Valuation Studies 2(1) 2014: 43–64

Uncertainty and the Development of Evidence-Based Guidelines

Esther van Loon* and Roland Bal

Abstract

This article explores how developers address uncertainty in the creation of an evidence-based guideline (EBG). As the aim of an EBG is to assist healthcare practitioners in situations of doubt, it is easy to assume that uncertainty has no place in guidelines. However, as we discovered, guideline development does not ignore uncertainty but seeks to accept it while establishing credible recommendations for healthcare. Dealing with omissions in knowledge, ignorance, or challenges in valuating different sorts of knowledge form the core of the work of guideline developers. Interviewing guideline developers, we found three types of valuation work: classifying studies, grading types of knowledge, and involving expertise and clinical practice. These methods have consequences for the credibility, and amount and kind of uncertainty EBGs can include.

Key words: valuation; uncertainty; evidence-based guidelines; development

Introduction

> With a background in science, you are used to thinking you know it all. Ask me something about a disease and I'll tell you all about it. But I can't tell you what I don't know. I think we need to make that more transparent, that we also don't know a lot. (Guideline developer involved in guidelines for infectious diseases)

Developing an evidence-based guideline (EBG) is a process of valuating and bringing order into a plethora of knowledge. Guideline making is collective work in which core issues are related to what knowledge is available, how this knowledge should be valued, which actors should be involved in the process, and how recommendations can be justified

Esther van Loon, Erasmus University Rotterdam, Institute of Health Policy and Management, the Netherlands, vanloon@bmg.eur.nl

LiU Electronic Press, DOI 10.3384/vs.2001-5992.142143
http://valuationstudies.liu.se

(Moreira 2005; Moreira, May, and Bond 2009; van Loon, Zuiderent-Jerak, and Bal 2013). This valuation inevitably meets uncertainty. Yet, at face value, uncertainty contradicts the EBG movement. EBGs are developed to provide recommendations that assist healthcare workers make the right decisions about patient care. These recommendations are based upon "a systematic review of the evidence and an assessment of the benefits and harms of alternative care options." (Graham et al. 2011, 4). The rhetoric of EBG is that guidelines provide *certainty* for healthcare workers who are faced with patients with ambiguous complaints and treatment choices with unpredictable outcomes. That such strong rhetoric works is understandable, as healthcare workers are increasingly held accountable for their decisions. Decision-making in healthcare has become more complex due to increased options for treatment and increased awareness of diseases. Yet, the idea that guidelines are free of uncertainty or the solution to clinical uncertainty is not realistic. Timmermans and Angell, for example, have shown that using EGBs in the socialisation of doctors sometimes helps to solve clinical uncertainty, but it also reproduces new kinds of uncertainty that need to be dealt with accordingly (Timmermans and Angell 2001). Uncertainty thus remains an aspect of clinical work, despite EBGs.

Rather than focusing on uncertainties in clinical work, in this paper we focus on uncertainties inherent in creating EBGs. We are interested in finding out how uncertainty manifests itself in this process, and what kind of valuation work is undertaken to engage with uncertainty. Valuation work is the social practice of bringing order into all kinds of information and signifying or giving worth to this information (Kjellberg and Mallard 2013; Helgesson and Muniesa 2013). It involves both the assessment of values (i.e., literature, opinion, expertise) and the reproduction of values into recommendations for EBGs. Uncertainty is an inevitable element in this process. We suggest that uncertainty in guidelines is not always detrimental. Uncertainty invokes reflection, and as we have discussed elsewhere, reflexivity in healthcare standards help practitioners to achieve good care (van Loon and Zuiderent-Jerak 2011). Yet expressing uncertainty makes one vulnerable. Therefore, as Gross puts it, "the challenge is how to knowingly and increasingly also publicly deal with what is not known without losing one's credibility or 'scientific authority'." (Gross 2010, 3). The focus in this article is on how the EBG can balance between recognising and accepting uncertainty while producing reliable and credible recommendations to guide healthcare practitioners. Our research question is: *How is valuation work done to balance between acknowledging uncertainty and remaining credible in guideline development?*

To answer this question, we held semi-structured interviews with Dutch guideline developers from a wide range of healthcare organisations. The interviews focused on the struggles, debates, and

valuation work of guideline developers in striving to create reliable and realistic recommendations and engage with uncertainty.

The remainder of this article is structured as follows: first, we define uncertainty and distinguish three ways in which it manifests itself in guideline development. After elaborating on the methods, we provide an analysis of our empirical findings. We discuss three valuation practices in creating EBGs: classifying studies, grading different types of knowledge and those involving expertise and clinical practice. We conclude by showing that different valuation practices have different consequences for acknowledging uncertainty.

On Uncertainty

Uncertainty gains a great deal of attention in the social science literature. Studies of decision-making on environmental issues, the practice of futurists, public involvement in science and health care are some examples (Wynne 1996; Callon, Lascoumes, and Barthe 2009; van Asselt, Mesman, and van 't Klooster 2007; Shackley and Wynne 1996; Mesman 2008). This section clarifies our approach towards uncertainty and discusses three forms in which it manifests itself in relation to the EBG. Further, we pay attention to the relation between uncertainty and ignorance.

Uncertainty is everywhere. It is part of scientific work, decision-making, and everyday life. As the opening quote of the introduction highlights, there is a general tendency to focus on certainty, rather than uncertainty. This makes uncertainty invisible to an extent (Star 1985; Shackley and Wynne 1996; Mesman 2008). The term also tends to have a negative connotation. Melse argues that it is an un-word, indicating that something is absent or missing (Melse 2003; van Asselt 2005).

In searching for a definition of uncertainty, we follow the work of Moreira who defined it as "the non-determinate or unsettled quality of a statement or knowledge claim" (Moreira 2011, 1335). Moreira's definition is highly suitable for us as his study investigated uncertainty in healthcare rationing. The reference to "unsettled" addresses the collective character of uncertainty. Uncertainty gets meaning in collaboration and discussion within a collective. However, "unsettled" also implies that work is needed to reveal uncertainties (or keep them hidden). Hence, "quality" in Moreira's definition underlines that knowledge valuation is not just the application of comparative techniques, but involves collective work. This combination, at the heart of valuation work, is what we want to study in connection with uncertainty in guidelines.

As uncertainty is often invisible and valued negatively, people are likely to avoid it, work around it or to try to overcome it. However, several authors who study ways of dealing with uncertainty point out

that we should try to accept uncertainty. Jerak-Zuiderent studied patient safety and argues that healthcare practitioners must engage with uncertainty to deal with all kinds of demands. She refers to this as "living with uncertainty" (Jerak-Zuiderent 2012). Living with uncertainty has to do with the acceptance of a given degree of uncertainty in medical work, but also points to a healthcare practitioner's mind-set, to always be aware of the uncertain aspects in their work. The challenge is how to do this, and keep doing it when collective decisions must be made. For example, studying the work of futurists, van Asselt et al. refer to "certainification"—uncertainties initially acknowledged in the decision-making process eventually vanished from the definitive documents (van Asselt, Mesman, and van 't Klooster 2007). Whereas decision-makers may recognise uncertainties, these do not have to be included in the final decisions, and thus disappear into the background.

We follow Jerak-Zuiderent in considering that ignoring or banning uncertainty is not productive. To a great extent, however, it is still unknown how we can include uncertainty in EBGs so that coherent and clear recommendations that support healthcare decision-making are made. Certainification is not the solution, but the question we explore is how guideline developers balance between uncertainty and credibility. In the following section we will discuss credibility in relation to uncertainty.

Credibility Needs Uncertainty

Credibility is a key issue in developing EBG. Expressing uncertainty seems to have a great impact on becoming or remaining credible. Wynne's eminent work on Cumbrian sheep farmers shows how distrust can grow when uncertainties are ignored. Wynne's study deals with environmental hazards for farmers after the Chernobyl disaster. Accustomed to all the uncertainties of farming, those farmers had a flexible and adaptable way of life. The environmental experts advised them on how to deal with the possible hazards with a putative high degree of certainty, ignoring the farmers' knowledge, whereby the farmers' trust in the experts' expertise declined (Wynne 2000). In contrast, Gross discusses a redevelopment project for a former coal mining area in Germany. In this case uncertainties were seen as a normal part of the decision-making process and so it was easy for the experts to acknowledge them without losing credibility. This generated lots of space for finding the right solution for the issues involved (Gross 2010). By studying the interaction between scientists and policy makers in debates about the future of climate change, Shackley and Wynne (1996) argue that boundary work helps to establish the authority of science, despite expressing uncertain knowledge, and that it helps to create a common ground for discussing uncertainties in the science-policy domain (Shackley and Wynne 1996).

When creating trustworthy EBGs, it seems essential to accept a degree of uncertainty. Knaapen speaks of evidence-searched guidelines, as she shows how the essence of guideline development is to deal with absences of evidence (Knaapen 2013). In accepting uncertainty, the credibility of a guideline is ensured. In another study, Knaapen et al. observed a guideline development programme and concluded that strong evidence and deployed methods do not ensure the credibility of a guideline. Instead they argue: "[The guideline's] legitimacy rests on the articulation of heterogeneous types of expert knowledge and judgements, both within the guideline development group, and vis-a-vis an external world of textual documents." (Knaapen et al. 2010, 691). As we now go on to show, valuation work, or the work of giving meaning to several types of knowledge, is essential in guideline development.

Uncertainty in Evidence-Based Guidelines

Uncertainty manifests itself in three ways in an EBG. First, there is uncertainty that is inherent in knowledge. Generally, scientific articles and reports are concerned with presenting the facts and omitting all the struggles, insecurities and adaptations that were necessary to create these facts (Latour and Woolgar 1986; Star 1985; Shackley and Wynne 1996). New knowledge brings new insights, but it also brings new areas of ignorance and uncertainty to the forefront (Gross 2010; Jasanoff 2007). Guideline developers must find ways to deal with these (hidden) uncertainties and gaps in knowledge.

A second way in which EBGs are confronted with uncertainty is that they make use of heterogeneous knowledge, such as (cost) effectiveness studies, clinical trials, clinical expertise, patient experiences, often completed with ethical considerations and more. The various types of knowledge have different strengths and weaknesses. All these "knowledges" should be combined, assessed and weighed to be explicitly included or left out of the guidelines (Moreira 2005; Knaapen et al. 2010). As stated above, decision-making processes are full of uncertainties (van Asselt 2005; Wynne 1996; Jasanoff 2007). Many uncertainties must be resolved as guideline development constantly involves decisions on which practical problem to attend to, how to address the problem, which knowledge to leave in or out, and which experts to consult.

The final way in which uncertainty manifest itself in guideline development is in the translation of evidence into recommendations. Knowledge does not arrange a specific action by itself. Instead, knowledge must be actively translated to be of practical use. This work is done in guideline development, and has consequences for uncertainty.

Guideline Development as Valuation

Guideline development is a process of valuation. According to Kjellberg and Mallard (2013) valuation is a process of ordering. Guideline developers bring order into different knowledge sources and types of information. Guideline development is collective work. It is the work of classifying knowledge and giving value to this knowledge. This signification of knowledge is what happens in guideline collectives (Knaapen 2013). A multidisciplinary group of actors is involved in establishing the content of the guideline, supported by methodologists experienced in selecting evidence and writing guideline texts. The whole process of selecting a guideline development group, determining the focus, selecting and weighing the evidence, and deciding how to formulate recommendations has crucial consequences for the outcome of the guideline. Developing guidelines can take years.

Moreira observed these negotiations in a guideline development group and, based on Boltanski and Thevenot's work on justifications, distinguished four repertoires of evaluation in guideline development decision-making (Moreira 2005). These are science, practice, politics, and process. Science involves choices based upon the technical robustness of evidence, practice is about the usability of a recommendation for health care delivery, politics deals with the acceptability of recommendations for stakeholders, and process is about the way in which discussions in the guideline group are adequately represented (Moreira 2005). Moreira's work shows that these considerations engage with each other in the development of guidelines. Although it is not the aim of this article, it is likely that uncertainties play a role in such valuation work, and influence the choice of a repertoire. Knaapen argues that the core struggle of guideline development groups is to find ways to deal with the absence of knowledge. A central question that needs answering is what counts as evidence and what does not (Knaapen 2013). This discussion is the core of valuation work that emphasises signification (Kjellberg and Mallard 2013).

One way to do valuation work is to follow specific procedures for weighing and selecting knowledge. Such methods are important to give meaning to uncertainties (Knaapen et al. 2010). This article analyses some of these methods and explore how they deal with uncertainty. Specifically, we focus on the kinds of valuation work guideline developers engage in to create credible guidelines.

Research Methods

For this article, we interviewed fourteen medical guideline developers from eleven Dutch national organisations. Interviewing guideline developers gave us the opportunity to reflect on their methods and make their experiences central in the analysis. In the Netherlands, various groups and organisations, such as governmental organisations,

associations for specific professionals or disease groups, and research institutes all make EBGs. The wide range of organisations involved in guideline development results in a broad variety of guidelines, both for single professional groups and multidisciplinary groups. There is no specific education for becoming a guideline developer in the Netherlands. Instead, guideline developers have different backgrounds. There are epidemiologists, healthcare practitioners with degrees in education, health scientists, and quality managers. Combinations are possible, such as medical doctor/epidemiologist. All the interviewed guideline developers have at least ten years' experience in developing guidelines. One guideline developer has been in the field for over 20 years. For some in this group, developing guidelines is their core task, whereas others combine it with other part-time work, such as being a practicing physician. We chose this wide selection of respondents as we believed the breadth would bring deeper insights into what happens to uncertainty in the development of EBGs.

The respondents were asked how they developed guidelines, which problems and uncertainties they encountered, and how they dealt with these situations. Colleagues from the institute of Health Policy and Management conducted half of the interviews, in relation to another project on guideline development (Zuiderent-Jerak et al. 2011). All interviews were recorded and transcribed verbatim. The results were analysed both inductively and deductively, in the latter case with a focus on ways of dealing with uncertainty. The empirical section starts with an explanation of guideline development, and then discusses the relation between classification systems and alternative methods for guideline development. We go on to explore the relation between ignorance and guideline development. This empirical section ends with an analysis on how credibility is accomplished in guidelines.

Guideline Development

In this section, we outline the guideline development process, as described by the guideline developers we spoke with. According to our respondents, their procedures are very similar to what is known from the literature (Knaapen et al. 2010; Moreira 2005), although there are differences between different Dutch guideline organisations.

Guideline development starts when there is a reason to develop a guideline for a certain problem. Reasons vary. At the start of the evidence-based medicine movement in the Netherlands, resolving uncertainty in medical practice was the reason to develop a guideline. A guideline developer involved in the field for some twenty years, provides an example:

> The guideline on oral contraception, the pill, was about abolishing check-ups for the pill. In those days, we still had pill check-ups and all women on the pill had to see the doctor twice a year for a smear test. The pill was first perceived as a

> risky thing, which needed to be examined regularly. Over time people started doubting the effectiveness of these check-ups, but how do you organise a stop to this? (Guideline developer/general practitioner involved in guidelines for general practitioners)

These first guidelines were developed to solve uncertainties in medical practice and/or reduce ignorance, according to our respondents. Over time, when the most striking problems had been addressed, the reasons for developing guidelines changed. Gradually guidelines became repositories of how medical work should be done. The same guideline developer remarks:

> Then the question for developing a guideline changed into 'What do guidelines lack? What common problem should we tackle next?' So that raises the question of what we want to achieve with these guidelines. Do we want to describe the entire medical terrain? Then it becomes a sort of handbook. Or do we focus on situations where something is going on, where doctors don't know what to do? (Guideline developer/general practitioner involved in guidelines for general practitioners)

Notably, most guideline developers criticise the idea of making guidelines for situations *without* uncertainties. This does not always mean that no guidelines are made. Interestingly, though, "good" guidelines, according to guideline developers, seem to include some degree of uncertainty; otherwise, the need for a guideline is questioned.

Reasons for developing guidelines change over time, according to our respondents. Sometimes, any new situation determines the need for a guideline. In infectious diseases, every new possible outbreak of a disease is a reason to develop a guideline. A consistent approach towards infectious diseases is essential to tackle the situation and guidelines are the way to reach the healthcare workers involved. Other guideline developers noted that the need for a guideline is determined on the basis of explicit criteria, including the prevalence of the problem, potentially achievable health benefits, solving controversies in practice, satisfying demands from professionals or patient groups and the availability of (at least some) evidence for the problem. These criteria help guideline developers to select relevant topics or to justify to others that such a topic is suitable for a guideline. In contrast, justifying that a topic is not suitable also occurs:

> We must be able to say this is not a subject for a guideline. For example, the geriatric society consulted us for a guideline on medical care for frail elderly on psychiatric wards. This could be a guideline topic. But when we investigated the source of the problem, we discovered that those geriatric beds in many psychiatric hospitals were under pressure due to financial problems. This affected the position of the geriatric doctors. How the medical care was to be given was not the question. Then you should rethink if this is a guideline topic. (Guideline developer/epidemiologist involved in clinical guidelines)

After selecting the topic, guideline developers establish the starting questions of the guideline. These are generally based upon the struggles, uncertainties, or bottlenecks in healthcare practice that are identified by consulting actors in the healthcare field. Who is consulted differs. Most often healthcare workers directly involved in the issue are asked, but for more complex or controversial issues, some guideline development organisations ask a broader range of stakeholders:

> In the guideline we made for intensive care we not only included practitioners, but also health insurers, academic hospitals, the local hospitals, the health inspectorate, health care spokespersons for political parties. We consulted everyone prior to developing the guideline, and asked what we should include, so that we knew what subjects to address and why. (Guideline developer/ epidemiologist involved in clinical guidelines)

Such an approach aims to ensure that most of the relevant issues are known up front, so that further development does not meet too many surprises.

After defining the starting questions, the core of the work of guideline development starts. This includes systematic searching, assessing, and selecting relevant knowledge, and translating various "knowledges" into guideline recommendations. Knowledge comes from scientific publications, reports and documents, international guidelines on the topic, experiences, and expertise, and also often from systematic reviews made, for example, by the Cochrane collaboration or the National Health Institute. The latter type helps translate large amounts of literature and makes it easier to apply in decision-making (Chalmers 1993). However, reviews still need valuation processes to be applicable in guideline development:

> Most of the Dutch guidelines are developed from scratch. We call it "de novo." Of course, we make use of international guidelines and reviews by, for example, the IHI or National Health Institute. They make good evidence reviews, which are also published in the literature. But this knowledge is not always applicable for the guideline we intend to make. So this kind of knowledge has limited use. (Guideline developer involved in GRADE working group)

Any kind of knowledge needs to be assessed for a guideline. This is done in guideline development groups and by guideline methodologists. Guideline development groups, consisting of various representatives with specific expertise and involvement in the issue, discuss the selected knowledge, judge its relevance, check its robustness, and deal with and (at times) resolve any omissions in the knowledge. This valuation work can take months or even years. The guideline drafts are the main focus of the debate. When the guideline is eventually finalised, it is introduced in healthcare practice. Often guideline development organisations have an infrastructure for

implementation, such as websites, periodical publication of a book containing all guidelines, and a network of healthcare practitioners.

Classification Systems: A Curse or a Blessing for Accepting Uncertainty?

The core of the work of guideline developers is classification or ordering of knowledge, often done with classification systems or levels of evidence tables (Gugiu and Ristei Gugiu 2010; Knaapen et al. 2010). These frequently used methods are often criticised by guideline developers. Evidence tables have different levels but their hierarchy is predominantly based on study designs, with level 1 on top and level 4 or 5 on bottom. In such tables, meta-analysis of randomised clinical trials (RCTs) are on top and patients' and practitioners' experiences are considered the least form of evidence. Classification systems help demarcate between "stronger" evidence and more "anecdotal" evidence, as they enable guideline developers to indicate with how much certainty a claim is made. The strength of evidence is made transparent. The levels are a means to accept uncertainty, as they allow demarcating between more and less certain claims. However, levels of evidence tables only help deal with the uncertainty inherent in knowledge (i.e. the first kind of uncertainty discussed earlier). Uncertainties in knowledge valuation and uncertainties in knowledge translation are not resolved with levels of evidence tables. The following two examples clarify our point.

First, classification systems are based upon study design. Strong study designs such as meta-analysis or RCTs tell something about the robustness of the evidence supporting a claim. However, they do not say anything about the *quality* of knowledge for making recommendations in a particular guideline. One guideline developer expressed this as follows:

> If you want to compare two pills, then you use a RCT, if you want to know how to best organise care for a specific group of patients then you might use a qualitative research design. Depending on the purpose of the guideline different knowledge is seen as hard evidence. If you use the same classification schemes for both kinds of research, then the qualitative research is valued less and you might make recommendations that are less firm. Well, as guideline developers we need to pay more attention to these things. (Guideline developer/epidemiologist involved in clinical guidelines)

What knowledge should be rated higher or lower in the hierarchy depends on which question the guideline aims to answer. Levels of evidence tables do not allow for such specificity. The valuation of the quality of knowledge remains the work of the guideline development group.

Secondly, classification systems cannot deal with omissions in knowledge. They can only categorise available knowledge; unknowns

cannot be valued in the levels. One guideline developer referred to this problem in terms of "thoughtless empiricism":

> On the one hand, if there is no trial, then you can say there is no evidence, nothing has been proven. If you are really strict, this means that you can no longer treat numerous patient groups. For example, if you look at osteoporosis, you see that almost all the trials have been done amongst women. So, what to do with men? Well, you could argue that it would work somewhat similar with men, and you could just give them the same pills. You could also argue that nothing has been proven for men, so you stop [prescribing]… But on the other hand, we sometimes face this [situation] if you are too restrictive. For example with heart attacks, certain medications are recommended, especially for the first six months. There are about six pills on the market but only two have been studied in decent trials. Should we then say use only those two, and not the others? What complicates the matter is that this treatment is prescribed for both diabetes and heart failure, and maybe other pills are being studied. So, you see, it's always a diffuse thing. It's what I call thoughtless empiricism; it depends strongly on what study has been done. We definitely need to find compromises. (Guideline developer/general practitioner involved in guidelines for general practitioners)

This quote shows, that omissions in knowledge need to be dealt with. Ignoring these unknowns leads to all sorts of partial recommendations, while the question is how to include the omissions. Classification systems focus only on available knowledge, so as the above guideline developer remarked, compromises are needed to solve these situations.

To sum up, classification systems are an aspect of valuation as they assist guideline developers to classify knowledge based upon study design and source of knowledge. While they rate knowledge, they do not tell anything about its quality. One guideline developer noted:

> Levels of evidence are like the star rating of a restaurant, but you only find out what a restaurant is like when you go and eat there. So the justification is more important than the rating. (Guideline developer at Dutch College of General Practitioners)

In short, classification systems can help to categorise knowledge in more or less proven claims. But that is all they do. They can neither deal with unknowns nor take the relevance of the knowledge to a particular context into account. Classification systems need other valuation practices, such as consensus-making amongst experts, to interpret the meaning of the classification. Classification systems therefore provide only modest assistance in dealing with uncertainty.

Grading Types of Knowledge

An alternative method that many of the interviewed guideline developers mentioned is GRADE, the Grading of Recommendations Assessment, Development, and Evaluation. Responding to some of the criticism of classification systems, the international GRADE working

group has come up with a systematic approach to rate heterogeneous types of knowledge, which is based on more criteria than study design alone. In terms of Moreira's repertoires, GRADE offers a legitimate way to include more of the repertoires of practice, politics, and process, instead of only science (Moreira 2005). A guideline developer with experience in using GRADE explains:

> The advantage is that you can select on subjects that are clinically relevant. You look at results and not the study design. (Guideline developer/policy adviser involved in guidelines for elderly care physicians)

The GRADE method involves five factors that downgrade and three factors that upgrade the quality of evidence (Guyatt et al. 2011). Such factors as "inconsistency" and "indirectness" lower the quality and "large effects" increases the quality. By including more relevant factors in the decision-making process, GRADE tries to suit the valuation processes better. One guideline developer involved in the international GRADE working group explains:

> GRADE is a real step forward, but one of the consequences is that the strength of recommendations generally decreases. There are more considerations to take into account, and they generally turn out to give a lower recommendation. (Guideline developer involved in GRADE working group)

GRADE tries to give more space for valuating "other" (i.e. not considered hard evidence) knowledge, and for expressing uncertainties. By taking more aspects into account, GRADE offers more opportunities to deal with unknowns and uncertainties in guideline development. Guideline developers involved in guidelines for the frail elderly explained that especially in the case of ignorance and uncertainty, this method had advantages:

> Well, the point is that relatively little research is done on the frail elderly. Often there are no RCTs available. So you search for alternatives to find evidence that is clinically relevant for this group. (Guideline developer/policy adviser involved in guidelines for elderly care physicians)

Generally, if studies match only partially with the focus of the guideline, the strength of the recommendations decreases with GRADE. However, in some cases, when a lot of risk is involved, the strength of recommendations can increase. For example:

> One of the best is the WHO guideline on avian flu. It's good as it specifies the considerations and choices. But, if you look at the proof for the advice you can see many unknowns. One factor that influenced their decisions was the considerable risk of disaster, with high mortality and morbidity. This risk and probable low side-effects made the recommendations strong, although there was only indirect evidence. (Guideline developer in GRADE working group)

By including other and more criteria for weighing knowledge than just study design, GRADE brings a broader ground for valuating knowledge. GRADE seems to support decision-making involving uncertainty in valuation and uncertainty in the translation of knowledge into recommendations, while allowing for the uncertainties inherent in knowledge to be addressed. However, at the time of the interviews, most guideline developers had no or only limited experience in using GRADE. Some guideline developers expected GRADE to make their work more complex, as the more formal valuation procedures would make decision-making more technical and time consuming. We have yet to see what these reservations mean to the use of GRADE and its credibility in healthcare practice.

Involving Expertise from the Healthcare Field

> One issue in guideline development is that you can't solve every question with evidence. If we are too strict, there will be hardly anything left in the guideline, especially since we focus on nursing care for the elderly. (Guideline developer involved in guidelines for nurses and geriatric assistants)

Guidelines cannot be made without experiential knowledge; i.e. the knowledge of healthcare practitioners and patients in the healthcare field. However, as we have discussed above, this most anecdotal kind of knowledge forms the bottom level of the evidence system. It risks being seen as individualised information, which is difficult to make relevant to the guideline. This section explores how such knowledge is used and what happens with uncertainty.

Including the expertise of healthcare professionals and patients is assumed to have several benefits, as it brings different information about healthcare delivery to the fore. For example, one of the epidemiologists developing clinical guidelines remarks:

> Surgeons and orthopaedists have different policies on anti-coagulants for some conditions. They argue that the guidelines don't need to mention this, as they agree to disagree on this point. Yet, a focus group revealed that patients in a shared room find it troublesome to be getting different treatment for the same complication. (Guideline developer/epidemiologist involved in clinical guidelines)

Such experiences are important to include in a guideline. Patients' and healthcare practitioners' knowledge not only fills in important unknowns, it also explores whether guideline recommendations are feasible and accepted.

But how should this knowledge be included in guidelines? Guideline developers have little experience with methods for including experiential knowledge. Some guideline developers have used Delphi-like methods, but regard them as time consuming and expensive. One

guideline developer refers to experiential knowledge as "impressionistic":

> It's like you say something, I say something and we put it together, but it's not systematic. (Guideline developer at Dutch Institute for Healthcare Improvement)

Interestingly, while guideline developers are highly systematic when it comes to knowledge assessment in general, they tend to be less systematic when it involves including more experiential knowledge (Zuiderent-Jerak, Forland, and Macbeth 2012). So how do guideline developers ensure that experiential knowledge is not too anecdotal? Generally, they rely on a large number of (patient) representatives:

> If there is a good patient-representing association we will contact it. They have investigated their members' demands and know what they want. Otherwise we often use focus groups of patients. If, for example, I make a guideline for emergency surgery, well there isn't a patient association for that, so then we'd consult a focus group. But we should evaluate if this is the best approach although I don't know how we could do it differently. (Guideline developer/epidemiologist involved in clinical guidelines)

The interviews revealed numerous cases of a request for a guideline, despite the absence of knowledge. As discussed above, uncertainties are often the reason to start developing a guideline. One example comes from the guidelines developed in youth health care:

> Very often there is no literature on our subjects, since we work in preventive care. It's on a different level. For example, we deal with screening programmes, how to screen for children that fall behind or don't function well. Well, you don't find this directly in the literature. […] So a huge part of our guidelines is practice- or expert-based. That's justified by grey literature, handbooks, expert opinions, focus groups etcetera. (Guideline developer/physician involved in guidelines for youth health care)

Another telling example is guidelines for new infectious diseases, made by a governmental organisation for infection prevention. With an outbreak of a new infectious disease (or the threat of one), such as the swine flu pandemic or SARS, there is a lot of uncertainty due to both ignorance and public reactions. A developer of the swine flu guideline explains:

> In the beginning we knew nothing. Something started in Mexico, but if and how it would affect us in the Netherlands was unknown. Our boss explained that it was severe in Mexico. The Spanish flu used to be severe as well, and that was our only frame of reference. (Guideline developer involved in guidelines for infectious diseases)

In the absence of knowledge and in the presence of the risk of an outbreak, guideline development becomes a delicate situation. The

public is highly involved in this situation, and may reaction with fear, indifference, and criticism:

> We got a lot of flak, as if we were taking it [i.e. reaction to a possible swine flu pandemic] out of proportion out of our own interests, since people suspected us of having stakes in the vaccine industry. Based on this criticism, you'd think that people would refuse the vaccine because, they argued, we made a problem out of nothing. But, people did take the vaccine, despite the fact that they also thought we made a big fuss about it. (Guideline developer involved in guidelines for infectious diseases)

The study by Gross point out that in situations of ignorance, communication of uncertainty is accepted (Gross 2010), but here the expertise of the governmental organisation was questioned and criticised. The guideline developer explained that they felt that acknowledging uncertainty was not an option, as there was a lot of pressure on them to come up with "an answer." She reflects:

> We concluded that maybe we should say explicitly that we don't know either. But people assume they'll get an answer from us. So we're almost forced to say something. And if we don't know either, then what should we do? Then we say "take all possible measures." It is actually impossible if you think about it. (Guideline developer involved in guidelines for infectious diseases)

In the absence of knowledge on the infectious agent or possible remedies, the governmental organisation for infectious diseases follows another approach to develop their guidelines. Especially with novel infectious diseases there is often a lack of knowledge on the disease as it is too new. Therefore guideline developers include the literature on viruses that look similar and—until more knowledge becomes available—they adapt the interventions suggested to deal with similar viruses. Of course, for "older" infectious diseases, such as hepatitis, rabies, or measles, specific literature is more widely available. Besides this literature search, experts and healthcare professionals in the Community Health Services[1] are intensively involved in guideline development. An external expert (a medical specialist, biologist, or virologist) is consulted to write the text and the texts are subsequently discussed with fifty representatives, one from each Community Health Service. The group reflects on all the comments and the result is the definitive guideline recommendations.

This organisation of guideline development ensures that experiential knowledge becomes known and can be included at a relatively early stage. After the guideline is finished and published, the governmental organisation encourages feedback. Guideline users can report all their new knowledge and experiences of using the guideline

[1] Dutch Public health policy is executed by regional Community Health Services serving a varying number of municipalities.

on a special 24/7 telephone service. This feedback not only enables the guideline developers to adjust their advice, but at the same time informs them about new knowledge and the practical usefulness of their recommendations. If they conclude, from this information, that the guideline should be changed, then this is done immediately. Acknowledging uncertainty thereby becomes an open and collective effort between guideline developers and practitioners. It is achieved by creating feedback moments, when the comments and experiences of users can be inserted in the guideline, even after its publication. The highly interactive process shows how uncertainty is fully integrated into the process of making guidelines. This approach not only improves development and the fine-tuning after publication, it also deals with uncertainties involving the implementation and use of guidelines. Feedback brings important insights into how the guideline is used and interpreted.

To sum up, guideline developers are very aware that they need experiential knowledge from healthcare practitioners, patients and specialised experts to create guidelines. There are, however, still great challenges in including this knowledge in ways that move beyond the overly "impressionistic." The feedback system used in guidelines for infectious diseases is a promising example of how uncertainties can be addressed collectively.

Ensuring Credibility of Guidelines

A core concern of developers is how their guidelines are received and used in healthcare practice. How can guidelines remain credible and express uncertainty at the same time? The "evidence-based" label gives the impression that evidence makes guidelines credible. However, as Knaapen argues, evidence-based medicine is more often about how to deal with the absence of evidence (Knaapen 2013). When asking guideline developers what "evidence-based" means, they answered that it deals congruently with working *systematically* and *transparently*:

> For me, a guideline is evidence-based when we have followed the process. So, when you define the focus and the limits at the start, and then you search the literature systematically, in all the databases. Evidence-based is when you select and assess the literature systematically, so that you come to a systematic conclusion. (Guideline developer/epidemiologist involved in clinical guidelines)

And another guideline developer explains:

> An evidence-based guideline is one where you can see if each recommendation is based on consensus or the literature. You can see that the literature has been searched in depth, so you can repeat a search. And you can see the justification for the recommendation, like "Jansen says this, Pietersen says that, and we chose this because..." (Guideline developer involved in guidelines for infectious diseases)

Evidence-based does not refer to the strength of the evidence found, but to the process of making guidelines (see also Knaapen 2013). In terms of uncertainty, the procedures for making EBGs involve systematic searches to ensure that there is indeed evidence, and if not found, that there is "truly" no evidence (Knaapen 2013). In other words, doing things systematically and transparently ensures the "evidence-basedness." As earlier work on guideline development has shown, guideline credibility is determined by the inclusion of a diversity of knowledge sources and comparisons to similar reports and documents. At times, therefore, we found that strong evidence is presented with softening nuances, otherwise it would reduce the credibility of the guideline. For example:

> I was involved in a guideline on sedation policy. There was very strong evidence that it's good to have an extra professional monitoring a patient during sedation. But we don't have these professionals and it involves training. It's unclear who should pay and how many of these professionals are needed. So it's worthwhile knowing this all, but to keep the actual recommendations a bit loose. Otherwise it leads to all kinds of problems in acceptance of the guideline. This then affects the trust in the whole guideline, not just this recommendation alone. (Guideline developer/epidemiologist involved in clinical guidelines)

Despite the strong evidence, the guideline developers chose to soften the recommendation a bit, since recommending unfeasible things can affect the acceptance of the whole guideline. In contrast to Wynne's sheep farmers, the situation here shows that uncertainties in practice are not ignored, but form a part of the rationale for deciding which evidence to include and how to include and present it.

One way to ensure credibility is to use a systematic evidence-based working method:

> We often get attacked for the recommendations we make. As a governmental organisation, we're under attack anyhow. That's why we need to make evidence-based guidelines. If we can't make well-founded statements, based on good knowledge, we're in trouble. We are very conscious of that. (Guideline developer involved in guidelines for infectious diseases)

A systematic evidence-based working method legitimises the credibility of the governmental organisation in making their guidelines. False certainty or certainification does not take place, according to our respondents.

Trust and credibility affect decision-making in the development of guidelines and are one of the many considerations that must be taken into account. Guideline development seems a practice that inherently addresses uncertainty, and therefore does not run into credibility issues, as Wynne describes. Instead, as we have tried to show in the empirical sections, guideline making is reflexive work that seek optimal ways to reflect what is known and what is uncertain, and to

do this in such way that it can retain credibility and guide healthcare practice.

Conclusion

This article explored the valuation work that guideline developers undertake to develop EBGs and how uncertainty is addressed in the process. We distinguished three valuation practices, based on empirical findings: classification of studies, grading types of knowledge, and those involving expertise and clinical practice. These three valuation practices differed in the types and amount of uncertainty they could endorse. Classification studies seem helpful for guideline developers in dealing with uncertainties inherent in knowledge, but cannot deal with ignorance and do not help to relate knowledge to a particular context. Thus, guideline developers need other valuation practices to interpret and include knowledge than solely classification systems. Grading different types of knowledge is, in the guideline developers' view, slightly better equipped to assist in valuation practices and to live with uncertainties. GRADE seems to better allow one to include various kinds of uncertainty and provides a ground for legitimising the choices made in the guideline development process. Involving expertise and practice endorses all three types of uncertainty, but risks being too anecdotal.

The type of valuation practice has consequences for the outcome; some types are better capable of accepting uncertainty than others. What seems essential is that the valuation practices that work better seem better capable of including various kinds of uncertainty and provide the grounds to legitimately justify the choices made in the decision-making process. This combination—allowing for uncertainty and yet being able to justify choices made through some form of systematic way of working—enabled guideline developers to deal with uncertainty.

The reflexive aspects of valuation work are particularly interesting. Valuation work in guideline development not only involves input (assessment of knowledge) but also the output (how users perceive the result). A telling example is the case of guideline making for infectious diseases. Feedback from users helped the developers improve the guideline and gain insight into how the guideline was used. A feedback system is likely to prevent some of the uncertainties that tend to occur in guideline implementation, such as uncertainties in the uptake of recommendations and the spread of the guideline.

The question how to remain credible can be solved by including heterogeneous types of knowledge (Knaapen et al. 2010). Wynne's study showed that ignoring fundamental aspects of knowledge (sheep farmers' local knowledge) leads to distrust and unrest (Wynne 2000). The challenge for guideline developers is thus to include relevant knowledge from various sources and of different strengths, and doing

this systematically and transparently. Justifying choices is essential and guideline development methods seem to offer a formal way to do this justification.

Our data showed that guideline development seems to be most systematic with knowledge that is more certain, and least systematic when knowledge is less certain. That is, knowledge stemming from patient experiences and expertise of professionals is generally not collected and included following a systematic approach, but as one guideline developer argued, it is "impressionistic." The knowledge that is most uncertain, in relation to its external validity, is included least systematically. How to approach this situation is one of the challenges for the future of guideline development.

For this article we interviewed guideline developers. We selected guideline developers working for different organisations and with different personal backgrounds. The benefit of this choice is that we could explore a broad range of valuation practices, and also see which elements of the evidence-based approach were common in all the different places. Dutch guideline development is likely to be done differently than in other countries, and this should be taken into consideration interpreting these results. We relied on the interviews as our main research method. Observation of guideline-making practices might produce different findings.

In studying valuation practices in guideline development we found that uncertainty is in many ways inherent and is essential to create EBGs. We conclude that guideline developers use different valuation practices to deal with this inherent tension in their work and these practices have different consequences for the types of uncertainties that can be taken on board. Studying guideline development as valuation work enabled us to move beyond a more rational investigation of classification of knowledge. Instead valuation serves as a valuable notion to study how heterogeneous and divergent knowledge can be connected, and how and where uncertainties are acknowledged.

Acknowledgements. We would like to thank the respondents for the open conversations and the valuable insights into their work. Thanks to Teun Zuiderent-Jerak, Sonja Jerak-Zuiderent and Hester van de Bovenkamp for sharing some of their data on guideline development with us. Finally, we would like to thank Marjolein van Asselt, Charles Bosk, Alena Bleichner and colleagues of the Health Care Governance group, especially Teun Zuiderent-Jerak, for comments on earlier version of this article.

References

Callon, Michel, Pierre Lascoumes, and Yannick Barthe. 2009. *Acting in an Uncertain World*, translated by Graham Burchell. Cambridge, Mass.: The MIT Press.

Chalmers, Iain. 1993. "The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care." *Annals of the New York Academy of Sciences* 703 (1): 156–165.

Graham, Robin, Michelle Mancher, Dianne Miller Wolman, Sheldon Greenfield, and Earl Steinberg. 2011. *Clinical Practice Guidelines We Can Trust*. Institute of Medicine. Washington, D.C.: The National Academies Press.

Gross, Matthias. 2010. *Ignorance and Surprise*. Cambridge, Mass.: The MIT Press.

Gugiu, Cristian P., and Mihaiela Ristei Gugiu. 2010. "A Critical Appraisal of Standard Guidelines for Grading Levels of Evidence." *Evaluation & the Health Professions* 33 (3): 233–255.

Guyatt, Gordon, Andrew D. Oxman, Elie A. Akl, Regina Kunz, Gunn Vist, Jan Brozek, Susan Norris, Yngve Falck-Ytter, Paul Glasziou, Hans deBeer, Roman Jaeschke, David Rind, Joerg Meerpohl, Philipp Dahm, and Holger J. Schünemann. 2011. "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables." *Journal of Clinical Epidemiology* 64 (4): 383–394.

Helgesson, Claes-Fredrik, and Fabian Muniesa. 2013. "For What It's Worth: An Introduction to Valuation Studies." *Valuation Studies* 1 (1): 1–10.

Jasanoff, Sheila. 2007. "Technologies of Humility." *Nature* 450 (1): 33–33.

Jerak-Zuiderent, Sonja. 2012. "Certain Uncertainties: Modes of Patient Safety in Healthcare." *Social Studies of Science* 42 (5): 732–752.

Kjellberg, Hans, and Alexandre Mallard. 2013. "Valuation Studies? Our Collective Two Cents." *Valuation Studies* 1 (1): 11–30.

Knaapen, Loes. 2013. "Being Evidence-Based in the Absence of Evidence: The Management of Non-Evidence in Guideline Development." *Social Studies of Science* 43 (5): 681–706.

Knaapen, Loes, Hervé Cazeneuve, Alberto Cambrosio, Patrick Castel, and Beatrice Fervers. 2010. "Pragmatic Evidence and Textual Arrangements: A Case Study of French Clinical Cancer Guidelines." *Social Science & Medicine* 71 (4): 685–692.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

Melse, Johan. 2003. "Van Onzekerheid Weg Denken." In *Niet Bang Voor Onzekerheid*, edited by Marjolein van Asselt and Arthur Petersen, 61–72. Utrecht: Lemma.

Mesman, Jessica. 2008. *Uncertainty in Medical Innovation: Experienced Pioneers in Neonatal Care*. Basingstoke: Palgrave Macmillan.

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

———. 2011. "Health Care Rationing in an Age of Uncertainty: A Conceptual Model." *Social Science & Medicine* 72 (8): 1333–1341.

Moreira, Tiago, Carl May, and John Bond. 2009. "Regulatory Objectivity in Action: Mild Cognitive Impairment and the Collective Production of Uncertainty." *Social Studies of Science* 39 (5): 665–690.

Shackley, Simon, and Brian Wynne. 1996. "Representing Uncertainty in Global Climate Change Science and Policy: Boundary-Ordering Devices and Authority." *Science, Technology & Human Values* 21 (3): 275–302.

Star, Susan Leigh. 1985. "Scientific Work and Uncertainty." *Social Studies of Science* 15 (3): 391–427.

Timmermans, Stefan, and Alison Angell. 2001. "Evidence-Based Medicine, Clinical Uncertainty, and Learning to Doctor." *Journal of Health and Social Behaviour* 42 (4): 342–359.

van Asselt, Marjolein B.A. 2005. "The Complex Significance of Uncertainty in a Risk Era: Logics, Manners and Strategies in Use." *International Journal of Risk Assessment and Management* 5 (2–4): 125–158.

van Asselt, Marjolein B.A., Jessica Mesman, and Susan A. van 't Klooster. 2007. "Dealing with Prognostic Uncertainty." *Futures* 39 (6): 669–684.

van Loon, Esther, and Teun Zuiderent-Jerak. 2011. "Framing Reflexivity in Quality Improvement Devices in the Care for Older People." *Health Care Analysis* 20 (2): 119–138.

van Loon, Esther, Teun Zuiderent-Jerak, and Roland Bal. 2013. "Diagnostic Work Through Evidence-Based Guidelines: Avoiding Gaps Between Development and Implementation of a Guideline for Problem Behaviour in Elderly Care." *Science as Culture* 23 (2): 153–176.

Wynne, Brian. 1996. "May the Sheep Safely Graze? A Reflexive View of the Expert-Lay Knowledge Divide." In *Risk, Environment and Modernity: Towards a New Ecology*, edited by Scott Lash, Bronislaw Szerszynski, and Brian Wynne, 44–83. London: Sage Publications.

———. 2000. "Misunderstood Misunderstanding: Social Identities and Public Uptake of Science." *Public Understanding of Science* 1 (3): 281–304.

Zuiderent-Jerak, Teun, Frode Forland, and Fergus Macbeth. 2012. "Guidelines Should Reflect All Knowledge, Not Just Clinical Trials." *BMJ* Vol. 345, Article Number e6702.

Zuiderent-Jerak, Teun, Sonja Jerak-Zuiderent, Hester van de Bovenkamp, Siok Swan Tan, Leona Hakkaart-van Roijen, Werner Brouwer, and Roland Bal. 2011. *Variatie in Richtlijnen: Wat Is Het Probleem?*. Instituut Beleid en Management Gezondheidszorg. Erasmus Universiteit: Rotterdam.

Esther van Loon is a PhD candidate in the Healthcare Governance group at Erasmus University Rotterdam, the Netherlands. She has a background in nursing and health sciences. Her research focuses on innovation practices in long-term healthcare and the development and implementation of standards and guidelines. She is especially interested in understanding the effects of the interaction between development and use of healthcare innovations and technologies. She has previously published in *Health Care Analysis* and *Science as Culture*.

Roland Bal is Professor of Healthcare Governance at Erasmus University Rotterdam. His research interests include science, policy, and the law and governance infrastructures in health care. With Wiebe Bijker and Ruud Hendriks, he published *Paradox of Scientific Authority* (MIT Press), on the role of science advisory councils in (health) policy. More recently, he has studied the creation of public accountability in healthcare, studying ways in which public services 'organise for transparency'. Recent publications have appeared in, amongst other journals *Social Science & Medicine*; *Science, Technology & Human Values*; and *Public Administration*.