falls into the group who should have about a 50% chance of success.

Physician 4: When most good units in Australia have got overall pregnancy rates, and that is putting the poor prognoses in with the tiny number of futile ones, overall per cycle of 35–40% with live birth rates of 35%. These people that public economists could argue we are throwing money away for is a tiny proportion of our overall people going through.

Physician 2: Smaller amount of money than the election advertising bill.

In the first paragraph of Extract 17, an account of ART treatment effectiveness is bolstered in a number of related ways. Firstly, and perhaps most significantly, is the selection of *pregnancy rate per cycle* as the key measure of treatment success ('we're generally looking at a 50% pregnancy rate per cycle'). As an indicator of outcomes, 'pregnancy rate per cycle' maximises treatment success as it divides the total number of pregnancy outcomes by the total number of cycles undertaken by all women. This calculation does

not include unsuccessful patients who drop out between cycles, whereas such attrition would be included in a calculation of the 'pregnancy rate of patients who commence IVF treatment'. Pregnancy rate per cycle also includes the outcomes of later-rank cycles, in which patients would have had the benefit of ongoing interventions aimed at increasing their chance of conception.

Further support for a maximised presentation of success is achieved in the extract by attaching a population pregnancy rate ("a 50% pregnancy rate per cycle") to individual patients ("a 50% pregnancy rate per cycle for the majority, the large group of women we treat"). It may be seen that this account conflates different levels of explanation: a success rate calculation based on the total number of pregnancies resulting from the total number of cycles undertaken by *all* women is presented as though it represents an *individual* woman's likelihood of success.

Additional maximisation may then be seen to underpin the 50% pregnancy rate outlined by Physician 4. Rather than presenting the percentage pregnancy outcomes of *all* women who attend for IVF, the calculation on which the claim is built divides the number of pregnancy outcomes by a reduced patient sample – the 50% success rate is attributed to the "majority of women that we treat". Ultimately, the account presented here suggests that approximately half of all women who undergo a cycle of IVF will become pregnant, yet relies upon a calculation from which a group of unsuccessful women (those labelled "poor prognosis" or "futile") have been removed (yet unsuccessfully treated). In turn, the statement "you generally do get a healthy child" (suggesting a 'better than chance' outcome and shifting the outcome measure to 'live birth') supports the claim that IVF represents an "amazing" return on "investment".

Throughout this extract, a range of descriptive selections are used to underpin an account of IVF success and value. These selections include a shift from population statistics to individual likelihood predictions, as well as an adjusted patient denominator being employed to maximise depictions of success. We also note a contrasting use of absolute figures ("50%") and the use of non-numerical relational descriptors (successful patients are described as the "large" "majority", while the group of those who are unsuccessful are minimised as "tiny"), as well as a shift between outcomes variously defined as "pregnancy", "a healthy child" (an outcome notably absent from official data repositories), or the more vaguely described "success".

Within disinvestment deliberations, such selections, categorisations and resultant evaluations provide support for specific funding outcomes and interpretations of value. For example, Physician 2's apparently technical claim that "someone who gets pregnant on their fourth cycle, may still have a good 50% chance of pregnancy. It randomly takes them – you know, a throw of the dice is four cycles for them" can be seen to support the moral argument in favour of giving all women a minimum of the four cycles it might "randomly take them" to conceive within this process. The presentation of four cycles as being necessary for an authentic assessment of their chances conflates the notion of 'chance' (what is random) with the notion of 'giving them a chance' (an ethical concern).

Likewise, Physician 4's ostensibly factual claim that the "people that public economists could argue we are throwing money away for is a tiny proportion of our overall people going through" supports an evaluative comparison against the cost of election advertising. Physician 2's statement that unsuccessful IVF treatments represent a "smaller amount of money than the election advertising bill" brings to bear not only a fiscal comparison, but a moral judgement about the relative expenditure: how can we think about restricting this important/effective procedure the costs of which are exceeded by other (morally questionable) outlays?

Discussion

Our analysis illustrates that engagement with physicians around a disinvestment question did not serve simply to append objective outcomes data with 'clinical judgement' to generate a 'more accurate' assessment of ART effectiveness that might inform relevant policy. Instead, physicians' constructions of legitimate evidence served to question and reframe both the terms of debate (whether ART should even be considered as a disinvestment candidate) and means of appraisal (what should be included/excluded from definitions of 'effectiveness'). Our illustration of the rhetorical, argumentative nature of physicians' contributions affirms the importance of opening the 'black box' of health policy decision-making particularly in contentious areas where stakeholder contributions are vital, yet processes for incorporating them remain unclear (Williams & Bryan, 2007).

As we have outlined, physicians' critique of the research team's data on the basis of currency, proximity, selectivity and bias undermined the validity of identifying ART as a disinvestment candidate. Given the authority afforded doctors within priority-setting deliberations (Duthie et al., 1999), the mobilisation of these strategies is significant, serving to elevate physicians' own account of evidence (based upon 'current, first-hand' experience) above that derived from clinical literature. Physicians' influence in this regard, and the notion that any priority-setting activity may be "doomed at the outset" (Mitton et al., 2003: p. 1659) without their engagement, appears to see them favoured by a deliberative evidence asymmetry within such processes. Indeed, we suggest that physicians' critiques have the potential to forestall disinvestment debates on the grounds of questions of evidence – a potentially problematic finding when one considers physicians' stake in priority-setting decisions that stand to impact upon them both professionally and financially.

While it may be deemed quite legitimate for physicians to be afforded a particular 'category entitlement' (Potter, 1996) to knowledge of clinical data, the role of their contribution to the ethical aspect of disinvestment debates is perhaps more contentious. While McKie, Shrimpton, Hurworth, Bell, and Richardson (2008) found that consumers desire a distinction to be made between clinical and ethical expertise within health priority-setting debates, these aspects may be discursively elided in deliberative contexts. As our analysis illustrated, physicians' accounts of ART treatment effectiveness blurred the line between clinical data (evidence of outcomes) and assessments of worth (arguably ethical concerns) within an ostensibly 'neutral' appraisal of evidence. The notion that physicians' moral arguments could be afforded the same credence as their clinical contributions may be of concern where health policymaking deliberations seek to include, and balance, a range of stakeholder perspectives on economically and socially complex issues.

Conclusion

A central criticism of deliberative democratic methods within policy contexts is the notion that outcomes are inevitably shaped by the evidence informing discussions (Russell & Greenhalgh, 2009). Interpretations of evidence, and subsequent deliberative outputs (Walmsley, 2009), are necessarily 'framed' by the selections and evaluations that underpin what is often represented as an 'objective' evidence base. Our analysis suggests that such concerns are legitimate. As we indicate, both policy 'problems' and 'solutions' rely upon the selection and interpretation of evidence, and decisions in this regard are inexorably linked to the social context in which they are negotiated.

We argue that the difficulty of negotiating questions of evidence should not be taken as cause to abandon engaged,

deliberative processes in the context of disinvestment decision-making. On the contrary: the accounts of physicians described above add vital nuance to debate around disinvestment options, their contribution highlighting an important perspective and supporting accountable decision-making. For example, physicians' capacity to catalogue potential 'unintended consequences' represents a valuable contribution when modelling the likely impact of any disinvestment initiative. Similarly, their pivotal role in implementing policy outcomes suggests the importance of facilitating physicians' buy-in at key decision-making junctures.

Ultimately, however, our analysis indicates that dispute around what constitutes 'appropriate evidence' in a disinvestment context has the potential to forestall true deliberation in this domain. For this reason, we argue that stakeholder-engaged deliberative processes might usefully begin with an explicit, meta-level discussion around evidence and 'what counts'. Our analysis suggests that once a technology or practice has been identified for disinvestment consideration, consensus from a range of stakeholders (whose interests and perspectives are explicitly acknowledged and appreciated for the value they add) around what *can*, *does* and *should* constitute evidence has the potential to support later, genuine, deliberation around funding policy.

Acknowledgements

The ASTUTE Health study is funded by the National Health and Medical Research Council of Australia (NHMRC Grant ID 565327). Adam Elshaug holds a Sidney Sax Fellowship from the NHMRC, ID 627061. We would like to thank the physicians who generously participated in this study.

Disclaimers

The views presented here are those of the authors and should not be attributed to The Commonwealth Fund, including directors, officers, or staff.

References

- Abelson, J., Giacomini, M., Lehoux, P., & Gauvin, F. (2007). Bringing 'the public' into health technology assessment and coverage policy decisions: from principles to practice. *Health Policy*, 82, 37–50.
- Assisted Reproductive Technologies Review Committee. (2006). *Report of the Independent Review of Assisted Reproductive Technologies*. Canberra: Department of Health and Ageing.
- Carter, D., & Braunack-Mayer, A. (2011). The appeal to nature implicit in certain restrictions on public funding for assisted reproductive technology. *Bioethics*, 25(8), 463–471.
- Davies, C., Wetherell, M., & Barnett, E. (2006). *Citizens at the centre: Deliberative participation in healthcare decisions*. London: Policy Press.
- Devlin, N., & Parkin, D. (2003). *Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A discrete choice analysis*. London: Department of Economics, City University of London.
- Dill, S. (2006). How Australia secured IVF funding. <http://www.iaac.ca/content/partnerships-politics-and-ivf> Accessed 04.01.11.
- Dobrow, M. J., Goel, V., & Upshur, R. E. G. (2004). Evidence-based health policy: context and utilisation. *Social Science & Medicine*, 58, 207–217.
- Dryzek, D. (2000). *Deliberative Democracy and Beyond*. Oxford: University Press.
- Duthie, T., Trueman, P., Chancellor, J., & Liez, L. (1999). Research into the use of health economics in decision making in the United Kingdom – phase II: is health economics for 'good or evil'? *Health Policy*, 46, 143–157.
- Fishkin, J., Luskin, R., & Jowell, R. (2000). Deliberative polling and public consultation. *Parliamentary Affairs*, 53(4), 657–666.
- Green, J. (2000). Epistemology, evidence and experience: evidence based health care in the work of accident alliances. *Sociology of Health and Illness*, 22(4), 453–476.
- Griffiths, A., Dyer, S., Lord, S., Pardy, C., Fraser, I., & Eckermann, S. (2010). A cost-effectiveness analysis of in-vitro fertilization by maternal age and number of treatment attempts. *Human Reproduction*, 25(4), 924–931.
- Hendriks, C. M., Dryzek, J. S., & Hunold, C. (2007). Turning up the heat: Partisanship in deliberative innovation. *Political Studies*, 55, 362–383.

- Jenkings, N.K., & Barber, N. (2004). What constitutes evidence in new hospital drug decision-making? *Social Science & Medicine*, 58, 1757–1766.
- Lomas, J., Culyer, T., McCutcheon, C., McAuley, L., & Law, S. (2005). *Conceptualizing and combining evidence for health system guidance*. Ottawa: Canadian Health Services Research Foundation.
- McKie, J., Shrimpton, B., Hurworth, R., Bell, C., & Richardson, J. (2008). Who should be involved in health care decision making? A qualitative study. *Health Care Analysis*, 16, 114–126.
- Mitton, C., Patten, S., Waldner, H., & Donaldson, C. (2003). Priority setting in health authorities: a novel approach to a historical activity. *Social Science & Medicine*, 57, 1653–1663.
- Murray, S., Holmes, D., & Rail, G. (2008). On the constitution and status of 'evidence' in the health sciences. *Journal of Research in Nursing*, 13(4), 272–280.
- Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction*. London: Sage.
- Russell, J., & Greenhalgh, T. (2009). *Rhetoric, evidence and policymaking: A case study of priority setting in primary care*. London: Research Department of Primary Care and Population Health, UCL, URL: <http://eprints.ucl.ac.uk/15560/>.
- Scutchfield, F. D., Ireson, C., & Hall, L. (2004). The voice of the public in public health policy and planning: the role of public judgement. *Journal of Public Health Policy*, 25(2), 197–205.
- Shapiro, I. (2003). *The state of democratic theory*. Princeton: Princeton University Press.
- Tritter, J. Q., & McCallum, A. (2006). The snakes and ladders of user involvement: moving beyond Arnstein. *Health Policy*, 76, 156–168.
- Walmsley, H. (2009). Mad scientists bend the frame of biobank governance in British Columbia. *Journal of Public Deliberation*, 5(1), 1–26.
- Wang, Y. A., Chambers, G. M., Dieng, M., & Sullivan, E. A. (2009). Assisted reproduction technology in Australia and New Zealand. In *Assisted reproduction technology series*. Sydney: AIHW National Perinatal Statistics Unit.
- Watt, A., Elshaug, A., Willis, C., & Hiller, J. (2011). Assisted reproductive technologies: a systematic review of safety and effectiveness to inform disinvestment policy. *Health Policy*, 102(2–3), 200–213.
- Wetherell, M. (1998). Positioning and interpretative repertoires: conversation analysis and post-structuralism in dialogue. *Discourse & Society*, 9(3), 387–412.
- Williams, I. P., & Bryan, S. (2007). Cost-effectiveness analysis and formulary decision making in England: findings from research. *Social Science & Medicine*, 65, 2116–2129.