

ORIGINAL ARTICLE

State Measurements of Benefit Fraud: Why Expert Elicitations Cannot Be Used to Measure Incorrect Personal Assistance Payments

Niklas Altermark and Hampus Nilsson

Lund University, SE

Corresponding author: Niklas Altermark (niklas.altermark@svet.lu.se)

During the last decade, personal assistance programs have suffered from cutbacks, resulting in a public debate about the future of the service. This debate has largely been framed by public commission estimations of the level of incorrect personal assistance payments, where the idea that circa 10% of all assistance payments are incorrect repeatedly has been used to justify measures designed to decrease costs. These estimates have been generated with the Expert Elicitation Method (EEM), originally developed to assess uncertainty and risks based on hard scientific data. In this study, we aim to answer two questions: (1) whether the method is suited to estimate incorrect payments and (2) whether the application of the method follows scientifically agreed upon methodological recommendations. By a systematic literature review about the method and when it can be used, we conclude that the EEM is not applicable to estimate incorrect payments. In a second step, we examine whether the applications of the Swedish public commission follow common methodological recommendations in the research literature, concluding that assessments of incorrect personal assistance payments do not meet basic requirements of how the method is supposed to be used. Hence, our overall conclusion is that the application of the EEM to estimate incorrect payments in personal assistance should not be relied upon to guide policy decisions.

Keywords: expert elicitation; expert elicitation method; personal assistance; austerity

In the last decade, the Swedish public debate about disability politics has been dominated by the costs of personal assistance, a cash payment service designed to provide disabled people with independent living. The debate about rising costs accelerated after the general election of 2014, when a new government consisting of the Social Democrats and the Green party, made clear their intention to trim the costs of the service. Between 2014 and the following election in 2018, there was a 10% decrease in the number of people getting the service, which largely is a result of how the responsible agency, the Swedish Social Insurance Agency (SSIA) (Sv: Försäkringskassan) today reject around 90% of all new applications for personal assistance (assistanskoll.se 2018). This, in turn, can be linked to the government steering of the SSIA, focusing on the rising costs of the service and urging the SSIA to contribute to decreasing the total number of assistance hours granted (see Socialdepartementet 2015; Altermark 2017).

A prominent feature of the government rhetoric has been the frequent references to 'over-use' and welfare crimes within personal assistance (see Altermark 2017). In fact, the majority of the articles and press releases signed by the responsible minister between 2014 and 2017 referred to personal assistance as haunted by rising costs and incorrect payments, which is an implicit reference to welfare fraud (e.g. Socialdepartementet 2015; Regnér 2016a; Regnér 2016b Begler & Lender 2016). The recurring number referred to in this context has been that about 10% of all cash payments do not correspond to real personal assistance needs. As is evidenced by the above referenced texts of the SSIA and the responsible minister, this estimation has framed the debate about personal assistance and has frequently been invoked as a justification of measures taken to reduce costs (see Altermark 2017). This estimation is the result of a number of state commissions which focus on all social insurance systems or on personal assistance specifically. These make use of a method called the Expert Elicitation Method (EEM), which was developed to measure insecurity through expert deliberations based on natural scientific data. The method has now been adopted in Norway, whilst still being used in Sweden. In order to have an informed discussion about the basis of current policy decisions and debates on personal assistance, it is necessary to make sense of the EEM and its application by the Swedish state.

In this article, we will evaluate the use of the EEM, specifically focusing on (1) the evidence of its suitability and (2) whether its application meet established methodological standards. We will do this by means of a systematic

review, including almost 1,300 texts either presenting applications of the method or describing its strengths and weaknesses, and an analysis of how the Swedish applications of the method relate to methodological standards. Our main conclusions are that there is no evidence that this method can be used to estimate incorrect welfare payments and that the applications of the method in Sweden disregard a number of central methodological recommendations.

Research design

This study is designed to answer two questions: whether the EEM is suited to estimate incorrect personal assistance payments and whether the Swedish public commissions use the method in accordance with methodological standards. Hence, it is necessary for us to assess the evidence base for using the method to estimate incorrect welfare payments, to pinpoint methodological recommendations in the relevant literature, and to compare these recommendations with how the method has been used to estimate incorrect personal assistance payments. Our research design is thus set up in three steps: (1) a systematic research review, (2) a narrative interpretation of review entries and (3) a text analysis. These are described below.

Systematic review of research using EEM

As noted by Boaz et al. (2002) systematic reviews of research developed as a tool to answer questions about 'what works,' often as concerns the evidence base of health care interventions or in the context of education. The basic idea is that systematic reviews can synthesize evidence concerning a specific topic of interest (Davies 2019). In order to do this, the reviewing strategy needs to be designed with reference to a specific research question (see Davies 2019, 1008; Siddaway et al. 2019), since different questions call for different review strategies. In this study, our concern is whether the EEM works to estimate incorrect welfare payments, and the first step in answering this question is to examine the evidence-base of the method in this area.

To find applications of the EEM, we used the Lund University search engine. It covers 200 databases of academic journals, 300,000 e-books, and 17,000 e-publications, and is considered to be the most extensive Swedish academic search engine, on par with any European equivalent. The search engine covers a broad range of research fields, which means that it is unlikely that entries from specific subfields are systematically excluded. This makes this search engine well-suited to generate a broad and inclusive sample of applications of the EEM.

As Boaz et al. (2002: 5) argues, being explicit with inclusion/exclusion criteria is central for the reliability of a systematic review. Hence, any review should start with the formulation of eligibility criteria operationalized as search terms (see Davies 2019: 1009–1010). This study has used the search terms 'expert elicitation' OR 'expert judgement elicitation' as inclusion/exclusion criteria. These are the established names of the method, which means that they are frequently used in abstracts and among keywords. Texts that contain any of the search term among keywords or in abstracts is included in the study. Following from the research questions of this study, the aim was to collect as many applications of the method as possible, irrespective of research field and topic area. When adding search terms, such as 'expert judgement' and 'expert panel', the number of search items found increased significantly. However, this also meant that a large number of entries were unrelated to the method examined in this article, a common problem of systematic reviews described by Davies (2019: 1010). Rather than adding search terms, in order to reduce the risk of missing important applications of the EEM, we did complementary searches on Google Scholar and Google, without finding any entries that our initial search had missed. After adjusting for double posts and papers not using this method, we ended up with 1,288 original articles, reports and chapters in full text.

There are two kinds of texts in the material: research using the EEM and reporting on the findings and texts discussing the method, how it should be used and what the benefits and downsides are. We were specifically searching for texts that used the method to estimate incorrect payments, welfare politics or similar questions or that discusses how the EEM can be applied. We read through all abstracts in order to map how the method is used. Since the purpose here is to simply describe how the method is used, it was not necessary to weigh studies or operationalize indicators of the robustness of results.

Narrative interpretation of results

There are qualitative as well as quantitative methods for systematic research reviews (see Boaz et al. 2002; Gough & Elbourne 2002). In order to determine whether the EEM is well-suited to estimate incorrect welfare payments and whether the Swedish applications follow methodological recommendations, it is insufficient to simply describe whether the method has been applied for this or similar purposes, it is also necessary to distil central methodological recommendations from the literature. This is akin to what Boaz et al. (2002) describe as a 'narrative interpretation,' where 'narrative synthesis brings together the results of the studies and looks at similarities and differences between the studies and their outcomes' (6). In other words, the research question requires a qualitative analysis to derive central methodological recommendations.

Here, it is important to note that the EEM exists in various forms and can be used for different purposes. There is also a debate about how the method should be set up. Siddaway et al. (2019) stresses that review methods can be an efficient means to make sense of complex and multifaceted research fields, with reference to a specific question. In congruence with this, for the purposes of this study, there is a need to transform the complexity of debates on EEM so that it will be

possible to evaluate the methodology of the Swedish public commissions. For the purpose of evaluating the Swedish applications, we settled on describing methodological recommendation and criteria that more or less all applications have in common, in other words, describing the overlapping consensus as concerns what EEM entails. Davies (2019, 1008) has pointed out that an advantage of systematic reviews is to eliminate the bias of single studies. Following from this, looking for methodological recommendations that are widely accepted is appropriate.

More specifically, we reviewed texts in our sample that either discuss the method of EEM or that have a methods section which discusses how the method should be applied. Texts that merely describe the set-up of the method in the specific study were excluded in this step. Thereafter, we read the text in full (if the purpose of the text was to discuss EEM as method) or relevant parts of the text (if it was an application of the EEM that discusses methodological recommendations). We were specifically looking for methodological recommendations that these texts shared.

Text analysis

The last step of this study consists of making an estimation of whether the Swedish applications of the EEM meet generally agreed upon methodological recommendations. Hence, it was necessary to compare the methodological recommendations that were distilled from the literature on the EEM with the Swedish public commissions in question.

Methodological choices and considerations of the Swedish public commissions are described in the publications where assessments of incorrect payments are presented. As is described in more detail below, there are three Swedish applications of relevance here: the reports *Vad kostar felen?* (Sw: 'What is the cost of the incorrect payments') (FUT 2007) authored by Delegationen mot felaktiga utbetalningar (Sw: the Delegation against incorrect payments) (FUT), the Swedish National Financial Management Authority's report *Samverkanuppdrag mot felaktiga utbetalningar från välfärdssystemen 2010* (Sw: 'Cooperation assignment against incorrect payments from the welfare systems 2010') (ESV 2011:11) and *Åtgärder mot fusk och felaktigheter med assistansersättningen* (Sw: 'Measures against frauds and incorrect payments within personal assistance') (SOU 2012:6). Although we have read these reports in full, the analytical focus has been on the sections devoted to the EEM, where we compared the applications of the method with the standards found in the scientific literature.

There is of course a possibility that these texts do not fully describe the methodology of the Swedish applications. To make sure that we did not miss design features not included in the published reports, we visited the Swedish national archive and went through the archived material related to the final report of the FUT delegation and SOU 2012:6 (there is no archive material about ESV 2011:11). This material did not add to the methodological descriptions in the published texts, although it made it possible to trace participating experts.

The text analysis consists of using the methodological recommendations of the EEM literature as yardstick to evaluate the methods descriptions in the public reports. As is described below, central methodological recommendations concern choice of experts and bias, background data, transparency, and interpretation of results. The analysis in this step consisted of analysing the public reports and evaluating whether they met the recommendations in these areas.

Is the EEM suited to estimate incorrect personal assistance payments?

The Swedish commissions and the EEM as a scientific method

To start with, it is necessary to introduce the public commissions that have used the EEM along with the basic features of the method. The background here is that welfare cheaters and overuse of social insurance was introduced as a topic of public debate in the 1990s and a few years later at the turn of the century (Lundström 2011: 92), primarily by the liberal and conservative parties of the Swedish parliament. When the liberal-conservative block was formalised in the lead up to the 2006 election – an election this block would go on to win – the fight against welfare cheaters was a central part of the political program (Johnson 2010: 222; Lundström 2011: 115–116). Just as in the UK and the US, a generous welfare insurance system was said to lead to 'benefit dependency' and 'overuse' of welfare services (see Fraser & Gordon 1997). Furthermore, stories about welfare crimes of people faking illness or disability to get social insurance payments started to appear in media, a tendency that has continued to this day. As a response to this, the Social democratic government, that would go on to lose the election in 2006, initiated a public commission – called the Commission of Incorrect Payments or the FUT-commission – with the task of investigating overuse and welfare criminality. For reasons not clarified in their publications or in the archived material, the commission settled on using EEM to estimate the costs of incorrect payments. As we shall elaborate on below, the idea is that expert estimations can be used to generate a general measure of the level of incorrect payments. The FUT-commission focused on 16 welfare systems, that were estimated at separate two-day seminars. In the report (FUT 2007), the results showed that personal assistance was one of the systems with the highest rates of overuse and fraud, estimated to be 10.9% of all cash payments. However, the results also contained an uncertainty interval ranging between 6–19%. This is not to be confused with a confidence interval, which is a statistical measure of the robustness of results when data is collected from a sample of a population. Here, the uncertainty interval rather functions as a measure of the uncertainty in the group of experts. However, despite the high degree of uncertainty as regards the result, the single estimate of 10.9% would be the result focused on by politicians and media.

The FUT-commission was followed up on four years later by Ekonomistyrningsverket (ESV), the public agency that the government assigned the role of coordinating with several other state agencies in the work against incorrect payments

and benefit frauds. As part of this work, the ESV presented new EEM-estimations of incorrect payments within a number of social insurance systems, stating that 12.2% of all assistance payments were incorrect (ESV 2011:11). This time, the uncertainty interval was wider, ranging between 1.8–27.6%. Nevertheless, also this time the level of uncertainty disappeared in the reports. The responsible minister, Maria Larsson, issued a new public commission lead by High Court Justice Susanne Billum, focusing specifically on incorrect personal assistance payments (Dir. 2011:26). In a statement, Larsson declares that the ESV has showed that ‘at least 12%’ of all assistance costs are incorrect (see Folcker Aschan 2011). The Billum-commission (SOU 2012:6), as it came to be called, did not proceed to make its own EEM-process, but reinterpreted the ESV-numbers. Also, their re-evaluation suggests that about 12% of all personal assistance payments were incorrect, although, for reasons we shall return to, with a narrowed uncertainty interval.

Before the public commissions that used the EEM, there was virtually no public discussion about disability services and welfare crime. In the last decade, on the other hand, this has been a main feature of discussions revolving around disability politics. A number of decisions affecting personal assistance has referred back to the results of these reports (see Socialdepartementet 2015; Näsman 2016). Whilst media and politicians were focusing on the eye-catching results, however, very little attention was paid to the method of EEM, which at most was described as a scientific measurement. However, as will be discussed at length below, the EEM was not an established method to estimate incorrect payments. The origins of systematic expert elicitations are commonly traced back to the Research and Development branch of the US Air Force (RAND), originally developed to produce scenarios and prognoses in situations of scarce information (Ayyub 2001; Knol et al. 2010; Morgan 2014). After the security classification of the RAND expert elicitations was lifted, the method came to be broadly applied within a number of areas in the 1960s and 70s. Although the method has since evolved, the method is still primarily used to answer questions where standard scientific measurements are not obtainable. Instead, the EEM is based on expert deliberations and judgements provided comprehensive background data and rigorous procedures concerning how summarised estimations are calculated (see de Franca Doria et al. 2009; Rietbergen et al. 2016). For example, because there is no experimental setting that can help us estimate how large evacuation areas should be after nuclear incidents, the EEM can be used to make calculations based on estimations by scientific experts and provided information about factors such as climate, the architecture of the power plant, and the properties of nuclear fuel. Although the results will not meet established scientific standards, it is nevertheless necessary to base policy decisions on some kind of approximation when formulating a security policy. Hence, the EEM is often applied to make estimations related to technology, biosecurity, engineering, and health risks, where it is necessary to provide decision-makers with background information despite uncertainty (see Cooke & Goossens 1999; Ouchi 2004; Landeta Rodriguez 2006; Butler, Thomas & Pintar 2015). In summary, it can be said that EEM is applicable if three conditions are met: (1) the question is impossible to answer by established scientific methods (such as experiments or analysis of empirical data), (2) trustworthy background data and scientific knowledge is available and can inform expert deliberations, and (3) there is a great societal need to answer the question.

More concretely, a typical EEM-process starts with the expert group being put together. The group shall be diverse, independent, and represent different viewpoints. These experts are provided with comprehensive background data in advance and are then asked to make estimations, often in the form of an interval, without knowledge as concerns who else participates (see de Franca Doria et al. 2009; Rietbergen et al. 2016). Thereafter follows a number of deliberative rounds where experts are asked to improve the evaluations. In its standard format, the goal of these rounds is to reach consensus among the experts (Knol et al. 2010), although more recent applications often skip this in favour of open processes where arguments and counterarguments are carefully documented and where experts are allowed to differ in their judgements. Oftentimes, the whole process is anonymous (Ayyub 2001: 3–4). The resulting estimations are used to create an uncertainty interval, which sometimes is used to derive a single estimation of the results. However, it is generally believed that the uncertainty interval is the most important result since it provides information about the level of agreement among experts. Of course, there are variations as concerns how this process is set up, based on methodological differences and the question examined, but the above description involves the most common general features of the EEM.

Literature review results

Our examination of the 1,288 publications in our sample shows that EEM is almost exclusively used to make estimations based on hard scientific data and established scientific knowledge, most often concerning engineering, radiation safety, health risks, and biosecurity. It is primarily used as a method to estimate risk and uncertainty based on indisputable natural scientific data, to make estimations on scenarios and situations that cannot be empirically measured. For example, and to return to the above example, it is not possible to empirically examine the appropriate size of the evacuation area in case of a nuclear power plant accident. Furthermore, this question is highly complex question, in the sense that must be based on considerations of a large number of factors, pertaining to previous accidents, local climate, the nuclear fuel in question, and so on. Here, the idea of EEM is that a group of experts, with a diverse expertise, can deliberate and reach a useful estimation provided as much and as good background data on the issue as possible.

In our sample of texts, there are no examples where the EEM has been previously applied to estimate incorrect welfare payments. Indeed, among the 1,288 studies, we found no application of the method that deals with criminality, welfare, social insurance, or even estimations of human behaviour in general. This is hardly surprising: whilst the EEM was

developed to measure uncertainty provided undisputable background data, there is no background data of a comparable kind as regards incorrect welfare payments. It is distinctively different to measure the behaviour of nuclear waste, where the properties of nuclear fuel can be measured and validated by scientific experiments in laboratory settings, and human behaviour as related to social institutions. In the Swedish applications of the EEM, the only established fact that informed the experts was the costs of incorrect payments that led to court rulings against the individual receiving assistance. Other than that, there is not background data for the experts to base their estimations on. When turning to the literature that actually deals with welfare crime and incorrect social insurance payments, we have failed to find any applications of the EEM. However, there is a general scepticism against methods that approximate unknown quantities based on estimations by experts or people working within the welfare systems in question (see Gee, Button & Brooks 2010; Brooks, Button & Gee 2012). The main reason is that such estimates are highly susceptible to bias, a phenomenon we shall discuss at length below. Instead, the literature on welfare crime tend to favour thorough controls of randomised probability samples.

Hence, neither the literature on EEM nor the literature on measurements of welfare criminality provides a basis for choosing this method to measure benefit frauds. Provided that large size of our sample and the total absence of social scientific applications, we conclude that the EEM is not an established social scientific method. In the actual public commission reports, there is also notable absence of explicit motives for choosing the method. There is no reference to previous applications of the EEM within this area. However, the FUT-Commission, which produced the first report on incorrect payments using the EEM, informs the readers us that the US Nuclear Regulatory Commission has used the method to estimate the suitability of various bedrock compositions for final deposit of nuclear waste (FUT 2007: 156–157). The reports that follow the FUT-commission refer to the original use of 2007. No arguments in regards to how the method translates to a social scientific setting are provided in any of the reports that we have read nor in the archive material.

In summary, we have failed to find any scientific evidence that the EEM is a suitable method to measure incorrect personal assistance payments. No prior social scientific studies are cited in the public commission reports and no arguments are provided for why the method can be transferred to answer a social scientific research question. No studies in this or related areas were found in our systematic review. It is also clear that the method was developed to make estimations based on natural scientific data, rather than answering questions about human behaviour and social institutions. In the public debate about personal assistance, EEM has been described as based in science, which added credibility to the estimations. Although it is true that EEM partly was developed by academics, for public use when traditional methods are not applicable, it was developed to answer fundamentally different questions as compared to incorrect payments from welfare systems.

A Methodological Examination of the Swedish Commissions

Now, we turn to discuss whether the Swedish applications of the EEM meet established methodological standards of the EEM. There are a number of different versions of the EEM, developed to meet specific types of questions (Morgan & Henrion 1992; Grigore et al. 2016). Despite this, there are some fundamentals that nearly all proponents of the method subscribe to. The fact that the method has been widely used within a number of areas means that there is a rich base of accumulated knowledge about expert elicitation. We have identified four key methodological recommendations in the literature, which we describe below and compare to the Swedish public commissions.

The composition of the expert group and bias

First, it is noted that EEM does not provide measurements of the specific question, but of the level of certainty among experts. For example, an expert group focusing on how far dangerous levels of radioactive waste travels after a nuclear disaster does not contribute with empirical facts about the matter, but with empirical facts about how experts on the matter view the question. If used with great care taken to what the method actually measures, this can then serve as a valuable input for policy decisions. It follows that the quality of expertise will be a key determinant of the quality of the results. In the scientific literature on the EEM, it is taken for granted that the method involves academics. First, because they are expected to have vast knowledge about their area of expertise, but importantly also because of their independence from the policy processes that they give input into. Knol et al. (2010: 7) argues that a first step here is to make a comprehensive reading of the scientific literature in order to set up criteria for the kinds of expertise requested.

This is related to the risk of biases, which is a main concern in the EEM-literature. It has been established over and over again that human beings, due to our propensity to resort to heuristic thinking, are prone to biases. In everyday life, this serves us well, where it is vital that we can make decisions based on scarce information. When facing more complex questions, however, heuristic thinking can easily lead us astray. Bias occurs when our heuristics tricks us to make incorrect judgements, which is a natural result of the fact that human brains are not equipped with objective and advanced statistical programs (Morgan 2014). This means that all of us, no matter our expertise, can resort to bias. In the methodological development of expert elicitation, this is seen as a starting point for how EEM-processes should be set up; by systematising the elicitation process and counter specific kinds of heuristic fallacies, the risk of bias can be minimised (see O'Leary et al. 2009: 381). In the composition of the expert group, it is central that the expert group is as diverse as possible, to avoid group think and that people enter the process with similar biases (Cooke & Goossens 1999: 29;

de Franca Doria et al. 2009: 812; Butler, Thomas & Pintar 2015; Rietbergen 2016: 170–171). Ideally, the experts have different backgrounds and areas of expertise and are not acquainted with each other, stated by Ayyub (2001: 6) to be essential to any elicitation processes. Similarly, Knol et al. (2010: 7) argues that diversity of the expert group is of special importance when the results of the elicitation are likely to have policy implications.

Against this background, it is notable that the Swedish applications of the EEM have expert groups consisting of bureaucrats, primarily employed by the SSIA (see FUT 2007: 171; ESV 2011: 85), which means that the expert groups are far from independent and relatively similar as concerns their professional expertise. They are part of the same organizational structure and are likely to know each other as colleagues. This makes it less likely that the elicitation process will make heuristic fallacies visible, since there is a significant risk that people will enter with similar biases. If employees of SSIA are likely to be biased in any direction, this tendency will be multiplied when many or all experts come from this public agency.

Another form of bias occurs when people are affected by the estimations that they make. According to the US Environmental Protection Agency (USEPA 2009: 39), this form of bias is one of the most important to take into consideration. For example, if a person is likely to be affected by their estimation concerning a certain question, they are less trustworthy than if they would be far removed from any potential consequences. This is completely uncontroversial and well-known among those that work with the EEM. Despite this, by merit of being bureaucrats working with social insurance and welfare fraud, the experts of the Swedish applications are almost certain to be affected by the results. For example, the commissions did lead to resource allocation focusing on fighting welfare crime, working routines were changed, and so on. According to the methodological literature, this should have served as a warning against using employees of state agencies working with welfare crime participating.

This relates to what in the literature is called 'accessibility bias.' We know that people tend to overestimate phenomena that they are familiar with, as seen in examinations of how highly specialised medical professionals tend to inflate the prevalence of diseases that they are working with. In parallel, there is a significant risk that experts working with benefit frauds over-estimate their incidence, as has also been shown by Lensvelt-Mulders et al. (2006: 306). In addition, phenomena that get more attention, in media and in the public debate, are likely to be overestimated as compared to the mundane and common phenomena (Morgan 2014: 7177). This also suggests that public debates and media reports about benefit fraud are likely to factor into results of expert elicitations.

A third form of bias worth mentioning here is called 'overconfidence,' which is based on the tendency to stick with previous judgements, despite new evidence being presented (Mosleh, Bier & Apostolakis 1988: 66–67; Speirs-Bridge et al. 2010; Burgman et al. 2016: 15). Since the most common forms of EEM build on several rounds of estimations that are gradually fine-tuned and adjusted, it is important that experts are not incentivised to put prestige behind their evaluations. It is vital that experts do not see themselves as unquestionable authorities regarding the present matters. It is also worth quoting a bureaucrat of the SSIA that defended the Swedish applications of EEM. Against the charge that there is a risk of bias, he states that knowledge and experience guarantee that estimations are not affected by the public debate (Assistanskoll.se 2017). This is a classic example of overconfidence and the kind of attitude that must be avoided. The idea that one's experience guarantees immunity against bias is actually a source of bias rather than a protection.

We conclude from this that the risk of bias is not accounted for in how the public commissions designed their estimations. In the reports, the discussions on bias are scarce and does not reflect the methodological literature. The composition of the expert groups represents a grave methodological shortcoming judging by the accumulated knowledge of the scientific community. Of course, this does not prove the actual occurrence of bias, but it certainly does expose deep methodological flaws of the public commissions.

Quality of background data

The foundation of EEM is that experts make estimations about questions characterised by uncertainty. This means that the elicitation process by itself cannot provide new knowledge. For the estimations to be of a high quality, there must be good and reliable data in place to base estimations on. When the EEM is used the way it is intended to, background data is often reliable and experimentally tested, as we are dealing with uncertainty related to natural phenomena that can be studied in laboratory settings. The job of the experts, then, is to draw on their scientific competence to generate as good estimations as possible. This, however, is very far from the procedure leading up to the estimates of the level of incorrect personal assistance payments.

To start with, the FUT-commission as well as the ESV started from unclear definitions of what was actually to be measured. This becomes evident in a special statement of the SSIA to the ESV commission (ESV 2011: 158–159), which stated that representatives of different agencies have interpreted the definitions related to incorrect payments differently. In addition to this, the background data that the experts were provided with suffered from several flaws. The only firm measure that the FUT-commission had was the detected incorrect payments by the SSIA, measured to be in the region between 0.8–1.7% of all assistance payments. This was based on probability sample controls, in the literature on benefit frauds regarded as a much more reliable design than methods based on expert judgements (Gee et al. 2010). Nevertheless, the FUT-commission arrived at an estimate of about 10.9% of incorrect payments. One reason for this may have to do with the other background materials that the experts of the FUT-commission were presented with. It is noted in the report of the FUT-commission (FUT 2007:11), that there is a lack of previous

studies that provide good background data. Therefore, the FUT-commission use a previous report, examining possible causes of incorrect payments, as a material to base EEM-estimations on. Hence, the experts used a typology of types of incorrect payments as a background for their estimations of their prevalence. This, of course, is itself a typical example of accessibility bias, resting on a logical flaw; there is no necessary link between the causes of a phenomenon and its occurrence. A second source of background information was provided by interview and survey studies with case managers and representatives of the general public, examining beliefs about the total level of incorrect payments (FUT 2007), suffering from a similar logical flaw: beliefs about the extent of incorrect payments can never be seen as data about the actual extent of incorrect payments.

The lack of reliable data was also pointed out in the commission report:

As pointed to already, there is a very limited amount of available data, both for the particular social insurance systems that we have been looking at and for these systems in total. (FUT 2007:59. *Our translation*).

Following the literature on the EEM, this should be taken as a good reason to avoid using the method; without reliable background data, the EEM is not applicable. However, throughout all of the Swedish applications, the lack of data is instead interpreted as a rationale for choosing the method in the first place. This implies that a serious misunderstanding concerning the method occurred. This is not a method to be used when there is a lack of empirical data, but a method to be used when the empirical data is complex as concerns how it should be interpreted with respect to risk and uncertainty.

Transparency

It comes with the nature of EEM that the precision of measurements are impossible to fully evaluate. This makes it all the more important to guarantee a high level of reliability, that is, making sure that it is possible to recreate how estimations were arrived at (Burgman 2006: 50). A key component of EEM, as described in the scientific literature, is that the estimation process should be as open and transparent as possible. The procedure and the key choices about how to design the process should be made explicit, as well as the individual estimations by the experts and their joint deliberations. For example, Hora and Jensen (2001: 1) argue that documenting and making the experts reasoning public must be seen as a minimum requirement in any application of the EEM, although it may be permissible to skip linking individual experts to specific estimations. Since the EEM builds on subjective judgements, transparency may counter the appearance that the results are arbitrary (USEPA 2009: 25).

The Swedish public commissions do not meet the standards of transparency found in the literature. None of the public commission reports let the readers know how the EEM-rounds have been carried out, which experts participated, how the experts were chosen, how their reasoning went, or how they adjusted their estimations after joint deliberation. In the scientific literature, these are things that are recommended to be presented. Ultimately, this is a question of legitimacy. Naturally, there is a risk that experts are accused of being biased. Making the process as transparent as possible is both a way of guaranteeing that they are not and to counter such accusations. Unfortunately, it is not possible to trace how the results of the Swedish applications of the EEM were drawn, which goes against what is recommended when the method is used in policy processes and by public agencies (USEPA 2009: 11). Ultimately, it makes it harder to judge the quality of the results.

Interpretation of results

The results of expert elicitations can be presented in two different ways: as an uncertainty interval, indicating a range within which it is likely that the actual result is, or as a single estimation, based on the uncertainty interval and the individual estimations of the experts. In the literature, it is generally held that the uncertainty interval is the most important result, because it also serves as an indication of the reliability of the estimation and the level of uncertainty among experts. (Burgman et al. 2006: 31; Grigore et al. 2016). A broad interval is indicative that there is no consensus among experts, suggesting that the results are less reliable than a narrower interval. This information is left out when a single estimation is presented. Indeed, in the literature, it is stressed that there may be a pressure from policymakers to present a set answer, rather than an interval, where the methodological recommendations strongly suggest that such pressures are resisted, to at the least, not compromise the scientific basis of the method. This concerns the fact that the EEM not is a method of measuring a specific empirical phenomenon, but a method measuring the certainty among experts regarding a specific empirical phenomenon. Single estimates convey the message of a scientific precision that the method does not aspire to.

The Billum Commission is of special interest here. As mentioned above, this commission did not proceed with its own EEM-process. Instead, their results were based on previous estimates. In this process, the Billum Commission kept the single estimate, but narrowed the uncertainty interval (SOU 2012:6, p. 320). This is peculiar for two reasons. First, the single estimate is mathematically generated by the uncertainty interval, which means that the method is not properly used if one is kept and the other changed. Furthermore, the rationales for narrowing the uncertainty interval were different as concerns the low and high end of the interval. This is evidence of a grave methodological misunderstanding. Any factor affecting the lower end will also affect the higher end and any factor affecting the higher end will also affect

the lower end. When the Billum Commission changes the high and lower ends of the uncertainty interval for different reasons, they reveal that they have not understood the basics of the method. Lastly, although leaning on the ESV results, the Billum Commission makes an ad hoc definition of what they measure, where they invent 'overuse' as a term included within the 12%. This term is not used in the ESV reports. From a scientific and methodological perspective, this is simply not possible to do. You cannot reinterpret the result of an EEM round to mean something else than was actually estimated by the partaking experts. Taken together, these flaws raise questions as concerns whether the results of the Billum report have any credibility at all.

The Swedish applications of the EEM to measure incorrect payments present results both as uncertainty intervals and as single estimations. As discussed above, the intervals were broad, indicating a correspondingly high level of uncertainty. Nevertheless, and not very surprisingly, the single estimates got most attention – in media and among policy makers. The reports quickly established that 10 or 12% were empirical facts, provided by a scientific method. In this context, the USEPA's (2009: 7) warning of single estimates generating a false sense of precision is important to bear in mind. These 'facts', in turn, came to influence policy making, for example by allocating resources to fight benefit fraud and initiating new commissions. More generally, these estimates have repeatedly been used in the political rhetoric in order to justify general cutbacks (Altermark 2017). But in actuality, the single estimates are not useful without the uncertainty intervals, indicating a high level of insecurity as concerns the results.

Conclusion

This article presents the first examination of how the EEM has been used by Swedish authorities to estimate incorrect personal assistance payments. These estimates have been important sources of justification for policy decisions on personal assistance, from 2014 and until today. As was highlighted in the introduction, expert elicitation is now spreading to other Scandinavian countries. Provided that a rhetoric focusing on incorrect payments is recurring in justifications of cutbacks on welfare, it is important to examine the quality of EEM-measurements in this area.

Our study finds that there is no evidence to support applying the method to estimate incorrect personal assistance payments. Within this area, there is a lack of reliable background data and we have found no previous studies that the public commissions could have learnt from. In addition, we find that the public commissions applications of the EEM fails to consider agreed upon principles of how the method should be used. Most problematic is the fact that partaking experts are not independent, but work within state agencies that are likely to be affected by results. This reflects a more general failure to account for the risk of bias in the research design. The fact that the last public commission drawing on EEM – the Billum Commission – misinterprets the results testifies to a lacking understanding of the method.

Our overall conclusion is that EEM estimations of incorrect payments in Sweden do not constitute a sound basis of policy decisions. The fact that estimations generated by this method has shaped the public debate on personal assistance and has been used to legitimise efforts to reduce costs is problematic provided a perspective that see a well-informed debate as paramount for democratic decision-making.

Competing Interests

The authors have no competing interests to declare.

References

- Altermark, Niklas. 2017. "Hur legitimeras socialpolitiska besparingar? Konstruktionen av personlig assistans som ett 'kostnadsproblem'". *Socialvetenskaplig tidskrift* 17(2): 107–126.
- Assistanskoll.se. 2017. "Thomas Falk, talesperson på Försäkringskassans kontrollenhet – 'Expertmetoden ger kvalificerade bedömningar av fusket'". March 15, 2017. Accessed December 4, 2018. <https://assistanskoll.se/20170317-Falk-Forsekringskassans-kontrollenhet-Expertmetoden-fusket.html>.
- Assistanskoll.se. 2018. "Statistik om personlig assistans". Accessed December 4. <https://assistanskoll.se/assistans-statistik.php>.
- Ayyub, Bilal M. 2001. *Elicitation of expert opinions for uncertainty and risks*. CRC press. DOI: <https://doi.org/10.1201/9781420040906>
- Begler, Ann-Marie, and Adriana Lender. 2016. "Bilag och brukare har fel om kris i assistansen". *Dagens Nyheter*, June 6, 2016. Accessed December 2, 2018. <https://www.dn.se/debatt/bilag-och-brukare-har-fel-om-kris-i-assistansen/>.
- Boaz, Anette, Deborah Ashby, and Ken Young. 2002. "Systematic Reviews: What have they go to offer evidence based policy and practice?". London: ESRC UK Centre for Evidence Based Policy and Practice.
- Brooks, Graham, Mark Button, and Jim Gee. 2012. "The scale of health-care fraud: A global evaluation". *Security Journal* 25(1): 76–87. DOI: <https://doi.org/10.1057/sj.2011.7>
- Butler, Ainslie J., Kate Thomas, and Katarina D. M. Pintar. 2015. "Systematic review of expert elicitation methods as a tool for source attribution of enteric illness". *Foodborne Pathogens and Disease* 12(5): 367–382. DOI: <https://doi.org/10.1089/fpd.2014.1844>
- Cooke, R. M., and L. J. H. Goossens. 1999. "Procedures guide for structured expert judgment". *Project Report to the European Commission*. Accessed December 4, 2018. http://rogermcooke.net/rogermcooke_files/PROCGEUR18820.pdf.

- Davies, Alan. 2019. "Carrying out systematic literature reviews: An introduction". *British Journal of Nursing* 28(15): 1008–1014. DOI: <https://doi.org/10.12968/bjon.2019.28.15.1008>
- de Franca Doria, Miguel, Emily Boyd, Emma L. Tompkins, and W. Neil Adger. 2009. "Using expert elicitation to define successful adaptation to climate change". *Environmental Science & Policy* 12(7): s. 810–819. DOI: <https://doi.org/10.1016/j.envsci.2009.04.001>
- Delegationen mot felaktiga utbetalningar. 2007. Vad kostar felet? Omfattning av felaktiga utbetalningar från trygghetssystemen. FUT 2007:7.
- Dir. 2011:26. Kommittédirektiv. *Utredning om åtgärder mot fusk, oegentligheter och överutnyttjande av den statliga assistansersättningen*. Accessed December 4, 2018. <https://www.regeringen.se/49bbc6/contentassets/847cbf87a2c4457fbfb8d3b8178df20c/utredning-om-atgarder-mot-fusk-oegentligheter-och-overutnyttjande-av-den-statliga-assistansersattningen-dir.-201126>.
- Ekonomistyrningsverket. 2011. Samverkansuppdrag mot felaktiga utbetalningar från välfärdssystemen. ESV 2011:11.
- Folcker Aschan, Annika. 2011. "Bedrägerier med assistenter utreds". *Dagens Nyheter*, March 25, 2011. Accessed December 4, 2018. <https://www.dn.se/arkiv/nyheter/bedragerier-med-assistenter-utreds/>.
- Fraser, Nancy, and Linda Gordon. 1997. "'Dependency' demystified. Inscriptions of power in a keyword of the welfare state". In *Contemporary political philosophy: An anthology*, edited by Philip Pettit, and R. E. Goodin. Chichester: Wiley.
- Gee, Jim, Mark Button, and Gramham. 2010. *The financial cost of UK Public Sector Fraud. A less painful way to reduce public expenditure*. Center for Counter Fraud Studies, University of Portsmouth.
- Gough, David, and Diana Elbourne. 2002. "Systematic Research Synthesis to Inform Policy, Practice and Democratic Debate". *Social Policy & Society* 1(3): 225–236. DOI: <https://doi.org/10.1017/S147474640200307X>
- Grigore, Bogdan, Jaime Peters, Christopher Hyde, and Ken Stein. 2016. "A comparison of two methods for expert elicitation in health technology assessments". *BMC Medical Research Methodology*. DOI: <https://doi.org/10.1186/s12874-016-0186-3>
- Johnson, Björn. 2010. *Kampen om sjukfrånvaron*. Lund: Arkiv förlag.
- Knol, Anne B., Pauline Slottje, Jeroen P. van der Sluijs, and Erik Lebret. 2010. "The use of expert elicitation in environmental health impact assessment: a seven step procedure". *Environmental Health* 9: 19; 1–16. DOI: <https://doi.org/10.1186/1476-069X-9-19>
- Landeta Rodriguez, Jon. 2006. "Current validity of the Delphi method in social sciences". *Technological Forecasting and Social Change* 73(5): 467–482. DOI: <https://doi.org/10.1016/j.techfore.2005.09.002>
- Lensvelt-Mulders, Gerty, Peter van der Heijden, Olav Laudy, and Ger van Gils. 2006. "A validation of computer-assisted randomized response survey to estimate the prevalence of fraud in social security." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 169(2): 305–318. DOI: <https://doi.org/10.1111/j.1467-985X.2006.00404.x>
- Lundström, Ragnar. 2011. *Den kalkylerande medborgaren: bidragsfusk i svensk välfärdsdebatt 1990–2010*. PhD Diss, Umeå universitet.
- Morgan, Granger. 2014. "Use (and abuse) of expert elicitation in support of decision making for public policy". *PNAS* 111(20): 7176–7184. DOI: <https://doi.org/10.1073/pnas.1319946111>
- Morgan, Granger, and Max Henrion. 1992. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.
- Mosleh, A., V. M. Bier, and G. Apostolakis. 1988. "A critique of current practice for the use of expert opinions in probabilistic risk assessment." *Reliability Engineering & System Safety* 20(1): 63–85. DOI: [https://doi.org/10.1016/0951-8320\(88\)90006-3](https://doi.org/10.1016/0951-8320(88)90006-3)
- Näsman, Mattias. 2016. "Konsekvenser av indragen assistansersättning. En studie om medikalisering och hur kostnader skiftar från samhället till brukaren och dess anhöriga. Rapport för STIL om förändrat stöd enligt LSS och SFB". *Umeå Papers in Economic History* No 46/2016.
- O'Leary, Rebecca A., Samantha Low Choy, Justine V. Murray, Mary Kynn, Robert Denham, Tara G. Martin, and Kerrie Mengersen. 2009. "Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby *Petrogale penicillata*". *Environmetrics* 20(4): 379–398. DOI: <https://doi.org/10.1002/env.935>
- Ouchi, Fumika. 2004. "A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis". *World Bank Policy Research Working Paper 3201*. DOI: <https://doi.org/10.1596/1813-9450-3201>
- Regné, Åsa. 2016a. "Assistanspengarna måste gå till rätt saker". *Dagens Samhälle*, March 17, 2015. Accessed December 4, 2018. <https://www.dagenssamhalle.se/debatt/assistanspengarna-maste-ga-till-raett-saker-23564>.
- Regné, Åsa. 2016b. "Välfärden får kosta – men pengarna ska gå till dem som behöver dem". *Metro*, March 31, 2016. Accessed December 4, 2018. <https://www.metro.se/artikel/valfarden-far-kosta-men-pengarna-ska-ga-till-dem-som-behoever-dem-xr>.
- Rietbergen, Charlotte, Rolf H. H. Groenwold, Herbert J. A. Hooijink, Karl G. M. Moons, and Irene Klugkist. "Expert elicitation of study weights for Bayesian analysis and meta-analysis". *Journal of Mixed Methods Research*, 10(2): 168–181. DOI: <https://doi.org/10.1177/1558689814553850>

- Siddaway, Andy P., Alex M. Wood, and Larry W. Hedges. 2019. *Annual Review of Psychology* 70: 747–770. DOI: <https://doi.org/10.1146/annurev-psych-010418-102803>
- Socialdepartementet. 2015. *Regleringsbrev för budgetåret 2016 avseende Försäkringskassan*.
- SOU. 2012:6. Åtgärder mot fusk och felaktigheter med assistansersättning. Accessed December 2, 2018. <https://www.regeringen.se/contentassets/c6e9db55b1aa4026be575ed62fdf829c/atgarder-mot-fusk-och-felaktigheter-med-assistansersattning-hela-dokumentet-sou-20126>.
- Speirs-Bridge, Andrew, Fiona Fidler, Marissa McBride, Louisa Flander, Geoff Cumming, and Mark Burgman. 2010. "Reducing Overconfidence in the Interval Judgements of Experts." *Risk Analysis: An International Journal* 30(3): 512–523. DOI: <https://doi.org/10.1111/j.1539-6924.2009.01337.x>
- USEPA. 2009. *Expert Elicitation Task Force White Paper*. Washington, DC: Environmental Protection Agency.

How to cite this article: Altermark, Niklas, and Hampus Nilsson. (2020). State Measurements of Benefit Fraud: Why Expert Elicitations Cannot Be Used to Measure Incorrect Personal Assistance Payments. *Scandinavian Journal of Disability Research*, 22(1), pp. 158–167. DOI: <https://doi.org/10.16993/sjdr.610>

Submitted: 13 December 2018 **Accepted:** 09 March 2020 **Published:** 19 May 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Scandinavian Journal of Disability Research is a peer-reviewed open access journal published by Stockholm University Press.

OPEN ACCESS The Open Access logo, consisting of the words 'OPEN ACCESS' followed by a circular icon containing a stylized 'O'.