Disease Control Priorities in Developing Countries, 3rd Edition
Working Paper #7

Quantifying uncertainty in intervention effectiveness: an application to obstetric fistula

Author (1): Abigail Colson

Affiliation: Center for Disease Dynamics, Economics & Policy, Washington, DC

Author (2): Sweta Adhikari

Affiliation: Center for Disease Dynamics, Economics & Policy, Washington, DC

Author (3): Ambereen Sleemi

Affiliation: Maimonides Medical Center, Brooklyn, NY
Columbia University Mailman School of Public Health, New York, NY

Author (4): Ramanan Laxminarayan

Affiliation: Center for Disease Dynamics, Economics & Policy, Washington, DC
Public Health Foundation of India, New Delhi, India
Princeton Environmental Institute, Princeton University, Princeton, NJ

Correspondence to: Ramanan Laxminarayan
ramanan@cddep.org

Keywords: structured expert judgment, expert elicitation, fistula

Abstract:
Intervention effectiveness is a major input of evidence-informed priority setting in healthcare, but information on intervention effectiveness is generally lacking. This is particularly problematic in the context of poorly resourced health care settings where even efficacious interventions fail to translate into improvements in health. The few intervention effectiveness studies focus on the experience of single facilities and do not consider the impact of multiple factors that may affect health outcomes.

Methods and Findings:
We use the Cooke method of structured expert judgment to assess the effectiveness of surgery to correct obstetric fistula in a low-income setting. This method has been used to quantify uncertainty in the areas of engineering and environmental risk assessment when data are unavailable. Expert judgment is the basis for irreversible decisions with far reaching consequences, but new to the realm of public health. Under this method, experts quantify their uncertainty about rates of long-term disability in fistula patients following treatment in different contexts, but the information content drawn from their responses is statistically conditioned upon the accuracy and informativeness of their responses to a set of calibration questions. Through this method, we develop best estimates and uncertainty bounds for the rate of disability associated with each treatment scenario and setting.

Conclusions:
Estimates developed using the Cooke method were statistically superior to those involving a simple averaging of expert responses. We find that the Cooke method is a viable approach to investigating the effectiveness of medical interventions where randomized controlled trials are not possible. Understanding the effectiveness of surgery performed at different types of facilities can guide program planning to increase access to fistula treatment.

Background

Obtaining accurate estimates of intervention effectiveness is a major challenge in evidence-based global health. Intervention efficacy is best estimated through randomized controlled trials, but costs and logistical challenges make these infeasible in many instances. Randomized controlled trials may also not be possible due to ethical concerns related to not treating or providing substandard treatment to a subset of patients. Questions concerning treatment effectiveness often persist following a randomized controlled trial because such studies may have limited external validity—efficacy in randomized trials may not always translate to effectiveness under real world health systems. Intervention efficacy does not necessarily translate from the study site and context to other settings. Efficacy studies focused on individual interventions also have limitations when applied to estimating the effectiveness of packages of interventions. The total effectiveness of a package may be less or greater than the sum effectiveness of the individual parts, but traditional efficacy studies offer no guidance on determining the effect size of the total package.

The Cooke method of structured expert judgment allows analysts to utilize expert opinion to overcome gaps in the literature to better understand the effectiveness of an intervention or set of interventions. The Cooke method is a technique for pooling expert opinion to create rational consensus on point estimates and quantify the surrounding uncertainty.[1] This study is the first application of the Cooke method of structured expert judgment to the determination of intervention effectiveness and long-term consequences. We apply the Cooke method to the question of long-term disability following treatment for obstetric fistula, in order to assess the method's applicability to quantifying uncertainty around intervention effectiveness.

Obstetric fistula is a major source of maternal morbidity in low and lower-middle income countries. An estimated 3·5 million women currently suffer from fistula, with 130,000 new cases occurring every year.[2] The 2010 Global Burden of Disease (GBD) study estimates that obstetric fistula results in 1·15 million years lived with disability (YLDs), approximately 64 percent of the total YLDs attributable to all maternal conditions.[3] This high burden is despite the fact that obstetric fistula is generally both preventable and treatable.

Fistula is typically a result of surgical error in high-income countries, but in lower-income settings fistula is largely a consequence of obstructed labor. Without access to emergency obstetric care, the internal trauma of an obstructed labor that may last several days causes tissue death in the woman's birth canal. The dead tissue sloughs away, leaving a fistula between the vagina and bladder, rectum, or both. A woman with fistula may suffer from constant urinary or fecal incontinence as well as other complications caused by the internal damage. The incontinence can also lead to social and sexual ostracism, introducing psychological complications. Women with fistula are often blamed for their condition, exacerbating the social and psychological harm.

Fistula can be successfully treated with surgery. Fistula centers, such as the Addis Ababa Fistula Hospital in Ethiopia, specialize in the care and treatment of fistula patients. Despite the high burden of disease and the existence of dedicated medical facilities, the literature on the efficacy and long-term consequences of fistula surgery is relatively sparse. Estimates of urinary incontinence, an important outcome following fistula surgery, in patients following fistula repair

vary widely, from 8-50 percent.[4–8] Recent literature reviews find that existing studies are poorly done, with few randomized studies, poorly defined outcome measures, and most research focusing on the experience of one clinic or even one surgeon.[9–11] Most studies also track outcomes only a couple of weeks post-surgery,[12,13] but incontinence can change over longer periods of time following discharge from the hospital.[5,11] While research examines the effect of patient and fistula characteristics on outcomes, no study explores the impact of contextual factors, such as surgeon experience, type of hospital, or whether a surgery was performed as part of a training or outreach program.[10] Understanding the role of these factors in determining surgery outcomes is necessary when planning programs that expand access to fistula surgery.

However, collecting long-term outcomes data on women who underwent fistula surgery is challenging. Fistula patients often live in remote areas, disconnected and far removed from the clinics where they sought treatment, making follow-up difficult.[5–7] To overcome the lack of existing data sources relevant to the long-term disability of fistula patients, we apply the Cooke method of structured expert judgment to explore two questions:

1. What is the long-term disability associated with different types of fistula following surgical treatment?
2. Do long-term outcomes differ for patients treated in different settings (such as a high-volume fistula center versus a district hospital that treats a low-volume of fistula patients)?

Methods

The Cooke method of structured expert judgment has been used to establish point-estimates and quantify uncertainty using experts in fields including aviation, engineering, environmental health, nuclear safety and climate change.[1,14,15] Whereas other forms of expert judgment—committees or the Delphi method, for example—seek to eliminate uncertainty by forcing agreement between expert participants, the Cooke method determines point estimates and then exploits experts' existing uncertainty to more accurately establish confidence intervals for the estimates. Experts are presented with a series of questions asking for potentially observable quantities, such as the rate of long-term disability in a given population of fistula patients. For each question experts provide a median estimate and 90% and 50% credible ranges (that is, a range the expert believes there is a 90% chance the true value falls within and a narrower range the expert believes there is a 50% chance within which the true value falls), creating the expert's uncertainty distribution for the question. Experts' distributions are combined in two ways: first assigning equal-weight to all experts and second assigning experts weight based on their performance.

Performance weights are determined by the experts' assessments on a set of calibration questions for which answers are known to the study team but unknown to the experts. Experts are scored according to the statistical accuracy and informativeness of their assessments. Statistical accuracy is based on testing the hypothesis that the true values of the calibration questions (that is, the actual answers) could be jointly drawn from an expert's provided distributions. An expert with high statistical accuracy captures the true values within his or her ranges at the expected frequency. That is, 90% of calibration question true values fall within the 90% credible range and 50% of the true values fall within the 50% credible range. Informativeness is a measure of

how peaked an expert's uncertainty distributions are, with more peaked distributions indicating a narrower range of values in the credible range and thus less uncertainty. Less peaked assessments encompass a wider range of credible values—indicating more uncertainty—and thus provide less information on the indicator in question.

Calibration questions allow analysts to assess an expert's ability to accurately quantify his or her uncertainty on questions in the field of interest. Calibration questions also create the basis for identifying the optimal performance-weighted combination of experts.[14] In this study, calibration questions focused on the epidemiology of obstetric fistula, including the prevalence of fistula in different subpopulations and rates of associated indicators of maternal well-being. Calibration questions were not designed to identify the expert most skilled at performing fistula surgery, but rather the experts able to best think about average outcomes and the surrounding uncertainty for a generic set of fistula patients.

We identified and recruited eight experts to participate in the study. Experts included a mix of urologists and gynecologists, most with experience predominately in Africa. Two members of the study team conducted in-person, two-on-one interviews at the International Society of Obstetric Fistula meeting held in Dhaka, Bangladesh in mid-November, 2012. In addition to gathering information on experts' uncertainty distributions for the set of questions, we asked experts to explain their thinking and provide rationales for their assessments. These narratives were captured alongside the uncertainty distribution data.

Results

The structured expert judgment protocol included 10 calibration questions and 23 variables of interest questions. Variables of interest questions focused on five scenarios (Table 1). For each scenario experts were asked to predict how many patients, given 1000 such cases, would develop long-term disability if treated in one of four settings: a hospital in a high-income country, a high-volume fistula center in a low-income country, a low-volume district hospital in a low-income country, and if untreated. For Scenario 3 experts were also asked how many of 1000 such cases would result in death if treated in the different settings.

Figures 1 through 6 show the equal-weight (EW) and performance-weight (PW) pooled combinations of experts, referred to as decision makers (DMs), for each variable of interest. The median estimate of the PWDM is lower than that of the EWDM for all 23 variables of interest, indicating lower rates of long-term disability and death for each scenario following any treatment or if untreated. The PWDM's 50% credible ranges are narrower than the EWDM's for all 10 calibration questions and 21 of the 23 variables of interest. The PWDM's 90% credible ranges are narrower for 9 of the 10 calibration questions and 21 of 23 variables of interest. This is the benefit of the Cooke method of structured expert judgment: performance weights typically yield pooled assessments that are more informative than the equal-weight combination while remaining at least as statistically accurate.[14]

The DMs indicated that experts were most certain about long-term outcomes when patients received treatment in high-income countries or high-volume fistula centers. Uncertainty

increased when experts thought about outcomes following treatment in low-volume district hospitals or if patients did not receive treatment.

Experts disagreed on the likelihood of long-term disability for women suffering from fistula who receive treatment in low-volume facilities or do not receive any treatment. Figure 2 shows the individual expert responses and pooled DMs for the question asking about Scenario 2 patients treated in a low-volume district hospital. The experts' assessments reflected two divergent viewpoints on the long-term outcomes. One group believed that even in a low-volume district hospital, facilities will have the resources necessary to adequately treat the patient, and thus long-term disability will not be common. A second group of experts were of the opinion that district hospital doctors would be too unskilled to prevent long-term disability in these patients (Supplement 1 includes individual expert assessments for all variable of interest questions).

Discussion

In this example application of the Cooke method as a method for estimating the effectiveness of health interventions, the PWDM indicates lower rates of disability and death for each scenario and treatment facility, including no treatment, than the EWDM. The PWDM also shows less uncertainty around its median estimates than the EWDM. Uncertainty increases, though, when experts shift from thinking about outcomes in high-volume centers to low-volume centers or among untreated women with fistula. The experts we interviewed work predominately in high-volume fistula centers. Existing short-term outcome research also focuses on patients who receive care at fistula centers.[10,12] Experts are most familiar with outcomes for patients in this

setting, and is reflected in the narrower uncertainty ranges given for patients treated in high-volume fistula centers. In contrast, the experts we interviewed do not see as many patients in low-volume district hospitals and, by definition, do not see any untreated women with fistula. Data on short- and long-term outcomes in these settings does not exist. Thus, the experts' assessments show greater uncertainty when considering long-term outcomes in these settings.

Experts expressed a large amount of uncertainty about long-term outcomes for patients who were not treated. Differences in the median PWDM estimates for untreated patients across the scenarios, though, suggest some untreated fistulas are more problematic than others. The low median values for untreated patients in Scenarios 2 and 3 indicates experts thought those cases could resolve even if untreated, whereas that was much less likely for the other scenarios.

In all scenarios, the PWDM signifies the rate of long-term disability or death is low for patients treated in high-volume fistula centers. The PWDM also shows low rates of disability or death for patients treated in low-volume district hospitals, but the higher upper bounds on the 50% and 90% credible ranges indicate experts are less certain about outcomes following treatment in this setting. Only one scenario, Scenario 1, has a PWDM median rate of disability substantially higher in low-volume versus high-volume hospitals, indicating this is a more complicated case that experts know benefits from specialized care. The PWDM for Scenario 2, in contrast, shows little difference between the ranges of rates for high- and low-volume settings, suggesting this case may be easier to repair. Identifying types of cases that require experienced, specialized care versus cases that can be successfully treated by less experienced surgeons is useful for planning and implementing fistula treatment programs.

A recent review by Arrowsmith and colleagues of the existing literature on fistula surgery success rates identifies reported closure rates of 53 to 97·5 percent, with an average of successful closure in 86 percent of patients. Defining success as "no incontinence" rather than closure lowers reported success rates. The average of reported rates of no incontinence is 70 percent, with a range of 42 to 92 percent.[11] Experts in our study were instructed to consider incontinence as a potential long-term disability following fistula repair surgery. Estimated surgery success rates across the five scenarios in this study are higher than those identified in existing research by Arrowsmith and colleagues. Here, the PWDM identifies median surgery success rates, based off of our definition of long-term disability, between 98·9 and 99·7 percent in our scenarios for patients treated in a high-volume fistula hospital. Lower bounds of success (based on the given 95-percentile rates of disability) range from 72·8 to 83·7 percent, and upper bounds (based on the 5-percentile rates of disability) are all near 100 percent.

Several factors explain why success rates in this study may be higher than those in existing studies. First, all of our cases focus on the first attempt to repair fistula. Surgery outcomes worsen if patients have a history of previously unsuccessful repair, and observational studies based on patients at a single fistula center may include a high proportion of patients referred from other facilities after a previously failed repair attempt.[11] Second, continence may improve time, and few studies look at long-term repair outcomes, due to the challenge of off-site follow-up.[5,11] Thus, existing studies may under-report the actual rate of no incontinence following fistula repair.

We were unable to compare our results to those from GBD 2010 because GBD disability weights sum the impact of all fistulas whereas we estimate outcomes within certain subsets of fistula. Without knowledge of the frequency with which these scenarios occur and the rate at which patients seek treatment for fistula, it is not possible or reasonable to translate our estimates into a single disability weight. Although SEJ can be a useful method for estimating disability weights, our objective in this study was slightly different. Rather than quantify the total disability attributable to obstetric fistula, we aimed to estimate the rate of disability associated with untreated and treated fistula, to better understand the effectiveness of fistula surgery.

Structured expert judgment provides estimates and uncertainty ranges for rates of long-term disability and death in a variety of treatment contexts. Such estimates would not be possible with existing data sources, and collecting the necessary data in the future would be difficult, unethical and/or expensive. Structured expert judgment also presents a new opportunity to estimate long-term success rates in context when long-term outcomes are difficult to track. The Cooke method presents a new opportunity to study questions of efficacy and effectiveness where these data limitations exist.

Acknowledgment

This work was funded by the Bill & Melinda Gates Foundation, through the Disease Control Priorities Network at the University of Washington.

References

1 Cooke RM. Experts in Uncertainty: Opinion and Subjective Probability in Science. New York, Oxford University Press, 1991.

2 Wall L. Obstetric vesicovaginal fistula as an international public-health problem. *Lancet* 2006; **368**: 1201–9.

3 Vos T, Flaxman AD, Naghavi M, *et al.* Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2163–96.

4 Browning A. Risk factors for developing residual urinary incontinence after obstetric fistula repair. *BJOG* 2006; **113**: 482–5.

5 Browning A, Menber B. Women with obstetric fistula in Ethiopia: a 6-month follow up after surgical treatment. *BJOG* 2008; **115**: 1564–9.

6 Nielsen HS, Lindberg L, Nygaard U, *et al.* A community-based long-term follow up of women undergoing obstetric fistula repair in rural Ethiopia. *BJOG* 2009; **116**: 1258–64.

7 Murray C, Goh JT, Fynes M, Carey MP. Urinary and faecal incontinence following delayed primary repair of obstetric genital fistula. *BJOG* 2002; **109**: 828–32.

8 Cook RJ, Dickens BM, Syed S. Obstetric fistula: the challenge to human rights. *Int J Gynaecol Obstet* 2004; **87**: 72–7.

9 Kelly J, Winter HR. Reflections on the knowledge base for obstetric fistula. *Int J Gynaecol Obstet* 2007; **99 Suppl 1**: S21–4.

10 Barone MA, Frajzyngier V, Ruminjo J, *et al.* Determinants of postoperative outcomes of female genital fistula repair surgery. *Obstet Gynecol* 2012; **120**: 524–31.

11 Arrowsmith SD, Barone MA, Ruminjo J. Outcomes in obstetric fistula care: a literature review. *Curr Opin Obstet Gynecol* 2013; **25**: 399–403.

12 Frajzyngier V, Ruminjo J, Barone MA. Factors influencing urinary fistula repair outcomes in developing countries: a systematic review. *Am J Obstet Gynecol* 2012; **207**: 248–58.

13 De Ridder D. An update on surgery for vesicovaginal and urethrovaginal fistulae. *Curr Opin Urol* 2011; **21**: 297–300.

14 Cooke RM, Goossens LLHJ. TU Delft expert judgment data base. *Reliab Eng Syst Saf* 2008; **93**: 657–74.

15 Bamber JL, Aspinall WP. An expert judgement assessment of future sea level rise from the ice sheets. *Nat Clim Chang* 2013; **2**: 1–4.

Table 1. Scenarios described in variables of interest questions. Questions asked about the rate of long-term disability and, for Scenario 3, the rate of death per 1000 cases treated at different facility types or untreated.

Scenario	Case Description
1	An 18 year-old woman had obstructed labor and delivery of a stillbirth one month ago. She has a large vesicovaginal fistula that obliterated the anterior vaginal wall, resulting in total loss of the urethra. Upon examination, she has involvement of both ureters, with partial obstruction of one. The main long-term complications are constant leakage of urine (urinary incontinence) and functional loss of a kidney.
2	A 22-year-old woman presents with urinary incontinence after having a stillbirth. She has a small fistula but bilateral footdrop. Long-term disability includes urinary incontinence with severe hip and leg pain, difficulty mobility and atrophy of the lower extremities.
3	A 27-year old patient presents after laboring for 3 days in her village during her 3rd pregnancy. She undergoes an emergency cesarean delivery of a stillbirth. Possible long-term outcomes include: loss of her uterus leading to infertility, and damage to the bladder and/or ureters due to the difficulty nature of the surgery.
4	A 17-year old woman presents in labor with obstruction for the last 24 hours. She delivers a live infant who has some transient respiratory depression, but otherwise appears well. She develops a rectovaginal fistula on post partum day one. Potential long-term disability includes severe fecal incontinence, imposed isolation from her family, and depression (This question asked experts about rates of both long-term disability and death).
5	A 30-year-old woman underwent an emergency cesarean delivery after presenting to a hospital with severe antepartum hemorrhage at term. Her recovery is complicated by constant leakage of urine during her post-operative course, despite placement of a foley catheter for drainage. She is diagnosed with a right-sided ureterovaginal fistula. Potential long-term disability is severe urinary incontinence, depression, isolation, and possible loss of kidney function.

Figure 1. Equal-weight (EW) and performance-weight (PW) decision makers' uncertainty ranges on the rate of long-term disability for Scenario 1 cases treated in different facilities.[1]

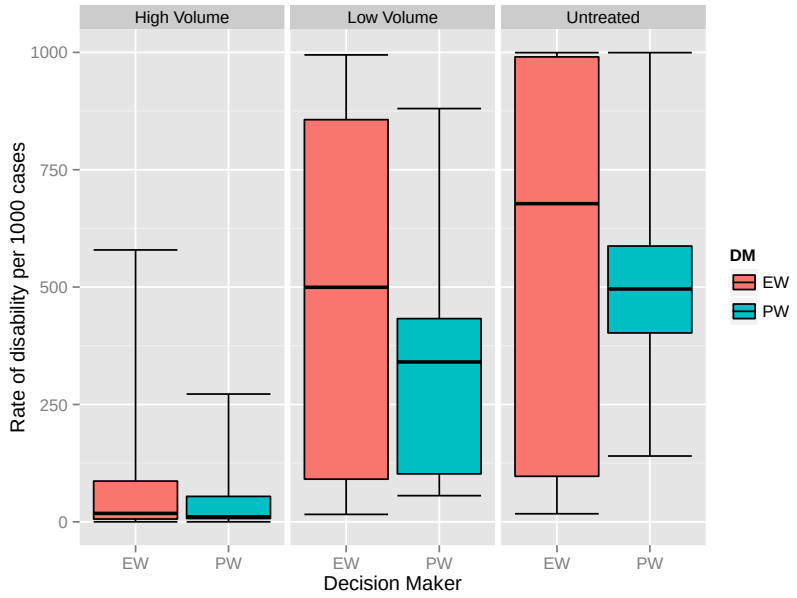
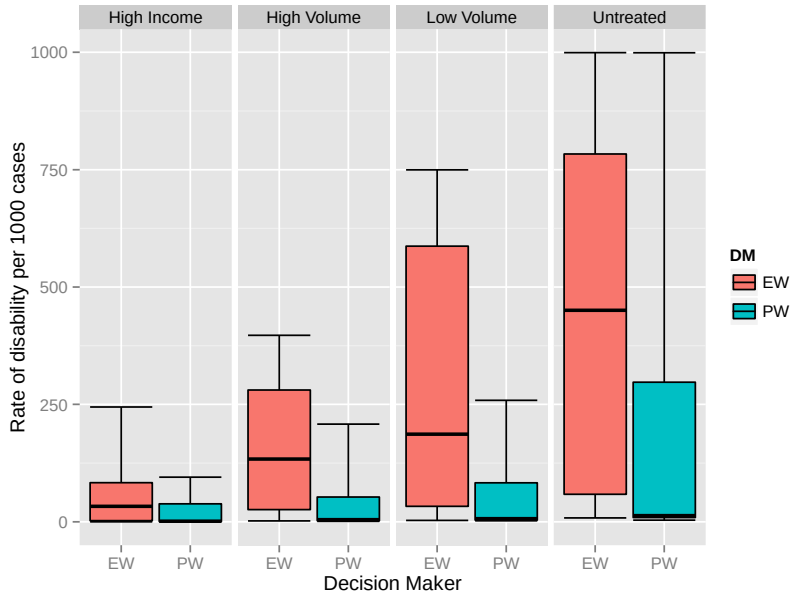


Figure 2. Equal-weight (EW) and performance-weight (PW) decision makers' uncertainty ranges on the rate of long-term disability for Scenario 2 cases treated in different facilities.



[1] For all figures, boxplots indicate the 5, 25, 50, 75, and 95 percentile values from the uncertainty distributions.

Figure 3. Equal-weight (EW) and performance-weight (PW) decision makers' uncertainty ranges on the rate of long-term disability for Scenario 3 cases treated in different facilities.

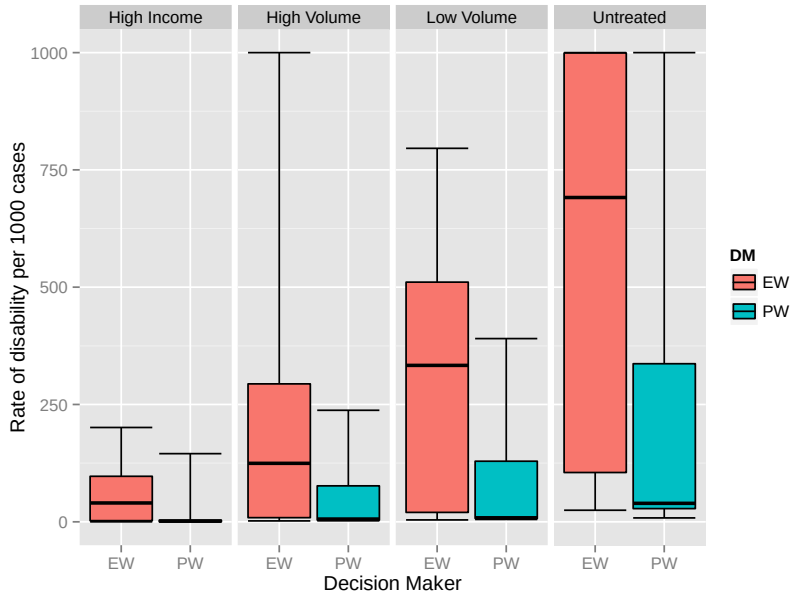


Figure 4. Equal-weight (EW) and performance-weight (PW) decision makers' uncertainty ranges on the rate of death for Scenario 3 cases treated in different facilities.

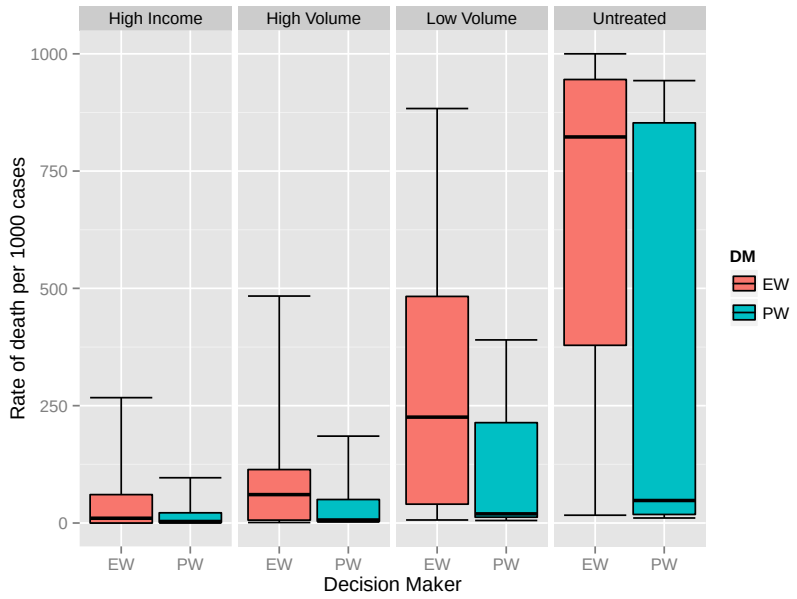


Figure 5. Equal-weight (EW) and performance-weight (PW) decision makers' uncertainty ranges on the rate of long-term disability for Scenario 4 cases treated in different facilities.

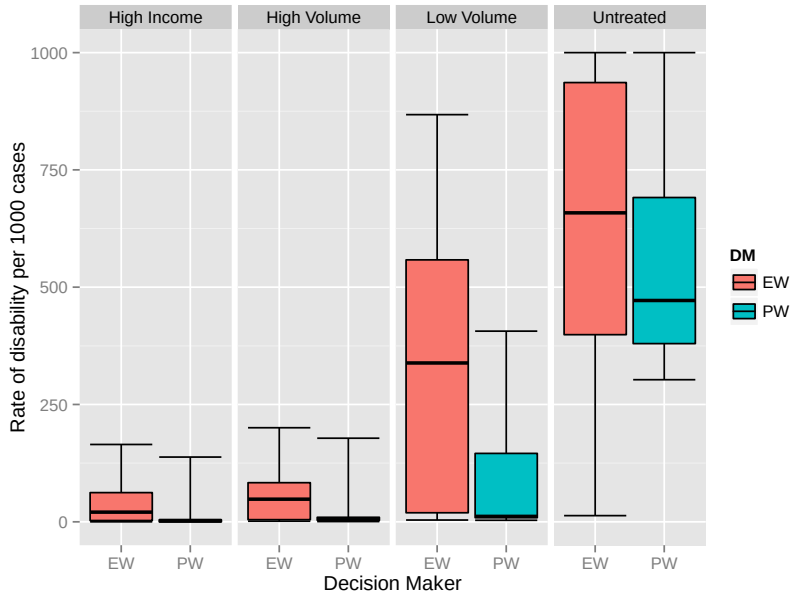


Figure 6. Equal-weight (EW) and performance-weight (PW) decision makers' uncertainty ranges on the rate of long-term disability for Scenario 5 cases treated in different facilities.

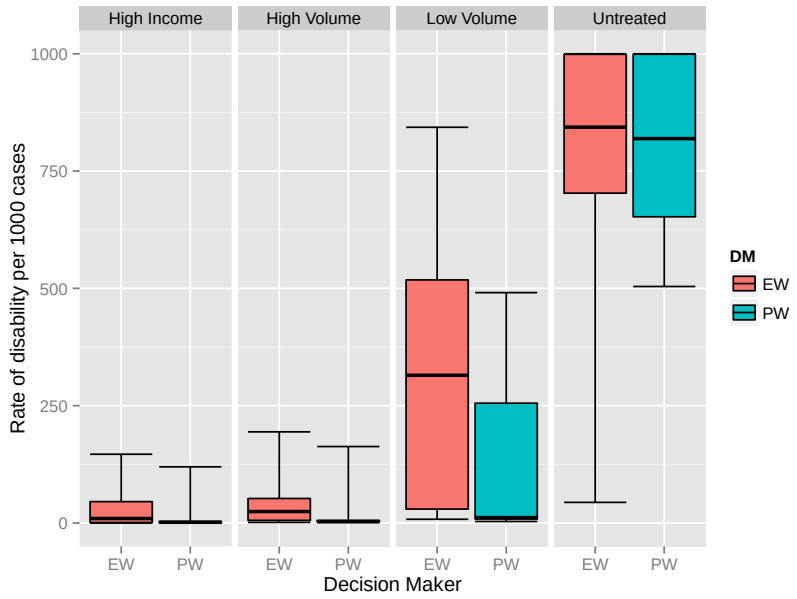


Figure 7. Individual expert and pooled decision maker assessments for Scenario 2 when patients receive treatment in a low-volume district hospital.