Necessity and health care coverage decisions

17th-century English statesman Oliver Cromwell held that it knows no law. As far as Greek tragedy poets Euripides and Sophocles were concerned, nothing had more strength. Even Ares, the god of war, could not battle it: necessity. These days, dictionaries speak of something required, unavoidable, indispensable, enforcing even. The philosophical definition runs along the same lines: "the principle according to which something *must be so*, by virtue either of logic or of natural law". As more contemporary idiom would have it, however, necessity is the mother of invention – you can even make a virtue out of it.

On this scale from enforcing to inspiring, many things have been considered a necessity: education, faith, fiction, happiness, music, and even private jets. As in the case of music, the person who deemed it so (in this instance, soul music legend Ray Charles) may add the nuance that it concerns a personal, rather than a general, necessity: "this is necessary for me". In the realm of health care, this dynamic rears its head too. Is a certain form of health care a necessity for one person – or does this also hold for more than one? Historically, medical doctors were responsible for making the decision for the individual patient, but in many countries, this has since been supplemented by general decisions made for more than one patient. These decisions on what is, and what is not, necessary health care taken on a more collective, often (sub) national level deal with the necessity of the health care technology generally. Examples of widely acknowledged *unnecessary* health care may seem relatively easy to come by. For forms of health care such as cosmetic surgery (Russell et al., 2014) and Viagra (Stolk et al., 2002), many people intuit that provision for all prospective recipients may not be a necessity. Then again, both erectile dysfunction and port-wine stains, the removal of which is often considered cosmetic and thus unnecessary, may cause impeded social functioning and severe psychosocial problems, which does indicate it may be a necessity for some patients. So what, if we think about it, should be the arguments pertaining to necessity to back up a collective health care coverage decision?

One of the few settings where this question is and has been tackled head-on is the Dutch National Health Care Institute (in Dutch: *Zorginstituut Nederland*, ZIN, in this text: the Institute). The Institute is responsible for delivering advice to the Minister of Health concerning which health care technologies should, and should not, be part of the Dutch basic benefits basket. This benefits basket specifies what Dutch citizens are entitled to through their collective health insurance, which is mandatory for all citizens. After receiving advice on the in- or exclusion of the health care technology from the benefits basket, the Minister of Health takes the final decision. The process for arriving at the advised decision at the Institute has on average, over the years, employed four criteria (effectiveness, cost-effectiveness, feasibility, and necessity), with these criteria representing questions that are to be answered in the final advice. For effectiveness, the question is: is this health care technology effective (enough)? Similarly, for cost-effectiveness: is this cost-effective enough? The necessity criterion likewise asks: is it necessary? This last question has also been phrased as: is a claim on the collective solidarity justified?

THIS DISSERTATION

This dissertation describes the outcomes of a study commissioned to operationalise the necessity criterion. The provenance of this project lies with the Institute. The project was part of a larger move towards more research substantiation and support for the Institute's work. This coincided with the set-up of the research network Health Technology Assessment Netherlands (in Dutch: *Academische Werkplaats Verzekerde Zorg*) to act as a "bridge between research and policy". This centre represents a collaboration between Erasmus University Rotterdam, Utrecht University, and the Institute and aims to support research designed together with policy advisors (Zorginstituut Nederland, 2020).

In 2014, a working group of Institute employees considered the necessity criterion likely to benefit from further operationalisation. This group had concerns about the extent to which the then-current version of the formalised necessity criterion reflected political and societal values pertinent to the case at hand. They noted that necessity was not a clear-cut criterion for making coverage decisions, despite multiple operationalisation and re-operationalisation steps that had been taken over the years. At the turn of the century, just over ten years after the term was first coined by the Dunning committee (Commissie-Dunning, 1991), scholars had already considered necessity a "difficult to measure, non-uniform unit" (Poley, 2002, p. 2313). Since then, multiple policy reports had dealt with the necessity question. Despite these efforts, and even though the formulation as a criterion did ensure that most, if not all, coverage advice documents contain a section 'necessity', further operationalisation was still deemed beneficial. This dissertation reports on the study that was done to fulfil this need.

A BRIEF HISTORY OF OPERATIONALISING NECESSITY

This section will provide a brief history of how necessity as a criterion has been operationalised thus far. This overview ranges from 1991 to 2013, which saw the publication of the last relevant policy report before the start of this project in March 2015. I distinguish two general strands of thought in these documents: first is the substance of the criterion, as several steps have been taken to 'operationalise' the necessity criterion, that is, to specify the exact question(s) this criterion should answer. Second, there is a recurring acknowledgement that necessity should be established in deliberation, concurrently with the other criteria and, possibly, other argumentations in the appraisal phase of coverage decisions. In health care coverage policy vernacular, the appraisal is treated as distinct from the establishment of the knowledge base through assessment of (a subset of) the knowledge types available, referred to as 'the assessment'. Such knowledge bases may range from randomised controlled trials to patient-reported outcome measures, which may include data harvested from social media as well as 'real-world data', but also live patient input (Kalf et al., 2018; Makady et al., 2017; Moes et al., 2016; Wiering et al., 2017). The appraisal provides a valuation of these knowledge bases, an exploration of additional pertinent factors, and a formulation of the

advice, and may or may not be set up as a distinct moment in time and space (Jansen et al., 2017; Oliver et al., 2004; Patera & Wild, 2014; Walley, 2007). With their focus on the appraisal as where necessity and its arguments should be brought together, these policy documents also highlight the importance of procedures to embed the necessity criterion in, in addition to the substance of the criterion.

In 1991, the term 'necessity criterion' was coined in the report '*Kiezen en delen; rapport van de commissie Keuzen in de zorg*', authored by the 'Dunning committee' (Commissie-Dunning, 1991). This report was written at the time of a reorganisation of health care more generally in the Netherlands (Helderman et al., 2014). The Dunning committee was the first to introduce a necessity criterion for making choices in health care. The report advocated the community approach (as opposed to the individual or medical-professional approach) to establishing necessity. This meant that necessary care was defined as all care that enabled, sustained, and where possible improved opportunities for individuals to share existence with other members of society. The core question of necessity was defined as what care would be considered necessary from a communal point of view, resulting in a potential 'ranking' of care according to level of necessity.

The Dunning committee specifically advocated the use of what they termed a funnel, which was to go down in history as 'Dunning's funnel' (in Dutch: *de trechter van Dunning*). This funnel contained four sieves (necessity, effectiveness, efficiency, and 'for own account' – the last of which will also become significant in due course). The idea was that forms of health care would either pass through all four sieves and end up in the basic benefits basket that was positioned underneath the funnel or get 'caught' in one of the sieves and therefore not be provided. Notably, this contrasts with the actual use of the four criteria in several examples that are given in the report. Helpful here is the case of in vitro fertilisation (IVF), where the Dunning committee was divided on the necessity, struggling, for example, with the question whether societal functioning was hampered by being childless. There were questions on financing as well in terms of own responsibility: IVF is expensive, but as expensive as adoption, which was (and is) not part of the benefits basket. All in all, the committee concluded, IVF is not highly necessary and should be low on the ranking. This showcases how criteria already in 1991 served less as a sieve and more like arguments to be weighed concertedly. In fact, as the committee poses later in the report, the goal of the whole exercise is to provide "arguments to base choices in health care on" (Commissie-Dunning, 1991, p. 109).

After its inception in the Dunning committee's report, the life of necessity as a formalised criterion was to last for nearly three decades. In this time, it went through several changes. Several reports by other government bodies followed Dunning's efforts, such as the *Contours of the Basic Health Benefit Package* report by the Health Council of the Netherlands (in Dutch: *Gezondheidsraad)*, which stated that package management should be based on both scientific and societal grounds and underlined the significance of following good procedures (Gezondheidsraad, 2003). This was resonated by the first of a duo of reports by the Council for Public Health and Care (in Dutch: *Raad voor de Volksgezondheid en Zorg*, RVZ, now *Raad voor de Volksgezondheid en Samenleving*, RVS). The first was the *Sensible and Durable Care* report (Raad voor de Volksgezondheid en Zorg, 2006),

published June 2006. This report was primarily concerned with the criteria for coverage decisions (whereas the next report, *Just and Durable Care* (Raad voor de Volksgezondheid en Zorg, 2007), dealt with the formalised procedures). The *Sensible and Durable Care* report noted the difficulties with operationalising necessity "not only in the Netherlands but elsewhere" (Raad voor de Volksgezondheid en Zorg, 2006, p. 6). It also names 'necessity/need for care' as a criterion, specifying that: "the higher this is, the sooner care would qualify to be paid for from collective means" (Raad voor de Volksgezondheid en Zorg, 2006, p. 6). This is operationalised as individual severity of illness only, specifying this as "the severity of illness for the 'average' individual as underwritten by society" (Raad voor de Volksgezondheid en Zorg, 2006, p. 15). The report also mentions the criterion 'justice/solidarity' – which is considered even more difficult to operationalise. Four potential concepts are given (egalitarianism, utilitarianism, rule of rescue, and the Maximin principle) but these are all "theoretical, without practical consequences." Usually, the report continues, the consequences come down to "justice based on need for care or based on equal access" (Raad voor de Volksgezondheid en Zorg, 2006, p. 20). This is considered to hang together with the 'solidarity' criterion, which is in turn related to the Dunning committee's 'for own account'. Here, it is important to note that collective funding is not necessary for those technologies patients can pay for themselves but that "necessary care must be given independent of financial carrying capacity" (Raad voor de Volksgezondheid en Zorg, 2006, p. 20). All in all, these considerations are related but it does not become clear how this should be worked out in practice. As to the working of criteria in general, the report notes that "filters or sieves do not work, because the world is not black-and-white" (Raad voor de Volksgezondheid en Zorg, 2006, p. 22). Accordingly, the report argues, these criteria must be weighed concomitantly. Notably, it names the appraisal phase as the place where non-quantifiable criteria feature, where a "*societal correction* on the technically achieved decision becomes possible" (Raad voor de Volksgezondheid en Zorg, 2006, p. 6, emphasis added). Moreover, the report states, if the two differ, the final choice needs to be justified explicitly.

Later that year, in December 2006, The Institute's predecessor, the *College of Health Care Insurances* (in Dutch: *College voor Zorgverzekeringen*, CVZ), was the first to mention package principles in their first *Package management in Practice* report (in Dutch: *Pakketbeheer in de Praktijk*, abbreviated PiP). Necessity was one such principle, which was defined as whether "the disease or required health care justify a claim on solidarity given the cultural context" (College voor Zorgverzekeringen, 2006). Notably, CVZ was also clear that these package principles and the specific criteria that underlay them should be weighed at the same time and without hierarchy – again in the strictest sense deviating from Dunning's sequential funnel (though not from how it was likely used in practice).

Necessity was operationalised in PiP1 as a combination of first, severity of illness and second, 'need for care', which were then to be combined with third, 'for own account': the costs of the intervention on the individual level (College voor Zorgverzekeringen, 2006, p. 36), previously visible in the Dunning report. These three are worked out further in the rest of the text (and the appendix also adds "public health argumentations, such as dangers to third parties" (College voor Zorgverzekeringen, 2006, p. appendix I)). The first element, severity of illness, was defined prefer-

ably quantitatively, in comparable units. The authors admit that this may not always be practical, or the data may not always be available, and this may therefore also be done qualitatively. The second element, need for care, was to "get a picture" of the "appeal to care" that this group of patients may do (College voor Zorgverzekeringen, 2006, p. 36). The third element, for own account, was part of the necessity criterion but to be used to value and nuance the necessity 'data' from the assessment phase during the appraisal phase.

> There is a severity of illness or need for care, and there is an adequate intervention, and despite this positive score, the judgement may still be that the costs do not justify a call on the social health care insurance. Because necessity will rarely be expressed as a resolute 'yes' or 'no', we weigh the question whether something can be for [the patient's] own account in the appraisal phase. (College voor Zorgverzekeringen, 2006, p. 42)

Regarding health care aids, the for own account question had been worked out further, to contain 'common use' and 'financial accessibility'.

> There are questions of 'common use aid', or 'substitution for a commonly used aid'. If it concerns a one-off purchase, with a long use, low costs, that does also not vary much from the provisions a regular citizen would have in their house, then the conclusion may be that the aid can be for own account [the patient can pay for it his or herself]. (College voor Zorgverzekeringen, 2006, p. 42)

In addition, 'financial accessibility', containing both contributions and/or accumulation of costs, were to be considered as part of necessity of insurance (see appendix 1 of the report). Overall, many elements of necessity were again specified, but the appraisal phase was given as the primary place where such elements would be brought together and weighed.

At the end of 2006, then, the necessity criterion was operationalised differently in different places. Most notable were the introduction of individual severity of illness, preferably quantified, and the specification of the Dunning committee's for own account as individual cost and/or common use. In addition, the reports suggest a host of different argumentations pertaining to solidarity, the cultural context, and public health. However, the difficulty of operationalising these is also commented on in several places. Moreover, we see an acknowledgement that criteria acting as filters or sieves does not work: criteria should be weighed concertedly and without hierarchy. In particular, the necessity question (the answer to which was considered to be rarely a terse yes or no) should be dealt with in an appraisal.

The following report of note was the second RVZ report, *Just and Durable Care* (Raad voor de Volksgezondheid en Zorg, 2007). It was, as far as I am aware, the first to propose the installation of a specific committee at CVZ: an appraisal committee. This report was followed by *Package management in Practice 2* (College voor Zorgverzekeringen, 2009). Necessity was defined the same as in

the first PiP report, but the criterion here comprised first, severity of illness, and second, necessity of insurance. The first element remained largely the same, but this second element deserves a closer look. It was defined as "whether it is necessary or due from a societal perspective to insure a health care intervention." The report continues,

> Experience teaches that it is difficult to judge during the assessment whether a health care intervention is necessary to insure. These considerations fit, after all, in the societal debate (appraisal). (College voor Zorgverzekeringen, 2009, p. 18)

Further on, the report states that the effectiveness and cost-effectiveness calculations will be weighed against the outcome on the rating on feasibility, the individual severity of illness, expressed as a percentage of individual loss of health for a patient due to this disease, and 'necessity of insurance'.

> This rating yields arguments in favour of or against incorporation into the [basic benefits] package. Arguments against incorporation will lower the chance of a positive advice; arguments in favour of incorporation into the package will heighten the chance of a positive advice. Examples of arguments in favour include: rarity of the disease (orphan indications) combined with a lack of alternative treatment options, informal care (a high level of informal care is given to the patient, which puts a high pressure on the environment of the patient), and risks for public health outside of the patient. Examples of arguments against incorporation into the package are: little overlap with the domain of health care, a high total budget impact, unsuitability of insurance due to high prevalence, and unsuitability of insurance due to high autonomy of the patient. The entirety of these argumentations determines the outcome of the advice. (College voor Zorgverzekeringen, 2009, p. 22)

In this report, the role of the appraisal committee (in Dutch: *Adviescommissie Pakket*, ACP) is also worked out for the first time by CVZ. This committee is specifically tasked with weighing the comments on the concept advice, making an inventory of the considerations, and determining the direction of the advice and priorities. Moreover, the report states, there should not be a sharp line between the assessment and appraisal phases.

By 2009, then, two significant steps had been taken. First, on the operationalisation of the necessity criterion side, we see the specification of Dunning's for own account in terms of insurance logic, where a combination of high individual cost and low risk indicates that *insuring* a health care technology makes sense. Second, on the procedural side, RVZ's suggestion to install an appraisal committee had been followed by CVZ.

Package management in Practice 3 was published another four years later (Zorginstituut Nederland, 2013). The funnel of Dunning is now named explicitly, with 'own risk and responsibility' as part of the necessity of insurance element of the necessity criterion (Zorginstituut Nederland, 2013,

p. 32). For the severity of illness element, the proportional shortfall method[1] was introduced, and the reasons why this was chosen: because the societal *opinions* on the distribution of health were now also considered of importance. Examples of such opinions may be that there may be no age discrimination, that people with an immediate need should be helped first, or that those with the worst health condition should receive most care (Zorginstituut Nederland, 2013, p. 33).

The Institute for Medical Technology & Assessment (iMTA) and the Institute for Health Policy & Management (in Dutch: *instituut Beleid en Management van de Gezondheidszorg*, iBMG) worked together with CVZ to develop the 'necessity of insurance' element, as in PiP2 it was noted to be in need of further operationalisation. Niëns and colleagues developed a '2x4 checklist', based on the Institute's assessment framework for medical care (in Dutch: *beoordelingskader hulpmiddelenzorg*):

Health insurance as an instrument

1. Is the intervention customary care?
2. Is the intervention foreseeable?
3. Might there be under-usage of an intervention if it is not insured?
4. Might there be over-usage of an intervention if it is insured (moral hazard)?

Financial accessibility

5. Does the intervention substitute for something that the majority of the population also uses?
6. Can the (additional) treatment costs be borne by the individual patient?
7. Can the patient expect relevant savings (offsetting the costs) due to the intervention?
8. Are treatment costs incurred only once or are they structural in character? (Niëns, 2014)

These questions, according to PiP3, were to be asked from an "insurance perspective" (Zorginstituut Nederland, 2013, p. 37). The first set of questions concerns the chance that something happens and the risk of moral hazard (over-usage), the second set considers the financial impact this may have for the individual. Questions 1 and 6 were expanded on in the report. For question 1, the report states: "This question is meant to delineate the insured care from the usual course of events in society." It specifies that if it is first, a generally customary provision, it will not be insured (such as a braille watch). Second, when it is customary care, it concerns the usual care that partners, inhabiting parents, or other house mates usually give one another. For question 6, the report notes

1 Proportional shortfall is a method currently in use in Dutch coverage decision-making practice to quantify the 'necessity of care' for a certain health care technology. Its objective is to create more equity in terms of severity of illness and (prospective or past) health than counting all Quality-adjusted Life Years (QALYs) as equal, as the latter does not reflect wider (societal) notions in terms of treating those in greatest need first. It seeks to do so through quantifying the proportion of QALYs patients lost due to the disease without the technology compared to the remaining QALYs these patients would have had without this disease (Reckers-Droog et al., 2018; Stolk et al., 2004). It is currently used in practice in setting reference values for cost-effectiveness thresholds (Reckers-Droog et al., 2019; Zorginstituut Nederland, 2018).

that it has been difficult in practice to state a maximum amount, and that this is perhaps even undesirable. It also holds that "whether certain package proposals exceed the capacity of citizens is, ultimately, a political consideration. From insurance theory, it is a legitimate question whether the costs of provision outweigh the costs of insurance." The report argues that for low costs, insurance is not indicated, because if it is not provided through the benefits package "the market will do its job, in the good sense of the word" (Zorginstituut Nederland, 2013, p. 38). This, the report continues, should be regarded in cohesion with other considerations.

PiP3 thus further operationalised the necessity criterion's first element, severity of illness, as calculated by means of the proportional shortfall method, and the second element, necessity of insurance, as a 2x4 checklist. Notably, these highly specified considerations should still be considered in cohesion with other argumentations but not much attention is given in PiP3 as to how this should take place.

Summarising, the two strands of thought that have dominated these policy reports have both evolved over the years. The substantive operationalisation of necessity has moved from relatively broadly defined to a high degree of specification in the shape of a complex calculation and a checklist. Over the years, the specified criterion has contained several elements, the most transient of which seems to have been solidarity and public health argumentations, though an acknowledgement of the context, be it cultural or otherwise, appears regularly. Second, there has been an almost continual appreciation of deliberative settings for the establishment of necessity so as to be able to weigh it concurrently with other criteria (and other argumentations), with necessity relatively quickly losing its primacy as the first 'sieve' in Dunning's funnel. Establishment of necessity or answering the necessity question, in particular, has over the years been specified as best taking place in the appraisal phase of decisions by a separate appraisal committee.

DUTCH COVERAGE DECISION-MAKING PRACTICE

Current Dutch coverage decision practice, follows the generic assessment-appraisal pattern outlined above, with a few additional steps. After a form of health care has been set on the agenda as a topic requiring consideration from the Institute, a scoping session is initiated, in which stakeholders are invited to contribute relevant considerations. Next, a wide variety of knowledge types is gathered by Institute employees, which are subsequently assessed by the assessment committee (in Dutch: *Wetenschappelijke Adviesraad*, WAR). This is written up into a 'discussion document' which features headings per criterion and other relevant considerations to benefit the appraisal committee (in Dutch: *Adviescommissie Pakket*, ACP). The appraisal aims for an explicit "societal weighing" of the knowledge established in the assessment (art. 14, Zorginstituut Nederland, 2016, 2017), resulting in an advised decision. This advised decision is discussed and approved by the Institute's board of directors and sent to the Minister.

To arrive at this advice, a combination of both Health Technology Assessment (HTA) and Accountability for Reasonableness (A4R) is reportedly used (Zorginstituut Nederland, 2017). HTA and A4R are both institutionalised frameworks aiming to benefit coverage decision-making practice and both have been extensively refined and their workings in practice studied by scientists. They are relevant as the two strands of thought I distinguished in the policy documents above have each been resonated by, and probably influenced by, these two scientific fields. On the one hand, the idea of providing well-established, consistently-applied substantive criteria for making decisions has been advocated by scholars in the field of Health Technology Assessment. Scholars working on Accountability for Reasonableness, on the other hand, would consider good procedures the primary guarantee for good decisions. Below, I give a brief introduction to both[2].

HTA aims to provide a "systematic evaluation of the properties and effects of a health technology" (INAHTA, 2020). It does so assuming that the total budget a country has available to spend on health care technologies is limited and it seeks to identify which technologies provide value for money as a consequence (Lehoux, 2014). HTA was at its inception defined as the evaluation of both the technical side of a health care technology as well as the societal impact, that is, both assessment and appraisal. In practice, HTA has largely been narrowed down to effectiveness and cost-effectiveness calculations to benefit the assessment phase (Giacomini, 1999; Lehoux & Blume, 2000), despite historic and recent calls to (re)integrate 'ethical issues' into HTA (Daniels et al., 2016; Jansen et al., 2017). HTA remains a hugely impactful enterprise and these technical assessments are now more and more achieved internationally (Guegan et al., 2014; Stolk et al., 2009). A large part of the work of the Institute currently entails achieving an HTA per health care technology, which feeds into decision making through the assessment phase in particular and serves to 'ground' the advice scientifically (Niezen, 2012). This grounding is directly based on notions of evidence-based medicine (Abrishami, 2017; Lehoux & Blume, 2000).

On the procedural side, Daniels and Sabin developed the Accountability for Reasonableness (A4R) framework in the 1990s based on Rawls' theory of justice as fairness. A4R responded to questions around the legitimacy and fairness of limit-setting decisions made by USA-based Managed Care Organizations and other insurers at the time. A4R sets procedural boundaries for coverage decisions: relevance, publicity, appeals, and enforcement (Daniels, 2000; Daniels & Sabin, 1997, 1998, 2008). These respectively aim to ensure that 1) the rationale for the decision is supported by reasons judged reasonable by "fair-minded people"; 2) this rationale is provided

2 Recent efforts have attempted to bring the two together into a comprehensive framework for 'evidence-informed deliberative processes' (EDPs) (Bærøe & Baltussen, 2014; Baltussen et al., 2016; Baltussen et al., 2017; Oortwijn et al., 2020). Such EDPs should, these authors argue, "learn about the relevant social values" through early stakeholder involvement and evaluate these values in a manner that is informed by evidence. They denote five important implications for the ideal organisation of the processes at HTA agencies. First, these agencies should organise stakeholder involvement well. Second, they should integrate the assessment and appraisal phases. Third, it is suggested that the criteria for making decisions are subjected to public scrutiny. Fourth, HTA agencies are advised to formulate and use a checklist of criteria that are considered potentially relevant and, in the justification or rationale for decisions, outline how each criterion affected this decision. Fifth, the authors suggest making these justifications or rationales public and enabling appeals.

publicly; 3) this decision can be appealed; and 4) the first three criteria are enforced in some way. A4R thus prescribes criteria for coverage decision processes and holds that through satisfaction of these criteria, the process can be considered legitimate. A4R has been considered both acceptable and applicable in a wide variety of decision settings (Daniels & Sabin, 2008; Kapiriri et al., 2009; Kapiriri & Razavi, 2017; Martin et al., 2002). In essence, A4R holds that a decision outcome should be accepted as fair when it has been reached by a fair procedure (Daniels et al., 2016). As opposed to HTA, A4R thus sets procedures over content in terms of achieving health care coverage decisions. A recent policy report specifies how A4R's procedural boundaries are adhered to in practice at the Institute (Zorginstituut Nederland, 2017).

Of the whole coverage decision-making process, this dissertation will focus on the appraisal phase as this is where necessity is located according to the policy documents studied. Specifically, this dissertation will build on descriptive, inductive understandings of what appraisal entails. This will mean not situating it in either the HTA or A4R scientific tradition, as these are in essence both more prescriptive than descriptive, deriving as they do from larger principles for good decision making, namely justice as consistently-applied criteria and justice as fairness respectively. This means that these schools of thought are not primarily concerned with describing what happens in practice but with to what extent practice may adhere to the principles set out. Instead, I will take my cues from inductive studies which understand appraisal to be a deliberative process to come to a decision through interpretation of the knowledge input. These studies describe appraisals as featuring a plethora of different argumentations (Dakin et al., 2015; Franken et al., 2015; Gordon, 2006; Guindo et al., 2012; Jansen et al., 2017; Morrell et al., 2017; Shah, 2009), which may contradict (Martin et al., 2001; Singer et al., 2000), and the applicability of which may be under discussion (Kapiriri et al., 2009; Vuorenkoski et al., 2008). Two schools of thought have considered inductively how these types of decisions are arrived at: Health Services Research and Science and Technology Studies. I will discuss both in turn.

HEALTH SERVICES RESEARCH: ELEGANCE, RATIONALITY, EXPERTISE

This overview on the inductive work on coverage decision making in the Health Services Research (HSR) field starts in the middle of the 1990s. 'Muddling through', a key phrase in this field at this time, denoted decisions made 'not according to general rules', that is, decision criteria. Instead, decisions relied on the discretion of the decision makers (Hunter, 1995; Mechanic, 1997). The term muddling through is borrowed from the political scientist Lindblom (1959) who observed in his landmark study that even when you agree on both the (knowledge) input and the intended goals of a policy, you may still disagree on which policy is most appropriate. Preferences of decision makers vary, and they, therefore, value certain policies or policy instruments differently. This, however, Lindblom considered not so much a problem as a fact of policy life. Since its inception, muddling through has been criticised for resulting in inconsistency between decisions and even in

arbitrariness. Scholars in the coverage decision-making field, however, preferred it over the rigid application of universal decision criteria, which they considered to lead to too little attention for the case at hand (Entwistle et al., 1996; Ham, 1999). They described 'implicit' or individual-level decisions that relied on "discretion, flexibility, and ability to take account of emotions, aspirations, and preferences" (Mechanic, 1997, p. 90). Their answer to the arbitrariness charge was to muddle through *elegantly* with greater transparency concerning the grounds for a decision (Hunter, 1995; Mechanic, 1997). Much of the inductive work that followed this positioning described decision makers' subjectivities, that is, personal preferences and ways of deciding, showcasing "the practical circumstances of real-world decision making" (Hughes & Light, 2002, p. 1).

Since then, the field has moved towards describing the rationality of decisions, specifically favouring a both-and conceptualisation of rational decisions (Calnan et al., 2017; Gkeredakis et al., 2011; Hughes & Doheny, 2011). For rational decisions, decision makers not only take into account "contexts and occasions" but are accountable to criteria, such as those provided by evidence-based medicine, at the same time (Jenkings & Barber, 2004, p. 1765). The application of criteria and the incorporation of scientific knowledge has thus become part and parcel of making rational decisions, rather than being opposed to it. This has required decision makers to negotiate between and ultimately incorporate both those more formal rationalities and more 'local' ones in the decision-making process (Hughes & Doheny, 2011). This has resulted in an overall focus on the work that goes into combining; some scholars describe decision makers as being both 'rational' and 'human' (Russell & Greenhalgh, 2014); others how they display a "combining of strategies" (Calnan et al., 2017; Moreira, 2005). Such combining work has specifically been characterised as being *pragmatic* (Calnan et al., 2017; Hunter, 1995; Russell & Greenhalgh, 2014); resulting in decisions with a 'pragmatic rationality' (Russell, 2017, following a.o. Aristotle). Such rationality is juxtaposed with both instrumental rationality, where a complete set of decision rules guarantees the quality of the decision (embodied by HTA efforts), and institutional rationality, with transparency and good procedure acting as guarantors (as visible in the A4R framework) (Gkeredakis et al., 2011; Ham & Glenn, 2003; Russell, 2017). Pragmatic rationality carries a distinctly positive valuation as "a characteristic of expert judgement" (Russell, 2017, p. 60). Pragmatic rationality is found in the way experts combine not only experiential knowledge and emotional engagement but also scientific evidence and ethical principles for fair processes (Russell, 2017). Expertise or specifically experts who display pragmatic rationality in the rhetorical deliberative setting thus may potentially be held as an alternative guarantor of decision quality.

Concluding, although some authors (Hughes & Doheny, 2011; Jenkings & Barber, 2004) do reference the world outside in the decision-making process, the HSR field primarily focuses its efforts on the dynamics of the deliberative processes and describing the role and embodiment of expertise therein. For this dissertation, these processual dynamics of pragmatic decision making will be of great interest. Based on the overview above, I note three elements: 1) decision-making experts display an understanding of, and are able to work with, different types of knowledges; 2) experts know how to respond in a human way; and 3) experts adhere to formalised procedures. I will,

however, release the emphasis on combining work and instead, focus on the expertise displayed in appraisal specifically and what may distinguish it from earlier studies on coverage decision-making expertise. Moreover, as the use of the necessity criterion is evident not only in deliberation but also in documents containing the justification or rationale for decisions, this study requires more focus in terms of the substance and the outcome of the decision than much current HSR is giving it. To fill this gap, I am turning to Science and Technology Studies.

SCIENCE AND TECHNOLOGY STUDIES: EXPERTISE AND THE WORLD OUTSIDE

Science and Technology Studies (STS) has a long history studying the role of expert advice in 'the greater picture', often a political decision-making process (Collins & Evans, 2008; Fischer, 2011; Frey & Fontana, 1991; King et al., 2018; Rip, 1985). According to Rip (1986, 1992), expert advice should aim to orientate the recipients of the advice in favour of some option for action, involving the world outside with its complexity and uncertainty in this advice (Rip, 1985, p. 95). Rip notes (following Ezrahi, 1980) that when there is no scientific consensus and no agreement on social and political goals, public controversy is likely. The aim should therefore be pragmatic rationalism (see also Rip, 1992), which he defines as:

> Each solution must take account of the circumstances relevant to that unique situation. [Pragmatic rationalism] redefines the goals of expert advice towards stressing its ability to help decision-makers to produce robust outcomes in particular contexts, rather than just in terms of the quality of its scientific and technical content. (Rip, 1985, p. 108)

This definition of pragmatic rationalism differs notably from Russell's. Russell defines pragmatic rationalism as a specific process, a moment in time in which the decision is made, and highlights the skills of experts therein. In contrast, through his definition, Rip underlines the substantive input, the decision outcome, and what happens to the decision afterwards. He defines pragmatic rationalism as a process that is not only localised but in an active relationship with 'contexts'. These contexts affect the decision both through input into the decision as the circumstances to be considered and through testing the 'robustness' of the decision output.

Concerning input into policy decisions, much recent STS work has considered the production of (socially) robust knowledge (Nowotny, 2003; Nowotny et al., 2013; Rip, 2010). Such knowledge, Nowotny et al. hold, is produced by "democratising expertise", which involves opening the door to experts of other kinds, most notably experts-by-experience rather than academic training (cf. Moes, 2019). Context is thus given a voice, "society [is] speaking back to science" (Strathern, 2003), in line with a broader societal trend in many Western countries towards more public participation in policy making (e.g., Jasanoff, 2003). Some scholars have expressed a fear of this resulting in a

collapse of the concept of expertise. Starting with their landmark paper 'The Third Wave of Science Studies', Collins and Evans (2002) have sought to re-draw boundaries around technical-scientific expertise, aiming to explicate its status and role in decision making and sparking subsequent debate within STS (Collins et al., 2010, 2011; Epstein, 2011; Fischer, 2011; Jasanoff, 2003; Rip, 2003).

I would agree with Fischer as he argues that technical expertise inputs in but cannot give the final judgement on decisions affecting the public realm (in line with Lindblom) (Fischer, 2011). Nowotny also gives the many complexities of the social and political world as a major reason for involving experts-by-experience (Nowotny, 2003), thus likewise linking the input to the performance of the output 'out there'. In this, the justification or rationale for a decision likely plays a major role (Bal, 1999). This has also been acknowledged by Collins et al., who noted that political choices based on expert advice ought to be made explicit and public (Collins et al., 2010). As Moreira recommends on coverage decisions, "experts and stakeholders should be able to pre-emptively account for their reasoning to a non-expert audience" (Moreira, 2011, p. 1340). As stated above, the decision justification or rationale should be robust given the places and situations where the decision is to have an effect (Rip, 1985), and knowledge concerning this context is to be crucial in achieving this.

RESEARCH QUESTIONS

This research answers the following main research question: how is the necessity criterion used in practice? The first important observation I made during the early days of my field work, underlined by the policy documents described above, was that the necessity criterion as used in practice took the form of *argumentations*. I will define argumentations as explicated reasons, generally given on paper or in discussion by anyone reasoning about the potential coverage status of a health care technology. Consequently, I set up the following sub-questions:

1. Which primary types of necessity argumentations can be distinguished from the published literature and how does the user and place of use of necessity argumentations affect their use and outcome?

This question, answered in Chapter 1, aims to provide an overview, not comprehensive but to be used as a first insight, of necessity argumentations, to see if a classification can be made that might distinguish several specific argumentation types and if patterns of use may be elucidated based on user and place of use.

2. How are necessity argumentations used in (advised) decision documents in the Netherlands, Belgium, England, and Germany to construct this decision through contextualising the knowledge on a certain health care technology?

Chapter 2 gives the answer to this question, which compares the use of necessity argumentations in (advised) decision documents in Belgium, England, Germany, and the Netherlands to gain insight into how necessity argumentation types are used, for what purpose, and especially how their use compares across four specific case studies.

3. How are necessity argumentations used, both in deliberation and in the advised decision, to 'construct necessity' in the Dutch appraisal committee meetings and what does this say about societal weighing expertise?

For this question, we delved into the setting of the appraisal committee as that is where the use of necessity argumentations such as the ones identified through research question 1 were audible in practice. This question is answered in Chapter 3.

4. How are necessity argumentations used to contribute to a robust advised decision and what may be done to make such a decision more robust?

In Chapter 4, a subset of data from Chapter 3 was re-analysed to formulate a conceptual model for arriving at robust decisions.

Together, these four questions aim to clarify how the necessity criterion is used by examining usage-in-practice as described in scholarly literature, in different countries, in the Dutch appraisal specifically, and from a theoretical stance.

SET-UP AND METHODOLOGY

Chapter 1 comprises a realist review, that answers research question 1. The realist review is an established methodology used to review literatures and to gain insight into technologies-in-context. My team has extended the applicability of this method by applying it to text rather than technology – we are, as far as we know, the first to do so. The realist review method asks: what works, for whom, under what circumstances? This is in fact not difficult to apply to argumentations, as they, like technologies, are used in a particular place to achieve a particular outcome (Pawson et al., 2005). The realist review method was thus chosen to enlarge our understanding of not only the variety of potential argumentations but of their usage, as we hypothesised that not all argumentations would be used in the same way. We chose to review scholarly literature rather than policy documents as much of the previous scholarly work on necessity had in fact built on the latter. Moreover, we were hoping to provide insight into more than the argumentation types used in formalised decisions only. In terms of the review, particularly the identification of the argumentation types was an extended process methodologically. There are iterations of the Excel file that have three argumentation types (Burden of Disease, Necessity of Insurance, and Solidarity, dated 23 March

2016), and there are iterations that have more than the twenty we arrived at (e.g., Vulnerability/Compassion is a combined argumentation type in the final list, as these turned out to be difficult to distinguish in some texts). Chapter 1, then, provides an overview of the list of necessity argumentation types retrieved from scholarly literature, and second, zones in on the patterns of use of these argumentations.

Chapter 2 answers research question 2 by offering a comparative analysis of the use of necessity in decision-making processes generally and in four (advised) decisions specifically across four countries in Western Europe: Belgium, England, Germany, and the Netherlands. These countries were chosen as they share certain 'health care system objectives'; equity, affordability, and transparent decision-making among them (Franken et al., 2012). The reason for examining not just Dutch decision-making practices but comparing them to those in three other countries was that comparative analyses are considered helpful for comparing arguments and statements specifically, resulting in clear definitions and succinct follow-up research questions (Deville et al., 2016). To gain understanding of the general decision-making practices, we started with semi-structured, active group interviews (Holstein & Gubrium, 2016), which were augmented by document and web site analysis, with the results member-checked. This member check also resulted in the decision to additionally examine several decisions that were made in all four countries. Employing the case approach promises insight into patterns of social behaviour specifically (Creswell & Poth, 2017; Ragin, 2004) and is as such a good match for this type of research. The selection of cases adhered to specific criteria formulated by all the collaborators on that chapter (for more details, please see Chapter 2). Importantly, the cases could include any decision as long as a decision document (containing a justification or rationale for the decision) was available, as only this paper was to be examined. The cases that were ultimately selected represented a relatively wide variety of health care technologies, as we hypothesised that would also contribute to the validity of our findings, in addition to the decision outcomes per case varying across the countries. Throughout the analysis of the documentation pertaining to the four selected cases in the four countries, I have used the list of necessity argumentation types as 'sensitising concepts' (Bulmer, 1979) to gain insight into how necessity argumentations are used in coverage decision practice. A sensitising concept is used to make the researcher (more) aware of certain dynamics, and in my case, it guides the analysis through acting as a code. I coded the documents with these twenty necessity argumentation types as codes, transferring them to Microsoft Excel so as to see when, where, and by whom the different argumentation types were used in the documents studied. This has enabled insight not so much in the contents of the argumentations but in the patterns of use to be compared across the four countries.

Chapter 3 will answer research question 3 by homing in on the Dutch appraisal phase and analysing how this appraisal committee interprets its role in terms of societal weighing. Question 3 asks how necessity is constructed, what societal weighing looks like. Chapters 3 and 4, both examine the Dutch appraisal practices specifically, as this is where both the policy documents and Chapter 2 expressly located the use of necessity argumentations. These chapters build in part on

the same cases (namely eculizumab and paracetamol-vitamin D tablets), and case selection was more time-constrained than for Chapter 2. The reason is that I had to have been present at the appraisal committee meeting or at least be able to listen to the audio file, with the former having a strong preference (in the end, only the maternity case happened before my presence at appraisal committee meetings; this case was selected as interviewees stated it would be a particularly fruitful case in terms of studying the necessity criterion). Moreover, the cases were purposely chosen to represent a wide variety of health care technologies. Like in Chapter 2, I coded the transcribed audio files and the documents (in this case, the discussion documents provided as input for the appraisal meeting as well as the decision documents containing the final decision and the letter to the Minister) using the necessity argumentation types formulated in Chapter 1. These data were triangulated by interviews with Institute employees and appraisal committee members.

Sensitisation to the robustness of decisions through appreciation of what public controversy might engender is the way into Chapter 4. This chapter seeks to answer research question 4 to focus more on theorising on how these decisions are brought together in terms of content, asking the question: how are robust coverage decisions made? In this chapter, instead of choosing a primarily data-driven approach, I chose to profit from theoretical work within STS on controversy and robustness to come to a model for making robust decisions in appraisal. This model is illustrated by data from Chapter 3.

REFERENCES

Abrishami, P. (2017). Public Value of Medical Innovations. *A quest for all and for all seasons (PhD-thesis). Maastricht/Diemen: Maastricht University & Zorginstituut Nederland.*

Bærøe, K., & Baltussen, R. (2014). Legitimate Healthcare Limit Setting in a Real-World Setting: Integrating Accountability for Reasonableness and Multi-Criteria Decision Analysis. *Public Health Ethics, 7*(2), 98-111. doi: 10.1093/phe/phu006

Bal, R. A. (1999). Grenzenwerk; over Het Organiseren Van Normstelling Voor De Arbeidsplek.

Baltussen, R., Jansen, M. P., Mikkelsen, E., Tromp, N., Hontelez, J., Bijlmakers, L., & Van Der Wilt, G. J. (2016). Priority Setting for Universal Health Coverage: We Need Evidence-Informed Deliberative Processes, Not Just More Evidence on Cost-Effectiveness. *International Journal of Health Policy and Management, 5*(11), 615-618. doi: 10.15171/ijhpm.2016.83

Baltussen, R., Jansen, M. P. M., Bijlmakers, L., Grutters, J., Kluytmans, A., Reuzel, R. P., . . . Van Der Wilt, G. J. (2017). Value Assessment Frameworks for Hta Agencies: The Organization of Evidence-Informed Deliberative Processes. *Value in Health, 20*(2), 256-260.

Bulmer, M. (1979). Concepts in the Analysis of Qualitative Data. *The Sociological Review, 27*(4), 651-677.

Calnan, M., Hashem, F., & Brown, P. (2017). Still Elegantly Muddling Through? Nice and Uncertainty in Decision Making About the Rationing of Expensive Medicines in England. *International Journal of Health Services, 47*(3), 571-594.

College Voor Zorgverzekeringen. (2006). Pakketbeheer in De Praktijk. Diemen.

College Voor Zorgverzekeringen. (2009). *Pakketbeheer in De Praktijk 2*. Diemen: College voor Zorgverzekeringen.

Collins, H., & Evans, R. (2002). The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science, 32*(2), 235-296.

Collins, H., & Evans, R. (2008). *Rethinking Expertise*: University of Chicago Press.

Collins, H., Weinel, M., & Evans, R. (2010). The Politics and Policy of the Third Wave: New Technologies and Society. *Critical policy studies, 4*(2), 185-201.

Collins, H., Weinel, M., & Evans, R. (2011). Object and Shadow: Responses to the Cps Critiques of Collins, Weinel and Evans,'Politics and Policy of the Third Wave'. *Critical policy studies, 5*(3), 340-348.

Commissie-Dunning. (1991). Kiezen En Delen. *Rapport van de Commissie Keuzen in de Zorg*.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative Inquiry and Research Design: Choosing among Five Approaches*. USA: Sage publications.

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'neill, P., & Parkin, D. (2015). The Influence of Cost-Effectiveness and Other Factors on Nice Decisions. *Health Economics, 24*(10), 1256-1271. doi: doi:10.1002/hec.3086

Daniels, N. (2000). Accountability for Reasonableness. *Establishing a fair process for priority setting is easier than agreeing on principles, 321*(7272), 1300-1301. doi: 10.1136/bmj.321.7272.1300

Daniels, N., Porteny, T., & Urrutia, J. (2016). Expanded Hta: Enhancing Fairness and Legitimacy. *International Journal of Health Policy and Management, 5*(1), 1-3. doi: 10.15171/ijhpm.2015.187

Daniels, N., & Sabin, J. (1997). Limits to Health Care: Fair Procedures, Democratic Deliberation, and the Legitimacy Problem for Insurers. *Philosophy & Public Affairs, 26*(4), 303-350. doi: 10.1111/j.1088-4963.1997.tb00082.x

Daniels, N., & Sabin, J. (1998). The Ethics of Accountability in Managed Care Reform. *Health Affairs, 17*(5), 50-64. doi: 10.1377/hlthaff.17.5.50

Daniels, N., & Sabin, J. (2008). Accountability for Reasonableness: An Update. *Bmj, 337*. doi: 10.1136/bmj.a1850

Deville, J., Guggenheim, M., & Hrdličková, Z. (2016). Introduction: The Practices and Infrastructures of Comparison.

Entwistle, V. A., Watt, I. S., Bradbury, R., & Pehl, L. J. (1996). Media Coverage of the Child B Case. *Bmj, 312*(7046), 1587-1591.

Epstein, S. (2011). Misguided Boundary Work in Studies of Expertise: Time to Return to the Evidence. *Critical policy studies, 5*(3), 323-328.

Ezrahi, Y. (1980). Utopian and Pragmatic Rationalism: The Political Context of Scientific Advice. *Minerva, 18*(1), 111-131.

Fischer, F. (2011). The 'Policy Turn'in the Third Wave: Return to the Fact–Value Dichotomy? *Critical policy studies, 5*(3), 311-316.

Franken, M., Le Polain, M., Cleemput, I., & Koopmanschap, M. (2012). Similarities and Differences between Five European Drug Reimbursement Systems. *International Journal of Technology Assessment in Health Care, 28*(4), 349-357. doi: 10.1017/S0266462312000530

Franken, M., Stolk, E., Scharringhausen, T., De Boer, A., & Koopmanschap, M. (2015). A Comparative Study of the Role of Disease Severity in Drug Reimbursement Decision Making in Four European Countries. *Health Policy, 119*(2), 195-202. doi: https://doi.org/10.1016/j.healthpol.2014.10.007

Frey, J. H., & Fontana, A. (1991). The Group Interview in Social Research. *The Social Science Journal, 28*(2), 175-187.

Gezondheidsraad. (2003). Contouren Van Het Basispakket. The Hague.

Giacomini, M. K. (1999). The Which-Hunt: Assembling Health Technologies for Assessment and Rationing. *Journal of health politics, policy and law, 24*(4), 715-758.

Gkeredakis, E., Swan, J., Nicolini, D., & Scarbrough, H. (2011). *Rational Judgement Revisited: Practices of Deliberation in Healthcare Funding Decisions.* Paper presented at the OLKC conference, Hull University Business School.

Gordon, E. J. (2006). The Political Contexts of Evidence-Based Medicine: Policymaking for Daily Hemodialysis. *Social Science & Medicine, 62*(11), 2707-2719. doi: https://doi.org/10.1016/j.socscimed.2005.11.024

Guegan, E. W., Huić, M., & Teljeur, C. (2014). Eunethta: Further Steps Towards European Cooperation on Health Technology Assessment. *International Journal of Technology Assessment in Health Care, 30*(5), 475-477.

Guindo, L. A., Wagner, M., Baltussen, R., Rindress, D., Van Til, J., Kind, P., & Goetghebeur, M. M. (2012). From Efficacy to Equity: Literature Review of Decision Criteria for Resource Allocation and Healthcare Decisionmaking. *Cost effectiveness and resource allocation, 10*(1), 9. doi: 10.1186/1478-7547-10-9

Ham, C. (1999). Tragic Choices in Health Care: Lessons from the Child B Case. *Bmj, 319*(7219), 1258-1261. doi: 10.1136/bmj.319.7219.1258

Ham, C., & Glenn, R. (2003). *Reasonable Rationing*: McGraw-Hill Education (UK).

Helderman, J.-K., De Kruijf, J., Verhey, J., & Thiel, S. V. (2014). Dike-Reeve of the Health Care Polder. A Political-Sociological Analysis of the Realisation of the National Health Care Institute against a Backdrop of a Changing Policy Agenda and Changing Political-Dministrative and Societal Relations.

Holstein, J. A., & Gubrium, J. F. (2016). *Narrative Practice and the Active Interview* (Vol. 67): Sage London.

Hughes, D., & Doheny, S. (2011). Deliberating Tarceva: A Case Study of How British Nhs Managers Decide Whether to Purchase a High-Cost Drug in the Shadow of Nice Guidance. *Social Science & Medicine, 73*(10), 1460-1468.

Hughes, D., & Light, D. (2002). *Rationing: Constructed Realities and Professional Practices*. Oxford, UK: Wiley-Blackwell.

Hunter, D. J. (1995). Rationing: The Case for "Muddling through Elegantly". *Bmj, 311*(7008), 811.

Inahta. (2020). What Is Health Technology Assessment (Hta)?

Jansen, M. P., Helderman, J.-K., Boer, B., & Baltussen, R. (2017). Fair Processes for Priority Setting: Putting Theory into Practice : Comment on "Expanded Hta: Enhancing Fairness and Legitimacy". *International Journal of Health Policy and Management, 6*(1), 43-47. doi: 10.15171/ijhpm.2016.85

Jasanoff, S. (2003). Breaking the Waves in Science Studies: Comment on Hm Collins and Robert Evans,the Third Wave of Science Studies'. *Social Studies of Science, 33*(3), 389-400.

Jenkings, K. N., & Barber, N. (2004). What Constitutes Evidence in Hospital New Drug Decision Making? *Social Science & Medicine, 58*(9), 1757-1766.

Kalf, R. R., Makady, A., Ten Ham, R. M., Meijboom, K., Goettsch, W. G., & Workpackage, O. B. O. I.-G. (2018). Use of Social Media in the Assessment of Relative Effectiveness: Explorative Review with Examples from Oncology. *JMIR cancer, 4*(1), e11.

Kapiriri, L., Norheim, O. F., & Martin, D. K. (2009). Fairness and Accountability for Reasonableness. Do the Views of Priority Setting Decision Makers Differ across Health Systems and Levels of Decision Making? *Social Science & Medicine, 68*(4), 766-773. doi: https://doi.org/10.1016/j.socscimed.2008.11.011

Kapiriri, L., & Razavi, D. (2017). How Have Systematic Priority Setting Approaches Influenced Policy Making? A Synthesis of the Current Literature. *Health Policy, 121*(9), 937-946. doi: https://doi.org/10.1016/j.healthpol.2017.07.003

King, N., Horrocks, C., & Brooks, J. (2018). *Interviews in Qualitative Research*. UK: SAGE Publications Limited.

Lehoux, P. (2014). *The Problem of Health Technology*. USA: Routledge.

Lehoux, P., & Blume, S. (2000). Technology Assessment and the Sociopolitics of Health Technologies. *Journal of health politics, policy and law, 25*(6), 1083-1120.

Lindblom, C. E. (1959). The Science of Muddling Through. *PublicAdministrationReview, 19*(2), 79-88.

Makady, A., Ten Ham, R., De Boer, A., Hillege, H., Klungel, O., & Goettsch, W. (2017). Policies for Use of Real-World Data in Health Technology Assessment (Hta): A Comparative Study of Six Hta Agencies. *Value in Health, 20*(4), 520-532.

Martin, D., Giacomini, M., & Singer, P. (2002). Fairness, Accountability for Reasonableness, and the Views of Priority Setting Decision-Makers. *Health Policy, 61*(3), 279-290.

Martin, D., Pater, J., & Singer, P. (2001). Priority-Setting Decisions for New Cancer Drugs: A Qualitative Case Study. *The Lancet, 358*(9294), 1676-1681. doi: https://doi.org/10.1016/S0140-6736(01)06714-9

Mechanic, D. (1997). Muddling through Elegantly: Finding the Proper Balance in Rationing. *Health Affairs, 16*(5), 83-92.

Moes, F. (2019). *Experts and Idiots: An Ethnography of Epistemic Practices in Dutch Social Health Insurance*: Maastricht University.

Moes, F., Houwaart, E., Delnoij, D., & Horstman, K. (2016). Contested Evidence: A Dutch Reimbursement Decision Taken to Court. *Health Economics, Policy and Law, 12*(3), 325-344. doi: 10.1017/S1744133116000281

Moreira, T. (2005). Diversity in Clinical Guidelines: The Role of Repertoires of Evaluation. *Social Science & Medicine, 60*(9), 1975-1985.

Moreira, T. (2011). Health Care Rationing in an Age of Uncertainty: A Conceptual Model. *Social Science & Medicine, 72*(8), 1333-1341.

Morrell, L., Wordsworth, S., Fu, H., Rees, S., & Barker, R. (2017). Cancer Drug Funding Decisions in Scotland: Impact of New End-of-Life, Orphan and Ultra-Orphan Processes. *BMC health services research, 17*(1), 613.

Niëns, L. (2014). Affordability in Health Care: Operationalizations and Applications in Different Contexts.

Niezen, M. M. (2012). *Working Towards Legitimacy in Decision Making: On Ogoverning Appropriate Medicine Use and Reimbursement in Health Care*.

Nowotny, H. (2003). Democratising Expertise and Socially Robust Knowledge. *Science and public policy, 30*(3), 151-156.

Nowotny, H., Scott, P. B., & Gibbons, M. T. (2013). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, UK: John Wiley & Sons.

Oliver, A., Mossialos, E., & Robinson, R. (2004). Health Technology Assessment and Its Influence on Health-Care Priority Setting. *International Journal of Technology Assessment in Health Care, 20*(1), 1-10. doi: 10.1017/S026646230400073X

Oortwijn, W., Jansen, M., & Baltussen, R. (2020). Use of Evidence-Informed Deliberative Processes by Health Technology Assessment Agencies around

the Globe. *International Journal of Health Policy and Management, 9*(1), 27.

Patera, N., & Wild, C. (2014). Assessment – Appraisal – Decision. Lbi-Hta Decision Support Document Nr.: 72. Wien: Ludwig Boltzmann Institut für Health Technology Assessment.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist Review–a New Method of Systematic Review Designed for Complex Policy Interventions. *Journal of Health Services Research & Policy, 10*(suppl 1), 21-34.

Poley, M. J., Stolk, E.A., Brouwer, W. B. F., Van Busschbach,J. J. . (2002). Ziektelast Als Uitwerking Van Het Criterium 'Noodzakelijkheid' Bij Het Makenvan Keuzen in De Zorg. *Ned Tijdschr Geneeskd, 146*(4), 2312-2315.

Raad Voor De Volksgezondheid En Zorg. (2006). Zinnige En Duurzame Zorg. Zoetermeer.

Raad Voor De Volksgezondheid En Zorg. (2007). Rechtvaardige En Duurzame Zorg. Zoetermeer.

Ragin, C. C. (2004). Turning the Tables: How Case-Oriented Research Challenges. *Rethinking social inquiry: Diverse tools, shared standards, 123*.

Reckers-Droog, V., Van Exel, J., & Brouwer, W. (2018). Looking Back and Moving Forward. *Health Policy*.

Reckers-Droog, V., Van Exel, J., & Brouwer, W. (2019). Equity Weights for Priority Setting in Healthcare: Severity, Age, or Both? *Value in Health, 22*(12), 1441-1449.

Rip, A. (1985). Experts in Public Arenas. *Regulating industrial risks: Science, hazards and public protection*, 94-110.

Rip, A. (1986). Controversies as Informal Technology a Ssessment. *Knowledge, 8*(2), 349-371.

Rip, A. (1992). Expert Advice and Pragmatic Rationality. *The Culture and power of knowledge: inquiries into contemporary societies*, 363.

Rip, A. (2003). Constructing Expertise: In a Third Wave of Science Studies? *Social Studies of Science, 33*(3), 419-434.

Rip, A. (2010). Social Robustness and the Mode 2 Diagnosis. *Science, Technology & Innovation Studies, 6*(1). doi: http://dx.doi.org/10.17877/DE290R-13458

Russell, J. (2017). *The Rationality of Rationing: A Rhetorical Policy Analysis of Deliberations About Resource Allocation in the Nhs.* University of Oxford, UK.

Russell, J., & Greenhalgh, T. (2014). Being 'Rational'and Being 'Human': How National Health Service Rationing Decisions Are Constructed as Rational by Resource Allocation Panels. *Health:, 18*(5), 441-457.

Russell, J., Swinglehurst, D., & Greenhalgh, T. (2014). 'Cosmetic Boob Jobs' or Evidence-Based Breast Surgery: An Interpretive Policy Analysis of the Rationing of 'Low Value'treatments in the English National Health Service. *BMC health services research, 14*(1), 413.

Shah, K. K. (2009). Severity of Illness and Priority Setting in Healthcare: A Review of the Literature. *Health Policy, 93*(2-3), 77-84.

Singer, P. A., Martin, D., Giacomini Bhatia, M. V., & Purdy, L. (2000). Priority Setting for New Technologies in Medicine: Qualitative Case Study. *Bmj, 321*(7272), 1316-1318. doi: 10.1136/bmj.321.7272.1316

Stolk, E., Brouwer, W., & Busschbach, J. (2002). Rationalising Rationing: Economic and Other Considerations in the Debate About Funding of Viagra. *Health Policy, 59*(1), 53-63.

Stolk, E., De Bont, A., Van Halteren, A., Bijlmer, R., & Poley, M. (2009). Role of Health Technology Assessment in Shaping the Benefits Package in the Netherlands. *Expert review of pharmacoeconomics & outcomes research, 9*(1), 85-94.

Stolk, E., Van Donselaar, G., Brouwer, W., & Busschbach, J. (2004). Reconciliation of Economic Concerns and Health Policy. *PharmacoEconomics, 22*(17), 1097-1107.

Strathern, M. (2003). Re-Describing Society. *Minerva, 41*(3), 263-276.

Vuorenkoski, L., Toiviainen, H., & Hemminki, E. (2008). Decision-Making in Priority Setting for Medicines—a Review of Empirical Studies. *Health Policy, 86*(1), 1-9.

Walley, T. (2007). Health Technology Assessment in England: Assessment and Appraisal. *Medical Journal of Australia, 187*(5), 283.

Wiering, B., De Boer, D., & Delnoij, D. (2017). Patient Involvement in the Development of Patient-Reported Outcome Measures: A Scoping Review. *Health Expectations, 20*(1), 11-23.

Zorginstituut Nederland. (2013). Pakketbeheer in De Praktijk Deel 3 (pp. 1-106). Diemen.

Zorginstituut Nederland. (2016). Reglement Adviescommissie Pakket. In Z. Nederland (Ed.). https://www.zorginstituutnederland.nl/over-ons/publicaties/besluit/2016/12/05/reglement-adviescommissie-pakket-zorginstituut-nederland (accessed 11 February 2020).

Zorginstituut Nederland. (2017). Pakketadvies in De Praktijk: Wikken En Wegen Voor Een Rechtvaardig Pakket. Diemen.

Zorginstituut Nederland. (2018). Ziektelast in De Praktijk. Diemen.

Zorginstituut Nederland. (2020). Academische Werkplaats Hta&A. Retrieved 28 January, 2020