

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Carvalho, N;Fish, D;Grant, GM;Salomon, JA;Studdert, DM

Title:

Feasibility of a Health-Utility Approach to Quantifying Noneconomic Losses from Personal Injury

Date:

2018-06-01

Citation:

Carvalho, N., Fish, D., Grant, G. M., Salomon, J. A. & Studdert, D. M. (2018). Feasibility of a Health-Utility Approach to Quantifying Noneconomic Losses from Personal Injury. *Journal of Empirical Legal Studies*, 15 (2), pp.278-319. <https://doi.org/10.1111/jels.12179>.

Persistent Link:

<https://hdl.handle.net/11343/283863>

FEASIBILITY OF A HEALTH-UTILITY APPROACH TO QUANTIFYING NONECONOMIC LOSSES FROM PERSONAL INJURY

Natalie Carvalho, David Fish, Genevieve M. Grant, Joshua A. Salomon, David M. Studdert *

Address correspondence to David Studdert, Stanford Law School, 559 Nathan Abbott Way, Stanford, CA 94305; studdert@stanford.edu. Dr. Carvalho is a Research Fellow at the Melbourne School of Population and Global Health, University of Melbourne, Australia. Dr. Fish is a Senior Lecturer in the School of Public Health and Preventative Medicine, Monash University, Melbourne, Australia. Dr Grant is a Senior Lecturer in the Faculty of Law at Monash University, Clayton, Australia. Dr Salomon is a Professor of Medicine at the Center for Health Policy/PCOR, Stanford University, Stanford, California. Dr. Studdert is Professor of Law and Medicine at Stanford University, Stanford, California.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of record](#). Please cite this article as [doi:10.1111/jels.12179](https://doi.org/10.1111/jels.12179).

ABSTRACT

There is wide agreement that existing approaches to valuing noneconomic losses from personal injury lack coherence. “Health utility” measurement—an approach developed in health economics for valuing health outcomes in public health and medicine—holds considerable promise for bringing greater rationality and consistency to assessments of injury-related noneconomic loss. However, the feasibility of creating utility measures that are suitable for use in personal injury compensation has not been demonstrated. This study takes that step. We surveyed over 4,100 members of the general public in Australia to assess people’s preferences for a variety of non-fatal “health states”. The health states were selected to reflect harms commonly seen in claims to compensation schemes for transport and workplace accidents. We then followed established methods for transforming the survey responses into a “severity weight” for each health state. We show how these severity weights can be used to define tiers in a schedule for guiding noneconomic damages determinations. We also discuss the strengths and limitations of the approach, and consider implementation challenges.

I. Introduction

We have come to accept almost without question the monetary evaluation of the immeasurable perturbations of the spirit. But why should the law measure in monetary terms a loss which has no monetary dimensions? . . . To put a monetary value on the unpleasant emotional characteristics of experience is to function without any intelligible guiding premise (Jaffe 1953:222).

Injury compensation schemes, casualty insurers, courts, lawyers, and legal scholars have long struggled with problems of whether and how to compensate noneconomic loss arising from personal injury. Opposition to the very existence of noneconomic damages comes from an unlikely coalition. Scholars on the left object to the baseness of commodifying human pain and suffering (Abel 1986; 2005). Law-and-economics scholars argue that the availability of general damages compels potential victims of injury to purchase a form of insurance for which there is negligible demand in the marketplace (Danzon 1984; Schwartz 1988; Croley & Hanson 1995). Some medical and public health researchers have posited that the availability of compensation for noneconomic loss impedes recovery from injury (Grant & Studdert 2009). Though staunch, such opposition has had little traction outside academia. Throughout Anglo-American legal systems, adjudications of injury claims make provision for the noneconomic consequences of harm.

How is it done? The approaches used to quantify compensation for noneconomic loss vary considerably, within and across legal systems, making generalizations difficult. With that caveat, some modal characterizations are reasonable. In personal injury litigation, the dominant approach is largely unfettered decision-making by judge or jury. Depending on the jurisdiction, courts may impose statutory damages caps to constrain high-end awards, but this is generally done *ex post*.

In administrative compensation systems—most notably, the schemes established to compensate workplace and automobile injury—lump-sum awards often accompany recompense for lost wages and medical expenses. Although the purpose of these awards is not always clearly articulated, award levels are generally indexed to the physical severity of the harm and understood to cover its non-pecuniary dimensions. For the task of indexing the severity of permanent injuries, workers' compensation and no-fault automobile schemes usually turn to the judgment of physicians, who often apply the American Medical Association's Guides to the Evaluation of Permanent Impairment (or some local adaptation of it). Over 45 years and six editions, the "Guides" has become the *de facto* standard for

such assessments (Seabury et al. 2013).¹ The Guides is essentially a patient evaluation protocol; it takes physicians through a structured decision-making process that leads to clinical ratings of a claimant's degree of impairment.² Schemes then use those ratings to determine levels of compensation for injury-related loss, including its noneconomic components.

Both court-based and administrative approaches lack coherence (Geistfeld 1995; Studdert et al. 2011). Judges and juries find noneconomic damages determinations vexing (Diamond 1993), a struggle that is borne out by wide variability in awards for injuries that appear to involve comparable levels of pain and suffering (Leebron 1989; Blumstein 2005; Abel 2006; Sanders 2008). Problems with horizontal equity have been well-documented (Bovbjerg et al. 1989; Wissler et al. 1999).³ Damages caps tend to reduce variability and improve horizontal equity, but do so at the expense of vertical equity and overall fairness (Studdert et al. 2004; Hyman et al. 2009; Chamallas and Wriggins 2010).

In administrative compensation systems, the Guides promises greater consistency of decision making, but falls short on several fronts. Concerns surround the instrument's lack of scientific foundation and the validity and reliability of the impairment ratings it produces (Spieler et al. 2000; Bonner 2008; Forst et al. 2010). The Guides blurs the distinction between "impairment" and "disability" (Bonner 2008; Holmes 2014).⁴ The instrument is incomplete and uneven, in the sense that some states of impairment are not covered, and certain activities, functions, and tasks are emphasized over others without justification. Furthermore, the fundamental presumption of a linear relationship between the kind of

¹ In the United States, nearly all states and several federal employee compensation systems either mandate or recommend use of the Guides to determine the amount and duration of workers' compensation permanent partial disability (PPD) benefits (American Medical Association 2008).

² The Guides defines "impairment" as "a significant deviation, loss, or loss of use of any body structure or body function in an individual with a health condition, disorder, or disease." A "permanent" impairment is defined as one that has reached maximum medical improvement and is well stabilized and unlikely to change substantially in the next year with or without medical treatment. "Disability" is defined as "activity limitations and/or participation restrictions in an individual with a health condition, disorder, or disease" (American Medical Association 2008).

³ Here "horizontal equity" refers to the principle that injuries of similar severity should receive similar levels of damages. "Vertical equity" refers to the principle that levels of damages should climb with injury severity.

⁴ The Guides appears to have a bet each way here. On the one hand, the preamble states that the system of impairment ratings "is not intended to be used for direct estimates of work participation restrictions." On the other hand, the Guides directs physicians to rate impairment based on the degree of impact it has on the injured person's ability to perform activities of daily living; it also claims to be "a physician-driven first approximation of a process that attempts to link impairment with a quantitative estimate of functional losses". Given this ambiguity, it is not surprising that many compensation schemes have elected to use Guides ratings as a direct estimation of the claimant's degree of disability (Holmes 2014).

functional limitations the Guides rates and the extent of a claimant's noneconomic loss is highly questionable (Studdert et al. 2011). All of these issues cast doubt on the appropriateness of relying on the Guides to grade claimants' losses (Spieler et al. 2000).

How can such incoherence so thoroughly pervade noneconomic damages assessment? And if existing approaches are so fundamentally flawed, why have they not been fixed or abandoned for better ones? The short if largely unspoken answer appears to be an entrenched view that nothing better is possible. If human suffering is essentially unknowable in any objective sense, one approach is surely as good as any another for quantifying what Jaffe called "immeasurable perturbations of the spirit" (Jaffe 1953).

This view ignores fifty years of theoretical and empirical work in developing methodologies for valuing health outcomes in public health and medicine. In an earlier paper, we reviewed a range of possible approaches for bringing greater coherence to assessments of noneconomic loss (Studdert et al. 2011).⁵ We argued that the most promising of them is a schedule-based approach in which tiers of severity are set by reference to "health-utility" measures. We are not alone in recognizing the potential for such measures to bring unprecedented rigor to the assessment of noneconomic damages. Several others have made the connection (Miller 2000; Ubel and Loewenstein 2008; Noah 2009; Avraham, forthcoming). Karapanou and Visscher have explored the idea in some depth (Karapanou and Visscher, 2010; Visscher and Karapanou, 2015; Karapanou and Visscher, 2014; Karapanou 2014). On the whole, this emerging body of scholarship is optimistic about potential to turn health-utility measures to this use, but the technical feasibility has not been empirically demonstrated to date. This study attempts to take that step.

We conducted a web-based survey of over 4,100 members of the general public in Australia to assess people's preferences for a variety of non-fatal "health states". The health states were selected to reflect conditions regularly observed as the basis of claims to transport accident and workers' compensation schemes. We then followed established methods for analysing the survey responses to produce a weight indicative of the severity of each health state. This paper argues that these "severity weights" are amenable for use as the basic building blocks of a schedule for guiding noneconomic damages determinations.

⁵ In a forthcoming book chapter, Ronen Avraham undertakes a similar review, and the options he discusses largely overlap with those we identified (Avraham, forthcoming).

The health utility-based approach we describe produces measures of loss severity with a variety of attractive qualities for damages assessment. Empirical rigor and internal consistency are achieved without abandoning attentiveness to subjective preferences. However, the approach is not a panacea. It does not provide monetary values for the losses; that must be done outside the framework we describe. And a schedule built from such measures would confront a raft of thorny implementation issues in civil justice systems. Nonetheless, we conclude that the approach developed and showcased in this paper is capable of taking a large stride toward coherent valuations of noneconomic loss.

The article proceeds as follows. Section II provides a brief description of the health-utility approach and considers theoretical aspects of their suitability for assessment of noneconomic damages. We also describe relevant methods from the 2010 Global Burden of Disease project, on which our data collection and analytical methods are based. Section III describes the study approach and Section IV presents our main findings. Section V discusses the findings and notes some limitations. Section VI reviews additional steps needed to create an implementable schedule. Section VII concludes.

II. Background

A. The Health-Utility Approach

Health-utility assessment involves the assignment of numerical weights to health outcomes for purposes of indicating their relative severity (Brazier et al. 2007; Drummond et al. 2005). The weights are typically expressed on a zero-to-one scale, reflecting a continuum of outcomes between perfect health and a state equivalent to being dead. The approach is used extensively in health economics and burden of disease assessments, and has been applied to many public health contexts. Its key strength is to produce summary health measures that enable comparison of a wide range of outcomes on a common scale.

Health-utility assessment is best understood as a family of techniques, rather than a single unified method. The approach several of us outlined in our earlier paper (Studdert et al. 2011), and which we pursue here, involves eliciting the preferences of a designated population for various injury-related health states, and then averaging those preferences to derive severity weights for each state. A severity weight reflects the seriousness of the injury's consequences, independent of purely economic effects.

The focus on health impacts and the exclusion of economic losses render utility-based severity weights a loose approximation of the legal concept of “pain and suffering”. To tighten the approximation for this study, we adapted the conventional preference elicitation methods in several ways, which are described below.⁶

B. Global Burden of Disease 2010, Injuries, and Health States

The 2010 Global Burden of Disease project (hereafter “GBD 2010”) was one of the largest studies ever undertaken in epidemiology and public health. It involved a decade-long collaboration between hundreds of scientists from over 300 institutions in 50 countries (Horton 2012). The study aimed to systematically quantify the magnitude of short- and long-term losses of health attributable to 291 different types of diseases and injuries (Murray et al. 2012; Salomon et al. 2012; Vos et al. 2012). From this set of 291 diseases and injuries, the GBD 2010 investigators identified a total of 1,160 sequelae—essentially all of the important health outcomes that may arise as consequences of these diseases and injuries. The 1,160 sequelae were then distilled into 220 unique “health states”. The health states were a parsimonious list of consequences of the diseases and injuries, formulated to capture key differences in symptoms and functioning associated with the sequelae (Salomon et al. 2012).

Our study used preference elicitation methods and stylized health states drawn directly from the GBD 2010 Disability Weights Measurement Study (Salomon et al. 2012). However, our focus differed from that study in two key respects. We limited the survey to a subset of the GBD states that reflect types of harm that commonly form the basis of injury compensation claims. Also, rather than asking participants to evaluate the severity of alternative conditions in terms of “health loss”, as the GBD 2010 had done, we were interested in negative non-financial impacts on “quality of life”—a construct that is notionally very close to the nature of the loss that noneconomic damages are intended to compensate.

III. Methods

A. Selection of Health States

We selected a subset of the 220 health states that had been developed as part of the GBD 2010. Three considerations guided our choices. First, we sought health states that resulted

⁶ See, in particular, subsections of Section III entitled, “Selection of Health States” and “Questionnaire Design and the ‘Quality of Life’ Construct”.

from injuries—specifically, the types of injuries for which compensation is frequently sought in compensation schemes for transport accidents and workplace injuries. Second, we restricted our choices to permanent health states. Finally, we aimed to include health states that spanned a wide severity spectrum. We elaborate on each consideration below.

The identification of health states commonly encountered in administrative compensation schemes was informed by the experience of three investigators: two (GG, DS) had experience in litigating and studying claims in these settings and one (DF) had experience in assessing and treating scheme claimants. We also reviewed detailed descriptions of the top 20 most common injuries that formed the basis of claims lodged over a 10-year period (2002-11) with the Transport Accident Commission and WorkSafe Victoria—respectively, the transport and workers' compensation schemes for the state of Victoria, Australia. Insights from both sources helped to ensure that the health states selected for inclusion covered a number of the most common types of non-fatal injuries adjudicated in these administrative compensation schemes.

We focused on health states that entailed permanent symptoms and functional limitations, conditions that were expected to endure in a more-or-less steady state over the rest of a person's lifetime. Although the GBD 2010 included both chronic and temporary conditions, we excluded the latter for practical reasons: Australian courts and statutory compensation systems generally do not provide noneconomic damages or lump sum payments for temporary conditions (Butler 2005; Safe Work Australia 2016), which diminished their relevance to our project.⁷ A related dimension of permanence, to which we return in describing the survey, is that participants were asked to consider the post-acute health state, as opposed to the symptoms and functional limitations that occur in the injury's immediate aftermath when the patient is recovering and possibly hospitalized. In other

⁷ The methods used to estimate severity weights for temporary and permanent injuries are similar. In both cases the magnitude of the loss is estimated for a single year and then multiplied by the expected duration of the condition, with the resulting summary measure of health typically expressed in units of quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs) (Gold et al. 1996). The key difference is that durations for permanent health states are bounded by life expectancy, whereas durations for temporary health states are bounded by a combination of remission and life expectancy. Since tort systems *do* compensate temporary injuries, future research might usefully extend exploration of the feasibility of a health utility-based damages schedule by assessing their severity by means of QALYs or DALYs. A Dutch team has recently described how this might be done (Karapanou and Visscher, 2010; Visscher and Karapanou, 2015).

words, the focus was on symptoms and limitations at a time when the injured person's health status had stabilized and recovery was as complete as it ever will be.⁸

The robustness of severity weights depends in part on obtaining a wide “spread” of health states across the severity spectrum, from the most minor conditions to the most debilitating ones (short of death). For our purposes, another advantage of a scale that covers the severity spectrum is that it provides the basis for a damages schedule capable of guiding compensation decisions on all types of injury.

Based on these criteria, we selected 47 of the 220 health states from the GBD 2010 (Table 1). We added one health state that had been developed as part of the European Disability Weights Study⁹ (Haagsma et al. 2015); this health state captured a permanent symptom (mild low back pain) that commonly results from workplace and transport injuries but was not included in the GBD 2010 set. Thus, our final analytical set consisted of 48 health states.

B. Questionnaire Design and the “Quality of Life” Construct

The survey questionnaire was designed around a series of pairwise comparisons. Each question described two hypothetical people, each of whom was living with one of the 48 health states. Participants were then asked to identify which person they thought had a higher quality of life.

The narrative descriptions of the health states largely followed those developed in the disability weights measurement component of the GBD 2010 (Salomon et al. 2012). The GBD 2010 investigators formulated the descriptions through an iterative consultation process with expert working groups; the wording aimed to capture the most salient symptoms and functional consequences of each condition in a manner that could be clearly and consistently communicated to laypersons (Salomon et al. 2012). The descriptions used simple, non-clinical language. Most described functional limitations and at least one consequence of the condition for the individual's activities of daily living. For 8 of the 48 health states we deviated slightly from the GBD 2010 descriptions, following modifications to the descriptions made as part of the European Disability Weights Study (Haagsma et al.

⁸ According to Safe Work Australia 2016, permanent impairment occurs at the point of maximum medical improvement, when the impairment has become stable, or is not likely to improve despite medical treatment (Safe Work Australia 2016).

⁹ The European Disability Weights Study involved a partial replication of the GBD 2010, with the goal of producing a set of disability weights for Europe. It estimated disability weights for a set of 255 health states, by eliciting preferences of representative population samples in four European countries.

2015).¹⁰ The language of the health state descriptions used in our questionnaire, together with their sources, is provided in the **Appendix**.

While the health state descriptions we used closely resembled those used in the GBD 2010, the question stems did not. As noted earlier, the GBD 2010 focused explicitly on the concept “health loss” (Salomon et al. 2012), whereas our study objectives called for a broader construct.

There is no consensus on the precise nature of the losses noneconomic damages are intended to compensate. Courts have described the scope and content of these losses in many different ways (Avraham, forthcoming). In most formulations, adverse effects on mental and physical health are central dimensions of the harm, but they do not occupy the entire construct. Injury afflicts other aspects of well-being, such as relationships, social life, and sexual and leisure activities. To be sure, those dimensions of the loss are not unrelated to health; they may even mediate (or be mediated by) health losses. Nonetheless, the point is that the multi-dimensional nature of life decrements caused by serious injury include reductions in health status but extend further.

The broader concept of welfare loss captures more completely the multifactorial nature of noneconomic harm in law. However, “welfare loss” is not a familiar term to most laypersons. We opted for “quality of life”, which comes as close as any term we could think of to capturing the construct of interest. Thus, the stem of all of the pairwise comparison questions asked participants to choose which of the two people described had a “better overall quality of life”.

Participants were given several pointers and instructions for answering the pairwise comparison questions, which were reiterated throughout the questionnaire. The goal was to ensure the severity weights derived from participants’ responses were aligned as closely as possible with the legal construct of noneconomic loss. One pointer encouraged participants to consider the full range of quality-of-life impacts, including effects on “everyday activities, relationships, and social and recreational activities”. We also sought to remove time and economic impacts from participants’ deliberations.

¹⁰ The European Disability Weights Study used 172 of the original GBD 2010 health states, modified an additional 33 of the GBD 2010 health states, and added 43 new health states (Haagsma et al. 2015). The language used in 8 of the modifications suited our study because it provided additional detail on the condition or used language that was more compatible with assessment in the context of a compensation system.

With respect to time, survey instructions explained that both of the health states described in each pair were chronic conditions that would persist over a lifetime, and the two people had the same number of years left to live. With respect to economic impacts, participants were asked—and reminded repeatedly—not to consider any of the financial consequences of the conditions when thinking about quality-of-life effects. Lost income and medical expenses were mentioned as examples of financial consequences to be excluded from consideration (**Appendix**). Because economic losses are usually quantified and compensated as a separate head of damages, exclusion of them from the participants' calculus should, in theory at least, produce a “purer” estimate of noneconomic loss.¹¹

Although the pairwise comparison questions were the centrepiece of the questionnaire, it also elicited other information. After completing their last pairwise comparison question, participants were asked to indicate how they had thought about which person had a better quality of life. The options given were: “put yourself in the shoes of each person”, “think of people you know with similar or related conditions”, “consider two imaginary people”, and “other”. We also collected some demographic information (e.g. education level), but the need to do so was limited by the fact that the survey research company we employed to administer the survey maintained a stock of demographic variables on all participants and provided us with this data.

C. Survey Sample

There has been debate in the decision science literature over *whose* preferences ought to be counted in valuing health states. A key pivot point in this debate is whether it is more appropriate to prioritize the views of the general public or individuals with direct experience of the conditions being valued. The latter constituency divides further into those with clinical expertise, such as physicians who treat many patients with the condition, and those with lived experience—patients, their families, and lay care-givers. Growing evidence that

¹¹ It is reasonable to question how well participants were able to make this separation. Interestingly, the same question has arisen in the general health utility literature, where (for other reasons) preference elicitation studies also seek to eliminate economic considerations, particularly when financial consequences are borne by persons other than the patient (Gold 1996; Drummond 2005). To the extent there is a strong positive linear relationship between levels of noneconomic loss and economic loss due to personal injury, the risk that severity ratings for the former would be polluted by consideration of the latter is attenuated. Decision makers in compensation systems often assume a strong correlation—evidenced, for example, by setting noneconomic damages as a proportion or multiplier of economic losses. However, there are clearly types of injuries for which the correlation is weak or non-existent. For example, consider scarring or disfigurement and mild-to-moderate sensory damage, both of which are conditions that tend to result in low economic losses but relatively high noneconomic losses.

patients characterize conditions they have experienced as less severe than do people who have not experienced those conditions (a phenomenon sometimes referred to as “hedonic adaptation”) has rekindled interest in the advantages and disadvantages of focusing on patients’ lived experience (Loewenstein and Ubel 2008). However, the predominant approach for estimating health-related quality of life values has been to turn to members of the general public (Gold et al. 1996;), especially when the assessments are destined to inform public policy (Gold et al. 1996; Hadorn 1991; Dolan 1996; Dolan et al. 2003; Neumann et al. 2016). We took this approach. Other considerations counselled in favour of this choice; we describe them in Section V.

Our sample was drawn from OzPanel, a large standing panel of survey participants maintained by Roy Morgan Research. OzPanel consists of Australian residents, recruited mainly through a national random door-to-door survey, who have agreed to be contacted to participate in online research surveys. Participation earns panel members “points”, which can later be converted to shopping vouchers or charitable donations. The panel is designed to have demographic characteristics representative of the Australian population.

The sample consisted of a sample of OzPanel members 18 years or older. Sample selection was based on non-interlocking quotas corresponding to several characteristics (age group, sex, and state/territory of residence) that were intended to match the demographic characteristics of the adult Australian population. Prior research has demonstrated that the use of even a small number of such quotas in internet survey samples can help ensure that the samples are representative of the general public on a much wider range of demographic characteristics (Bambrick et al. 2009).

D. Survey Pre-testing

To ensure the survey questions were clearly worded and would be well understood by participants, we conducted pre-testing of the questionnaire in two focus group sessions involving a total of 7 participants, as well as cognitive interviews with two people, in April 2013. The wording of several questions was modified based on feedback from these discussions. These sessions helped to confirm that participants understood the term “quality of life” and felt comfortable focusing on the non-financial impacts of the health states.

E. Survey Administration

Roy Morgan administered the questionnaire online in August and September 2013. Emails were sent to OzPanel members in the sample inviting them to participate in the survey. The email invitations were sent in sequential batches, with later batches focused on subgroups whose response rates were lower in earlier batches than was required to ensure national representativeness for the sample as a whole.

Participants were randomly assigned one of three different versions of the questionnaire.¹² Version 1 posed 15 pairwise comparison questions. Twenty percent of the participants who were assigned this version received a 16th pairwise comparison question that asked about the same health states as the first question they answered; this was done to allow evaluation of test-retest reliability. Versions 2 and 3 both posed 8 paired comparison questions, without repeated pairs for assessing reliability.¹³ In total, 45% of participants received version 1 of the questionnaire, 30% received version 2, and 25% received version 3.

The two health states that appeared in pairwise comparison questions were randomly selected from among the universe of 2,256 (48 x 47) possible pairs. When participants attempted to advance in the questionnaire after leaving a question unanswered, they were given up to two reminders to complete the missing response before being permitted to skip forward.

The overall response rate among all OzPanel members who received an initial invitation to participate in the survey was 45%. However, the method of sampling and recruitment we used does not permit calculation of a standard response rate, because of the variability in recruitment and follow-up strategies across subgroups (Bambrick et al. 2009). We began with firm targets for each questionnaire, aiming for a broadly representative sample on age, sex and state. Some target numbers were achieved without the need for reminders. Up to two reminders were sent to members in groups whose response levels were lower than required for representativeness. Overall, 30% of those who received an initial invitation to participate in the survey were sent one reminder and 27% were sent two reminders.

¹² The survey also included a fourth version that is not discussed here because it involved alternative measurement approaches that are not relevant to the current analysis.

¹³ The rest of survey versions 2 and 3 posed a series of questions asking participants to compare health states using a variety of different preference-elicitation techniques. Those questions are not relevant to this study and responses to them are not reported here.

F. Analysis

Our analytical method is very similar to the one used in the GBD 2010. Participants' responses to the pairwise comparison questions were analysed using a probit regression model. This approach is based on the conceptual framework developed by Thurstone and others (Thurstone 1927; Bradley 1976), which theorizes that comparisons of similar outcomes are likely to produce greater disagreement in the ordering of these outcomes than are comparisons of dissimilar outcomes (Salomon et al. 2012). Intuitively, if a large sample of people are comparing two states that are indistinguishable in terms of their severity, we expect that the responses will be evenly divided between the two states; on the other hand, we expect that comparisons of two states that are quite different will result in a relatively high proportion of respondents choosing one state over the other. The analytic model formalizes this intuition to infer 'distances' between states on a measurement scale based on the choice probabilities observed in the sample.

Within the probit regression model, choice responses were modelled as a binary response variable Y , where $Y = 1$ indicates that the first health state in a given pairwise comparison was chosen as the preferred option (Salomon et al. 2012). We included indicator variables (I_i) for each health state, which were specified as follows:

$$I_i = \begin{cases} 1 & \text{for the first health state in a given pairwise comparison} \\ -1 & \text{for the second health state the pairwise comparison} \\ 0 & \text{for all health states other than the pair being considered} \end{cases}$$

The probit regression coefficients, θ_i , were estimated by maximum likelihood. Under this specification and the assumption of constant variance, the difference in regression coefficients between two health states ($\theta_1 - \theta_2$) is linearly related to the difference in the mean quality of life value of these states. The resulting probit coefficients capture the relative differences in the levels of quality of life between health states, but on an arbitrary scale that must be transformed for scale and level to obtain values on a 0-to-1 scale.

Given our interest in using severity weights as the basis of a noneconomic damages schedule, we conducted a linear rescaling of the probit regression coefficients by anchoring coefficients on a 0-to-1 scale with endpoints identified *a priori*. The endpoints were set by reference to rules operating in the major administrative compensation schemes in Victoria

and New South Wales, Australia's two most populous states.¹⁴ Specifically, we used mild fracture as the lower-bound endpoint (i.e. scaled to 0); this injury resides at the borderline of eligibility for compensation, given the statutorily-prescribed minimum severity thresholds. A severe spinal cord lesion at the neck level was the upper-bound endpoint (i.e. scaled to 1); this injury would qualify for the maximum compensation allowed for noneconomic damages. Thus, our severity weights were scaled to span a scale between a lower-bound of no compensation for noneconomic loss and upper-bound of maximum compensation, and these bounds corresponded to the real-world compensatory status of the endpoint health states.

We estimated 95% uncertainty intervals by repeating the analysis on 1,000 bootstrapped samples from the full analytic dataset.

We assessed measurement error and the internal consistency of participants' ratings in several different ways. The responses to pairwise comparisons of health states were summarised with a heat map, which visually displays the aggregate choice probabilities for all pairs of health states (Salomon et al. 2012). In other words, for each pair we calculated the probability that the individual with the first health state was chosen as having the better overall quality of life, based on the responses of all participants who rated this pair.

We calculated intra-rater reliability using responses from the subset of participants who had been presented twice with the same pairwise comparison. We report intra-rater reliability in terms of percentage agreement and kappa scores (Landis and Koch 1977). Finally, we explored the variability of our estimates across different demographic groups by conducting stratified analyses according to various observed characteristics of participants and comparing the results of these analyses.

The statistical analyses were conducted using Stata (version 12) and R (version 2.15.2).

G. Comparison with AMA Guides Permanent Impairment Ratings

A final step in our analysis was to compare the severity weights with Guides ratings of the same health states.

¹⁴ See *Wrongs Act 1958* (Vic) s 28LB; *Workplace Injury Rehabilitation and Compensation Act 2013* (Vic) ss 325, 327; *Transport Accident Act 1987* (Vic) s 93; *Civil Liability Act 2002* (NSW) ss 16-17A; *Workers Compensation Act 1987* (NSW) s 66; *Motor Accidents Compensation Act 1999* (NSW) s 131.

As noted earlier, the Guides is routinely used for determining levels of permanent partial disability and impairment and, in many administrative compensation schemes, those determinations are influential in calculating benefits.¹⁵ The Guides calls for a physical examination of the injured patient by a suitably-qualified physician. The physician is led through a series of ratings on different parts and regions of the body. An algorithm then combines results from the regional assessments to produce a “whole person impairment” rating on a scale from 0% to 100%.

We could not apply the Guides in the conventional way. We did not have live patients for physical examination. Moreover, the health state descriptions were too general to support all of the judgments required to produce a whole-person impairment rating under the Guides.

Instead, one member of our team (DF), an occupational physician with extensive clinical experience in assessing and treating claimants and a regular user of the Guides, estimated ratings. He considered a “typical” clinical condition associated with each health state, and applied the 4th edition of the Guides (American Medical Association 1993) to rate that condition. Typicality related to both type and severity of the condition an “average” patient with the health state would be likely to present with. Uncertainties in the assessment process were handled by producing plausible ranges for the ratings wherever this was warranted.

Further details of the rating procedures are provided in the Appendix.

H. Ethics Review

The study was approved by the Human Research Ethics Committee at the University of Melbourne.

¹⁵ Schemes differ in the extent to which benefits calculated in this way are explicitly tagged as compensation for noneconomic loss. Workers' compensation schemes in Australia and Canada, which rely heavily on the Guides to determine compensation amounts, tend to be more explicit in linking components of the benefits to noneconomic loss (Safe Work Australia 2014; Association of Workers' Compensation Boards of Canada 2015). By contrast, although many US workers' compensation schemes mandate use of the Guides to assess impairment, and those assessments are integral to benefits calculations, the precise nature of the loss being addressed may not be explicitly stated (Barth 2004).

IV. Results

A total of 4,129 participants completed the survey. We dropped the responses of 158 participants who began the survey but did not complete and submit it. Among submitted responses, very few questions were skipped, resulting in only a small amount of missing data (<0.01% of assigned fields).

The analytic sample consisted of 47,912 ratings in total, an average of 1,934 (range 1,830-2,095) ratings per health state.

A. Sample Characteristics

Study participants closely resembled the general Australian population in relation to state of residence and sex (Table 2). Participants aged younger than 30 years and older than 70 years were slightly underrepresented, as were participants with lower levels of education and participants born outside Australia. Additionally, participants reported household incomes that were more likely to fall in the lowest income category (<\$400/week) and less likely to fall in the highest income category ($\geq$ \$2,000/week), compared with adults in the general population.

Ten percent of respondents reported working or volunteering in a health care setting.

B. Assessment Approaches

In describing how they “thought about the two health problems” presented in the pairwise comparison questions, three quarters of participants indicated that they put themselves in the shoes of the hypothetical persons described. Approximately one third of respondents reported thinking of people they knew with similar conditions, and one third said they considered “two imaginary people”.

C. Severity Weights

Figure 1 shows estimated severity weights and 95% uncertainty intervals for the 48 health states examined. Several of the weights are labelled for illustrative purposes. A complete list of weights and 95% uncertainty intervals for each health state is provided in the Appendix.

The health states were arrayed fairly smoothly between the minimum and maximum anchor states. Severe traumatic brain injury and various types of spinal cord lesion attracted the highest severity weights. Severe and moderate depressive disorders (0.88 and 0.72,

respectively) and severe anxiety disorders (0.76) also ranked highly, as did conditions associated with severe chronic pain (0.66-0.68). At the other end of the spectrum, mild hearing loss was the only health state with a severity weight below 0.20 (other than mild fractures, the minimum anchor state).

There were no patent inconsistencies in rankings of health states, in the sense that no severe version of a particular health state had a lower severity weight than the corresponding less severe version of the same state.

D. Data Quality

Figure 2 shows, in heat map form, response probabilities for the pairwise comparison questions. It provides a visual indication of the degree of agreement among respondents with respect to all possible combinations of pairwise comparisons. The amount of measurement error in the survey is reflected in the degree to which response probabilities follow an orderly transition from high to low between the upper left and the lower right corners of the heat map. A heat map with no measurement error and perfect internal consistency would have a smooth colour transition from blue to red along the diagonal, while one with 100% error would have a completely random assortment of coloured squares (Salomon et al. 2012). The shading pattern shown suggests a high level of internal consistency and a modest amount of measurement error.

E. Intra-rater Reliability

There was a high degree of consistency in the responses provided by the subset of participants (n=364) who completed the same pairwise comparison question twice. Intra-rater agreement was 80% in these questions and the kappa score was 0.599 (95% CI 0.497 - 0.702; $p < 0.0001$), indicating moderately good agreement after adjusting for agreement by chance.

Further analyses showed that when intra-rater discrepancies occurred they did so primarily among health states that were close in severity, according to the estimates derived from the totality of responses to these questions. Results of these analyses are presented in the Appendix.

F. Variability in ratings by sociodemographic characteristics

The ratings were extremely robust across different sociodemographic subgroups. For example, regression analyses based on responses from participants from high income

households produced severity weights that were very highly correlated ($r = 0.997$) with those produced by analyses of participants from low income households. The same was true for participants with high and low levels of educational attainment ($r = 0.996$) and participants who did and did not work or volunteer in health care settings ($r = 0.991$). The trivial amount of preference variability we observed across subgroups is consistent with findings from the GBD 2010, which found strong evidence of highly consistent disability assessment across samples with different cultural environments, educational attainment and socioeconomic status (Salomon et al. 2012).

G. Comparison of Severity Weights with AMA Guides Ratings

It proved feasible to assign a whole-person impairment rating under the Guides to 30 of the 48 health states. Seventeen health states could not be rated for the following reasons: the description of the health state was not complete enough to support a defensible rating ($n=8$); the health state was a mental health condition, for which Guides' ratings are not preferred in practice ($n=6$)¹⁶; or the Guides ratings clearly did not deal with crucial aspects of the condition ($n=3$).¹⁷ One additional health state could be rated, but was assigned a range that was too wide to be meaningful (max-min of >60 percentage points), so we excluded it from further analysis.

Figure 3 compares Guides ratings of the 30 health states that could be rated with the severity weights we estimated for those states. The severity weights are shown in blue and presented with 95% uncertainty intervals; the Guides ratings are shown in green and the bars indicate the rater's determination of plausible upper and lower bounds. To assist interpretation, the comparisons are arranged into six general types of conditions: (a) spinal cord lesions; (b) amputations; (c) fractures; (d) musculoskeletal problems; (e) hearing loss; and (f) neck and back pain.

The severity weight was higher than the Guides rating for 23 of the 30 health states, about the same for six health states, and lower for only one health state (amputation of an upper limb). The size of the differences between severity weights and Guides ratings tended to be

¹⁶ The Guides provides a method for rating mental health conditions, but many compensation schemes substitute alternative approaches (see, e.g., Safe Work Australia 2014). Therefore, the practical significance of Guides-based evaluations of these conditions was questionable.

¹⁷ All three of these health states involved disfigurement. The Guides' approach to rating this type of condition focuses on the effect it has on a person's activities of daily living. In our view, this did not permit a meaningful rating to be made under the Guides.

relatively small for spinal cord lesions, amputations, and hearing loss, and large for pain, fractures, and musculoskeletal problems.

V. Discussion

This study advances development of an approach with considerable potential to improve the rationality and coherence of assessments of compensation for noneconomic loss from injury in civil justice systems. In this Section, we discuss the study's contributions to new knowledge in three areas: (1) measurement of health utilities; (2) assessment of noneconomic loss; and (3) identification of problems with existing approaches to damages assessment. We also discuss some limitations of the study.

A. Utility Measurement

The health economics literature is replete with different types of interval scaled weights for grading the severity of health conditions. Utility weights and health-related quality-of-life weights, based on interviews eliciting people's preferences, are frequently used to evaluate the cost-effectiveness of interventions to treat or prevent disease. Disability weights, such as those developed in the of Global Burden of Disease series of studies (1990, 2000, 2010), have emerged as an influential method for quantifying the contributions of particular diseases and injuries to the overall burden of illness, globally and within countries.

Our study had a different goal: to demonstrate the development of a set of utility weights that could be used as the building blocks of a decision support tool for guiding noneconomic damages determinations. That goal led us to modify the conventional approach to health utility measurement in two main ways. First, rather than measuring benefits or decrements in terms of "health-related quality of life", "health", or "health loss", we sought to measure effects on the broader outcome of "overall quality of life", net of financial considerations. This brought us closer to the nature of the loss noneconomic damages are intended to compensate. Second, the severity weights we developed pertained to a set of health conditions often seen in injury compensation claims. To the best of our knowledge, no previous study has done either of these things.¹⁸

Our findings suggest that none of the tweaks we made to conventional methods of developing utility-based severity weights compromised the quality of the resultant

¹⁸ In fact, only a few studies have sought to develop severity weights targeted specifically at different types of injuries. See, e.g., Haagsma et al, 2009.

measures. Feedback during pre-testing of the survey questionnaire indicated that laypersons were comfortable comprehending brief written descriptions of injury-related health conditions and then opining on the relative negative impact of those conditions on quality of life. The survey response data were also encouraging: there was good consistency in the choices made, both between respondents and within a single respondent's answers; there was little systematic variation in these judgments by demographic subgroups; and there was no evidence that the technique was less well suited to measurement among respondents with lower levels of educational attainment.

B. Noneconomic Loss Assessment

Several features of health-utility measurement approach make it well-suited to quantifying noneconomic loss. One is its focus on noneconomic dimensions of the loss. Conventional preference elicitation techniques seek to eliminate such dimension from an individual's deliberations. This aligns with the goal of isolating noneconomic loss from economic loss in determinations of compensatory damages, in order to avoid imprecision attributable to double-counting or misallocation (Sharkey, 2005).

The approach we have outlined also allows a broad range of health states to be scaled. In theory, there is no limit to the number of conditions one could incorporate, although the costs associated with running a sufficiently large number of conditions through preference elicitation interviews to achieve stable estimates is likely to be prohibitive. A more typical strategy in the public health realm is to scale a limited number of conditions, as we did, and then rely on a separate qualitative judgment process to slot other health states (i.e. conditions that do not directly match those scaled) between appropriate neighbors.

Another desirable feature of the health-utility scale is that both the order and the cardinal values assigned to states along the scale have meaning. Moreover, in our adaptation of the method, the severity weights actually correspond to the strength of preferences for different types of noneconomic loss. By contrast, other approaches for valuing noneconomic loss focus on an expedient proxy for that loss, such as economic loss, percentage impairment, or degree of physical harm or disability.

Perhaps the most attractive feature of a health-utility approach is its fusion of subjective and objective elements in deriving the severity weights. Jury and judge valuations of noneconomic loss are heavily subjective and highly variable. Determinations made by experts, such as clinicians, purport to be scientific and rigorous, and thus derive their

authority from claims to objectivity (tenuous though those claims may be in some respects). Health utility-based measures rest on a hybrid approach.

On the subjective side, the weights come from individuals' preferences. One of the main criticisms levelled at the Guides (aside from scientific veracity) has been its failure to account for the views of the wide range of stakeholders in compensation settings (Spieler et al. 2000; Forst et al. 2010). By grounding severity weights in the perceptions and experiences of the community, the health-utility approach preserves the democratic nature of decision making that many find so attractive in juries.

The choice between members of the general public and individuals with lived experience of the condition as a reference group for developing the measures is less clearcut. Evidence showing systematically different valuations of specific health conditions among individuals with and without those conditions has raised new questions for both health valuation methods and compensation systems (Sunstein 2007; Loewenstein and Ubel 2008). There are theoretical arguments for and against deriving utility weights for a given condition from the subjective preferences of individuals afflicted with it. But in our study, practical considerations dictated the general population as the preferable reference group. The health state narratives and instrument we used drew heavily on the GBD 2010. The narratives were designed for a general audience, and pairwise comparison questions are not a suitable format for eliciting preferences in a survey of persons with specific conditions.¹⁹ It is also worth noting that compensation approaches in the civil justice system do not currently account for biases introduced by hedonic adaptation. Attempt to tackle them at the same time we presented another fundamental changes to the noneconomic damages assessment may well have eroded the political acceptability of the reform.

In our study, there were additional reasons to prefer the perspective of laypersons. One of the main criticisms of the Guides has been its failure to account for the views of the wide range of stakeholders in compensation settings (Spieler et al. 2000; Forst et al. 2010). Grounding severity weights in the perceptions and experiences of the community preserves the democratic nature of decision making that many find so attractive in juries.

On the objective side, the procedures for eliciting preferences and calculating severity weights both occur within a rigorous empirical framework. The pairwise comparison

¹⁹ Since few respondents would have both conditions, the only feasible use of pairwise comparisons may be to develop a set of "hedonic" adjusters for weights derived from a general population.

technique draws on discrete choice models from economics. It allows preferences to be probed within a constrained and structured format—far from the kind of open-ended noneconomic damages determinations courts are called upon to make in individual cases. At the analysis stage, preferences are averaged across persons and scaled by reference to a small number of absolute measures (in our case, these were the lower and upper endpoints that were fixed *a priori*). This analytical approach is grounded in the theory of welfare economics, and some of its assumptions are controversial. For example, utility-based measures require aggregation and averaging of preferences across individuals, and some commentators have argued that subjective responses lose meaning when unmoored from unique characteristics of a particular person or condition (Hammitt 2002).

C. Severity Weights Compared

A natural next step is to examine how the severity levels we estimated compare with severity levels that would be assigned to the same or similar conditions under the variety of other approaches currently used to grade noneconomic loss. That inquiry goes beyond the scope of this study, although we made one foray: a limited comparison to ratings under the AMA Guides. Even without more detailed comparisons, it is possible to make some preliminary observations about the size of our estimates.

The severity weights we estimated for chronic and severe mental health conditions were remarkably high. Psychiatric and psychological harms have long assumed “second fiddle” status in tort law and accident compensation regimes. They are often subjected to onerous thresholds, minimized, or deemed completely ineligible for compensation (Chamallas and Wriggins 2010; DePianto 2012). Our results suggest that community attitudes regarding the impact serious mental health conditions have on quality of life are out of step with how these conditions are handled in many injury compensation systems. The GBD studies and other research have also highlighted the large and underappreciated population health burden associated with these conditions (Ngui et al. 2010).

The comparison of severity weights with Guides ratings showed some gaping differences that warrant further investigation. The severity weights were markedly higher for three-quarters of the health states examined. The largest gaps occurred among chronic forms of neck and low back pain. The Guides do not emphasize pain in most impairment ratings, focusing instead on ‘objective’ indicators of injury, and dealing with pain in a separate Chapter (the use of which is legislatively excluded by many compensation systems) (Robinson et al. 2004; Safe Work Australia, 2016). The Guides’ approach to pain has been

the subject of controversy and criticism (Schultz and Chlebak, 2014; Robinson et al. 2004). The position adopted by many compensation schemes appears to be that the Guides is still capable of capturing the effects of pain through consideration of the underlying impairments that cause the pain. This is unconvincing. Our results deepen concerns that the Guides—and the way they are implemented in many compensation schemes—handle pain poorly.

The Guides inattention to effects of pain on health and well-being is symptomatic of a broader scepticism in compensation systems. Much like mental illness, the predominantly subjective nature of the pain-related symptoms has made them a target of reforms in Australia, Canada and the United Kingdom that seek to limit available compensation (see, e.g., Cameron et al. 2008; Cassidy et al. 2000; UK Ministry of Justice 2013). To be clear, these reforms may be driven less by perceptions regarding how much pain matters than by concerns about the evidentiary difficulties associated with proving its existence and magnitude. A health-utilities approach aids with the former but does not address the latter. There is growing speculation about the potential for neuroimaging techniques to be a game-changer on the evidentiary front for conditions like this, but penetration of such forms of evidence into compensation settings remains low (Pustilnik 2014; Greely 2015).

A large Canadian study of workers' compensation claimants in the 1990s examined the concordance between quality-of-life ratings from a sample of injured workers and Guides ratings of those workers' injuries. Findings from that study, like ours, showed a systematic under-rating by the Guides of some conditions (Sinclair and Burton 1994). However, the study's results were never published in the peer-review literature, apparently out of concern for the political fallout likely to have followed evidence suggesting systematic under-compensation in the Ontario workers' compensation scheme (Studdert et al. 2011).

It is specious, however, to automatically equate low ratings with under-compensation. The appropriateness of compensation for any given claim hinges on whether the amount of money paid is commensurate with the magnitude of the loss. As we discuss further below, neither Guides ratings nor our severity weights provide dollar amounts. Monetization is an additional step, about which both methodologies are silent. Consequently, it is more accurate to say that indexing noneconomic damages awards to Guides ratings may produce over- or under-compensation, depending on the injury under consideration and the outcome of the monetization step. What is evident from a comparison with the severity weights, however, is that the awards are inequitable relative to one another; in other

words, the Guides has a calibration problem. A better-calibrated scale, underpinned by health utility-based severity weights, would open the way to greater vertical and horizontal equity in awards.

D. Study Limitations

Limitations of our project fall into two categories: (1) limitations of the study design; and (2) feasibility issues with moving from the kind of severity weights we developed to a noneconomic damages schedule capable of guiding compensation decisions. We deal with the first category here and the second category in the next section.

One limitation of the study is that the generalizability of findings beyond our sample is not known. There are reasons to be optimistic. The sample was reasonably representative of the Australian population, and the results were robust across observable demographic and socioeconomic characteristics (although they may have varied according to individual or household characteristics that we did not observe). In addition, recent evidence from the GBD 2010 shows a high degree of consistency across countries and cultures in people's preferences for health states (Salomon et al. 2012).

Second, the health state descriptions survey participants rated were simple and quite brief. Although most included information regarding effects on activities of daily living, many lacked details of comorbidities that individuals living with these conditions may experience.²⁰ In valuing noneconomic losses in most compensation settings, adjudicators will have more information available to them regarding the nature and effects of a claimant's injury than survey participants did regarding the conditions they rated. The lack of detail in the health state descriptions may affect the precision of severity weights. It also creates an implementation challenge for a health-utility based schedule. Adjudicators will probably know more about the injuries they are seeking to position along the scale than they do about the reference conditions.

In theory, it would be possible to develop severity weights for more detailed set of health state descriptions, but there are some difficult trade-offs here. More detail won't

²⁰ This is a common limitation of studies eliciting preferences for health states, because of the sheer number of possible combinations of health states. There is no consensus on the most appropriate method to adjust for the presence of co-morbidities. The existing techniques (multiplicative, additive, or minimum methods) can produce very different estimates (Ara and Brazier 2012). The AMA Guides uses an additive formula in which the most severe condition is "fully" counted and proportions of the ratings for co-morbid conditions are added using a formula that weights these conditions according to their relative severity.

necessarily produce greater precision or ease implementation. The GBD 2010 investigators sought to strike a balance between, on the one hand, including enough detail to support meaningful valuations of the impact of the condition and, on the other hand, avoiding a level of detail that confused or fatigued participants (Salomon et al. 2012). Risks of the latter are real whenever preference elicitation methods turn to the views of laypersons and demand multiple comparisons. Moreover, scaling a highly specific set of conditions may complicate the adjudicators' task and invite error; such a schedule would require wide-ranging information on incident conditions to be compared to lengthy narratives of various reference conditions. Deciding what constitutes an optimal level of detail on the nature of the noneconomic loss—in both the reference conditions and the claims themselves—is an important question that warrants further investigation.

Third, the concept of “quality of life”, which was central to our valuation process, may not appropriately capture the harm noneconomic damages are supposed to compensate. It is difficult to refute this critique because of the rather inchoate nature of the noneconomic loss construct in personal injury law. It is a difficult critique to sustain for the same reason.

Fourth, a variety of factors may have led to systematic distortions in survey responses. Despite instructions not to, participants may have weighed financial consequences, or imagined the acute phase of the condition rather than its healed (or steady-state) phase. We do not consider adaptation effects. As discussed earlier, there is evidence that people who have lived for a period of time with a particular illness tend to rate it less severely than people with no direct experience of the condition (Ubel and Lowenstein, 2008; Karapanou 2011). This may have skewed responses to health states in our sample like anxiety and depression, which are relatively prevalent in the general population. It is debatable whether that skew is in the right or wrong direction.²¹

Fifth, the strength of health utility measures in calibrating losses rests in substantial part on the claim that they are capable of producing weights with both ordinal and cardinal meaning. This claim, in turn, rests on a couple of related technical assumptions. One is the assumption of constant variance that produces the “distances” between states on a measurement scale; the other is the key scaling assumption that these are linearly related to the magnitude of noneconomic loss. It may be argued that a linear re-scaling of our regression parameters to produce severity weights may not be appropriate, particularly if

²¹ In addition, it should be noted that this is a form of bias to which other methods of damages assessment (e.g. unconstrained determinations by judges and juries) are at least as vulnerable.

preferences for health states towards the ends of the severity spectrum do not have the same statistical properties as health states in the middle of the severity scale.

Sixth, our study produced severity weights only for permanent health states. This reflects the set of injuries that are eligible for non-economic payments in administrative compensation schemes in Australian. However, most tort systems allow non-economic damages for temporary injuries. Our approach could easily be extended to incorporate temporary health states.²² Finally, there is imprecision in ratings we made under the Guides, as evidenced by the wide ranges presented for some health states. Our inability to conduct physical examinations of injured patients with the underlying conditions, coupled with the generality and brevity of the health state descriptions, called for an unconventional rating procedure. Adherence to a more conventional approach may have produced different ratings, as may use of a panel of physician raters (as opposed to our one rater). None of those changes, however, are likely to alter the main findings from this part of the study—namely, Guides ratings frequently differ from the utility-based severity weights, they are lower in most instances, and they are especially low for conditions associated with pain, fractures, and musculoskeletal problems.²³

VI. Implementation and Feasibility

A collection of calibrated weights arrayed across a severity spectrum provides the empirical foundation for a decision aid capable of guiding noneconomic damages determinations, but it does not, of itself, create that aid. In this section, we review some further steps that are necessary to move from severity weights to a workable damages schedule. We also consider feasibility issues. This discussion is brief, and somewhat superficial; it focuses on identifying next steps and challenges for implementation, and providing a preliminary assessment of their difficulty. Jurisdictions or institutions contemplating introduction of a utility-based noneconomic damages schedule would undoubtedly benefit from a more thorough and context-specific analysis of these issues.

A. Tiering

We scaled only 48 health states. The severity weights we estimated for them had good spread across the spectrum and, thanks to approximately two thousand ratings available for

²² See footnote 7, above.

²³ The AMA Guides' underweighting of fractures and other musculoskeletal conditions is a recognized problem. Some Australian jurisdictions have sought to address it by making special upward adjustments for these conditions at the point of monetizing the impairment ratings.

each condition, relatively small standard errors. Nonetheless, nearly all of the weights still had confidence intervals that overlapped with those of their near neighbors on the scale. There is probably little or no true difference between survey participants' perceptions of the severity of health states whose severity weights were very close.

Clustering adjacent severity weights on the scale into bands, or tiers—about ten tiers with 4-5 weights in each may be reasonable—would confer some advantages. It would simplify application of the schedule. It would also avoid pressing the decision-maker into a degree of precision that the scaled values do not support. Also evident from the array of severity weights shown in Figure 1, however, is that there are few natural “cut points” to guide the creation of tiers.²⁴ Therefore, where the breaks are made may be quite arbitrary. The equity implications of that arbitrariness will turn on the size of the differences in the dollar values or ranges assigned to successive tiers.

B. Slotting

As previously noted, the 48 states, despite being fairly generic in their composition and selected to match common bases of claims, will not match all (or even most) of the health states an adjudicator will be called upon to assess. The pre-scaled health states that appear in each tier of the schedule should function as a collection of indicators of the severity of conditions that belong in that tier. The adjudicator's task then narrows to slotting an incident claim into the family of health states it best fits.

There is, of course, scope for variability and inconsistency to be introduced at this step, which cuts against the effort at precision in the original scaling. One way to mitigate this problem would be to invest in rigorous methods of decision making for a run-in period. In the context of an administrative compensation scheme, for example, consider entrusting the task of sorting an initial tranche of claims to panels of lay and expert decision makers, who would make the slotting decisions by consensus. The process would be resource intensive and almost certainly unsustainable for schemes with large caseloads. But committing to such an approach for a limited period would serve two purposes. It would permit expansion of an indicative set of health states within each tier (i.e. beyond those with established preference-based severity weights). It would also be an opportunity to expose adjudicators—either by including them on panels or allowing them to observe proceedings—to a range of perspectives on how to use the tool. Both are benefits because

²⁴ For exceptions, see the separations that occur above mild hearing loss in the lower section of the scale and above severe anxiety disorders in the upper section of the scale.

they help to ensure reliable uses of the schedule in future cases, when the determination process reverts to a more *ad hoc* and streamlined one.

C. Monetizing

What agencies, courts, and claimants ultimately care about is how much compensation should be awarded. Thus, by grading noneconomic losses only in non-monetary terms, a health-utility approach is patently incomplete (Karapanou and Visscher, 2010; Karapanou 2011; Studdert et al. 2011). A few alternative approaches to empirical assessment of noneconomic loss furnish dollar values (Studdert et al. 2011). Most notably, contingent valuation approaches, which are grounded in assessments of individuals' willingness to accept compensation for risk (or willingness to pay for risk reductions), can assign actual dollar figures to different health states (Berla et al. 1990; Andrews et al. 1996). However, contingent valuation methods remain controversial; they have been greeted with scepticism by many courts; and, for a variety of reasons described elsewhere, are ill-suited to guiding noneconomic damages determinations (Studdert et al. 2011).²⁵ Nonetheless, if the core goal of loss assessment is to quantify damages, severity weights must be monetized.

This is no mean feat. A number of studies have outlined difficulties adjudicators face in attempting to assign dollar values to levels of pain and suffering (Diamond 1993; Geistfeld 1995; Studdert et al. 2011). Additionally, a series of studies of punitive damages assessment have found that individuals achieve reasonable consensus regarding the severity of punishment appropriate for different cases, but that considerable divergence arises at the point of translating those judgments into dollar values (Kahneman et al. 1998; Sunstein et al. 1998; Schkade et al. 2000).²⁶

How does one move from severity weights on a health utility scale to monetary measures of noneconomic loss? The simplest approach would be a proportional transformation of severity weights to dollar values. Because severity weights have both ordinal and cardinal meaning, such a transformation reduces to determining and applying a constant "exchange rate". To derive this rate, dollar values for two health conditions on the scale are required.

²⁵ One reason for this is that contingent valuations of injuries in the compensation setting appear to be quite sensitive to respondents' levels of education and wealth (Jou & Chen 2015).

²⁶ This series of studies concluded that the monetization step in punitive damages assessment explained most of the erratic nature of these awards. However, the key measures used in these studies—in keeping with the researchers' focus on punitive (not compensatory) damages—related to the outrageousness of the defendant's conduct and the appropriate level of punishment. These are very different constructs than levels of pain and suffering, making their applicability in the current context questionable.

The obvious candidates are at the bottom and top of the scale: one health state worthy of zero compensation and another worthy of maximum compensation. (Many compensation systems already make a choice very like this one by designating minimum damages or injury severity thresholds and caps.) With dollar values for two health states in hand, it becomes straightforward to calculate an exchange rate and then apply it to the severity weights for the other health states up and down the scale to produce dollar values for them.

An objection to this approach is that the “true” exchange rate between quality of life and money may not be constant. If, for example, people tend to place a disproportionately large monetary value on severe quality of life losses, relative to less severe ones, a constant exchange rate will undervalue them. Studies comparing willingness-to-pay valuations with quality of life measures have shown evidence of such non-proportionality, although the extent to which this result is driven by artefacts of study design and cognitive biases is unknown (Pinto-Prades et al. 2009; Bobinac et al. 2010; Bobinac et al. 2012 Pennington et al. 2013; Ryen and Svensson, 2014).

One response to this objection is to assert, as a normative matter, that proportionality is nonetheless reasonable, desirable, and fair. An approach that awards \$100,000 in noneconomic damages to claimant A and \$50,000 claimant B, based on a careful determination that A’s condition is associated with a loss in quality of life that is twice as severe as B’s, and then consistently applies this rubric across cases, adheres to a defensible notion of vertical and horizontal equity; it is also thoroughly consistent with the goals of compensating non-economic loss. The results may deviate from a scale of vertical equity implied by alternative approaches to monetizing loss, such as contingent valuation methods. But those other approaches require their own normative justifications, and none, in our view, are manifestly superior.

Two other points speak to the normative merits of a constant exchange rate assumption. First, while it may be tempting to prefer a willingness-to-pay approach for its deftness in directly monetizing and circumventing proportionality assumptions, there is some sleight of hand here. Contingent valuation methods do not avoid the scaling and monetizing steps that confront the health utilities approach confronts. Rather, the health utilities approach isolates and tackles them separately, while contingent valuation conflates them—a move that invites various forms of mismeasurement and bias (Jou & Chen 2015; Venkatachalam 2004). Thus, comparisons between the two approaches should not be cast as a contest between one approach that is vulnerable to biases in translating non-monetary measures to

monetary ones (health utilities) and one that is not (willingness to pay), but as trade-offs between different forms of imprecision. Second, alternative approaches are not the only relevant comparator. The status quo remains the crucial benchmark. We believe that the simple proportional transformation described above would yield dollar values for pain and suffering that are much more consistent, valid, reliable, and defensible than those routinely arrived at today by juries and use of the Guides.

A different response to the non-proportionality objection is to consider applying an exchange rate that is not constant but varies across the severity spectrum. A related question—how to assign dollar values to QALYs—has become an active area of research in the last decade, chiefly because policymakers have sought cost-effective “thresholds” to guide resource allocation decisions (Pinto-Prades et al 2009; Robinson et al 2013; Ryen & Svensson, 2014). This research has not yet produced non-linear transformation functions, but it could do so in the foreseeable future. A function of this kind could be used as the basis of an exchange rate that varies with severity levels and is applied at the monetization step. Better still, creating both health utility and willingness-to-pay measures for a sample of health states like ours may provide the raw material needed for deriving this kind of exchange rate from scratch, and ensure it is well-matched to compensation-related injuries.

Whatever monetization approach is chosen, it is important to remember that compensation amounts need not be decided entirely by neutral application of the exchange rate.²⁷ Rather, the estimates can produce dollar ranges for tiers of health states. Adjudicators may then be given flexibility to choose within those ranges, or even go outside them on occasions, although affording broad discretion introduces disadvantages.

D. Regulating Discretion

Rules regarding the amount of discretion adjudicators have around the schedule—both in when it is used, and how closely its outputs must be adhered to—should be settled *a priori* and be clear. Will use of the schedule be mandatory, presumptive, or discretionary? Preserving some discretion may be politically advantageous and fairer. Karapanou has argued for case-specific factors to be taken into account when determining where to locate an individual plaintiff’s loss within a damages range (Karapanou 2011). At the very least, there is a strong case to be made for considering the claimant’s age or life expectancy, and,

²⁷ We have previously argued for embedding technical approaches to monetizing loss in a democratic, transparent, and deliberative process (Studdert et al. 2011).

all else equal, awarding more to claimants who are expected to live longer with the condition (Avraham 2006).²⁸

However, permitting discretion in when and how the schedule is used comes at a cost. It encourages adversarialism and may result in protracted disputes over determinations. Adjustments for individual and case-specific factors that occur outside the schedule may advance equity *within* particular cases, but *across* cases they reopen the door to a source of variability that may be inequitable. For that reason, integrating structured options for case-level adjustment into the schedule itself may be preferable. Again, fairness is likely to be served by strata or multipliers that deal with age or life-expectancy. Depending on the number of adjusters chosen, however, the schedule may quickly become complicated, thereby threatening the returns to efficiency and consistency from using schedule-guided decision making (Studdert et al. 2011; Engstrom 2015).

VII. Conclusion

Notwithstanding academic concerns about their very existence, noneconomic damages remain a form of relief that is woven into the fabric of injury compensation systems throughout the developed world. Controversy about how they are calculated will continue, most visibly in the ongoing debate over the merits of damages caps in the tort system.

Even though many stakeholders are dissatisfied with current approaches to determining noneconomic damages, moves to shift to new ones will inevitably encounter stiff opposition. A utility-based damages schedule won't be an easy sell to legislators and policymakers. Its attention to the views of the community is attractive, as is the predictability it promises. However, obscure concepts like "utilities", complex statistical steps for deriving weights, and the standardization, averaging, and prescriptiveness at the heart of the model will attract criticism, not least from trial lawyers.

²⁸ Our method of constructing severity weights was designed to make them time-neutral. This was achieved by limiting the sample to permanent health states and instructing survey participants that the individuals described in the pairwise comparison questions had an equal amount of time to live. In applying the schedule, however, claimants won't be identically situated. Two claimants who have an identical condition may have a wide age difference and/or a large difference in life expectancy. Considering the amount of time a claimant will spend in health state may be considered a legitimate "adjuster", and there is ample precedent for this in the civil justice system. For example, courts in a majority of US states already permit jurors to consider noneconomic damages using a *per diem* method in which quantum for a particular unit of time is estimated (e.g. a day or a year) and then multiplied by the time the injury is expected to last (King 2003).

But if merit counts, the utility-based schedule warrants consideration as a serious contender for changing the way decision makers in the civil justice system engage in assessments of noneconomic damages. It is a novel approach with considerable promise to bring greater rationality to these decisions. As a technical matter, it looks feasible. This is largely because the methods needed to derive robust health utility measures stand on the shoulders of decades of work in developing and refining such measures for other purposes in the fields of decision science, public health, and health economics.

A utility-based schedule is an imperfect solution to the problem quantifying noneconomic damages for reasons other than realpolitik. Principal among the approach's weaknesses are deflection of the critical question of dollar valuations, and the discretionary elements that will almost certainly need to be preserved to make use of the schedule palatable. These limitations attenuate some of the benefits of a schedule-based approach. But it is important to recognize that this is not new baggage introduced by the approach we have outlined. Those difficulties are embedded in every other approach to assessing noneconomic damages. The "fault" of the health utilities approach is to lay them bare.

Acknowledgements

The authors thank Ronen Avraham, Jeremy Goldhaber-Fiebert, and Robert Rabin for their helpful comments on an earlier draft of this manuscript.

References

Abel, Richard L. 1986. Should Tort Law Protect Property Against Accidental Loss? *San Diego Law Review* 23(1):79.

Abel, Richard. 2006. General Damages Are Incoherent, Incalculable, Incommensurable, and Incoherent (But Otherwise a Great Idea). *DePaul Law Review* 55(253):291-304.

American Medical Association. 1993. *Guides to the Evaluation of Permanent Impairment*. 4th ed. Chicago, IL: American Medical Association.

American Medical Association. 1990. *Guides to the Evaluation of Permanent Impairment*. 3rd ed. Chicago, IL: American Medical Association.

American Medical Association. 2008. *Guides to the Evaluation of Permanent Impairment*. 6th ed. Chicago, IL: American Medical Association.

Andrews, Paul., Robert G. Meyer, Edward P. Berlá. 1996. Development of the Lost Pleasure of Life Scale. *Law and Human Behavior* 20(1):99-111.

Ara, Roberta and John Brazier. 2012. Comparing EQ-5D Scores for Comorbid Health Conditions Estimating Using 5 Different Methods. *Medical Care* 50(5):452-459.

Association of Workers' Compensation Boards of Canada. 2015. Comparative Table: Permanent Disability Awards and Escalation Benefits – Summary – 2015. http://awcbc.org/wp-content/uploads/2013/12/Permanent_Disability.pdf

Avraham, Ronen. Forthcoming. Estimating Pain and Suffering Damages – Paths are Many, Loss is One. *Oxford Handbook of Law and Economics*. Oxford University Press.

Avraham, Ronen. 2006. Putting a Price on Pain-and-Suffering Damages: A Critique of the Current Approaches and a Preliminary Proposal for Change. *Northwestern University Law Review* 100(1):87-120.

Bambrick, Hilary., Josh Fear, and Richard Denniss. 2009. What Does \$50,000 Buy in a Population Survey? Characteristics of Internet Survey Participants Compared With a Random Telephone Sample. The Australia Institute Technical Brief No. 4.

Barth, Peter S. 2004. Compensating Workers for Permanent Partial Disabilities. *Social Security Bulletin* 65(4):16.

Berlá, Edward P. 1990. Hedonic Damages and Personal Injury: A Conceptual Approach. *Journal of Forensic Economics* 3:1-8.

Blumstein, James F. 2005. Making the System Work Better: Improving the Process for Determination of Noneconomic Loss. *New Mexico Law Review* 35: 401, 405, 410.

Bonner, Robert E. 2008. AMA Guides – Challenges in Application. *International Association of Industrial Accident Boards and Commissions Journal* 45:93–107

Bobinac, Ana., N. Job A. van Exel, Frans F.H. Rutten, Werner B.F. Brouwer. 2010. Willingness To Pay For a Quality-adjusted-life-year: the Individual Perspective. *Value in Health* 13(8):1046-1055.

Bobinac, Ana., N. Job A. van Exel, Frans F.H. Rutten, Werner B.F. Brouwer. 2012. Get More, Pay More? An Elaborate Test of Construct Validity of Willingness To Pay per QALY Estimates Obtained Through Contingent Valuation. *Journal of Health Economics* 31(1):158–168.

Bovbjerg, Randall R., Frank A. Sloan, and James F. Blumstein. 1989. Valuing Life and Limb in Tort: Scheduling Pain and Suffering." *Northwestern University Law Review* 83:908.

Bradley, Ralph A. 1976. Science, Statistics, and Paired Comparisons. *Biometrics* 32:213–39.

Brazier, John., Julie Ratcliffe, Joshua A. Salomon and Aki Tsuchiya. 2007. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press.

Butler, Desmond A. 2005. A Comparison of the Adoption of the Ipp Report Recommendations and Other Personal Injuries Liability Reforms. *Torts Law Journal* 13(2):203-215.

Cameron, Ian D., Trudy Rebeck, Doungkamol Sindhusake, George Rubin, Anne-Marie Feyer, John Walsh, and William Schofield. 2008. Legislative Change is Associated With Improved Health Status in People with Whiplash. *Spine* 33(3):250-254.

Cassidy, J. David, Linda J. Carroll, Pierre Côté, Mark Lemstra, Anita Berglund, and Åke Nygren. 2000. Effect of Eliminating Compensation for Pain and Suffering on the Outcome of Insurance Claims for Whiplash Injury. *New England Journal of Medicine* 342(16):1179-1186.

Chamallas, Martha and Jennifer B. Wriggins. 2010. *The Measure of Injury: Race, Gender, and Tort Law*. NYU Press.

Croley, Steven P. and Jon D. Hanson. 1995. The Nonpecuniary Costs of Accidents: Pain-and-Suffering Damages in Tort Law. *Harvard Law Review* 108:1785

DePianto, David. 2012. Hedonic Impact of Stand-Alone Emotional Harms - An Analysis of Survey Data. *Law & Psychology Review* 6:115.

Diamond, Shari S. 1993. *What Jurors Think: Expectations and Reactions of Citizens Who Serve as Jurors*. Pp. 282, 297 in *Verdict: Assessing the Civil Jury System*, edited by RE Litan. Brookings Institution Press.

Dolan, Paul. 1996. The Effect of Experience of Illness on Health State Valuations. *Journal of Clinical Epidemiology* 49:551-64.

Dolan, Paul., Jan Abel Olsen, Paul Mezel, and Jeff Richardson. 2003. An Inquiry Into the Different Perspectives That Can Be Used When Eliciting Preferences in Health. *Health Economics* 12:545-51.

Drummond Michael F., Mark J. Schulpher, George W. Torrance, Bernie J. O'Brien, and Greg L. Stoddart. 2005. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd Ed. Oxford University Press.

Engstrom, Nora Freeman. 2015. Exit, Adversarialism, and the Stubborn Persistence of Tort. *Journal of Tort Law* 6(1-2):75-114.

Faure, Michael and Louis Visscher. 2011. The Role of Experts in Assessing Damages - A Law and Economics Account. Rotterdam Institute of Law and Economic Working Paper Series No. 2011/09

Flatscher-Thöni, Magdalena, Andrea M. Leiter, Hannes Winner. 2013. Pricing Damages for Pain and Suffering in Court: The Impact of the Valuation Method (2013) 10(1) *Journal of Empirical Legal Studies* 104-119.

Forst, Linda, Lee Friedman, and Abraham Chukwu. 2010. Reliability of the AMA Guides to the Evaluation of Permanent Impairment. *Journal of Occupational and Environmental Medicine* 52(12):1201-1203.

Geistfeld, Mark. 1995. Placing a Price on Pain and Suffering: A Method for Helping Juries Determine Tort Damages for Nonmonetary Injuries. *California Law Review*. 83(3):773, 783-85, 791-92.

Gold Marthe R., Joanna E. Siegel, Louise B. Russell, and Milton C. Weinstein. 1996. *Cost-effectiveness in Health and Medicine: Report of the Panel on Cost-Effectiveness in Health and Medicine*. Oxford University Press.

Grant, Genevieve M. and David M. Studdert. 2009. Poisoned Chalice: A Critical Analysis of the Evidence Linking Personal Injury Processes with Adverse Health Outcomes. *Melbourne University Law Review* 33:865-85.

Greely, Hank T. 2015. Neuroscience, Mindreading, and the Courts: The Example of Pain. *Journal of Health Care Law and Policy* 18(2):171

Haagsma, Juanita A., Charline Maertens de Noordhout, Suzanne Polinder, Theo Vos, Arie H. Havelaar, Alessandro Cassini, Brecht Devleesschauwer, Mirjam E. Kretzschmar, Niko Speybroeck, and Joshua A. Salomon. 2015. Assessing Disability Weights Based on the Responses from 30,660 People from Four European Countries. *Population Health Metrics* 13:10.

Haagsma Juanita A., S. Polinder, E.F. van Beeck, S. Mulder, and G.J. Bonsel. 2009. Alternative Approaches to Derive Disability Weights in Injuries: Do They Make a Difference? *Quality of Life Research* 18(5):657-65.

Hadorn, David C. 1991. The Role of Public Values in Setting Health Care Priorities. *Social Science & Medicine* 32:773-81.

Hammit, James K. 2002. QALYs Versus WTP. *Risk Analysis* 22(5) 985, 987–89, 994–97.

Holmes, Edward B. 2016. Impairment Rating and Disability Determination. Available at: <http://emedicine.medscape.com/article/314195-overview> (accessed January 6 2016).

Horton R. 2012. GBD 2010: Understanding Disease, Injury and Risk. *Lancet* 380(9859):2053-54.

Hyman, David A., Bernard Black, Charles Silver, and William M. Sage. 2009. Estimating the Effect of Damages Caps in Medical Malpractice Cases: Evidence from Texas. *Journal of Legal Analysis* 1(1):355-409.

Jaffe, Louis L. 1953. Damages for Personal Injury: The Impact of Insurance. *Law & Contemporary Problems* 18:219,222.

Jou, Rong-Chang and Tzu-Ying Chen. 2015. The Willingness To Pay Of Parties To Traffic Accidents For Loss of Productivity and Consolation Compensation. *Accident Analysis and Prevention* 85: 1-12

Kahneman, Daniel, David Schkade, Cass R. Sunstein. 1998. Shared Outrage and Erratic Awards: The Psychology of Punitive Damages. *Journal of Risk and Uncertainty* 16:49--86.

Karapanou, Vaia. 2011. Pain and Suffering Damages Based on QALYs: Circumventing Victims' (and Juries') Hedonic Misperceptions. Social Science Research Network.

Karapanou, Vaia. 2014. *Towards a Better Assessment of Pain and Suffering Damages for Personal Injuries: A Proposal Based on Quality Adjusted Life Years*. Portland, OR: Intersentia.

Karapanou, Vaia and Louis Visscher. 2010. Towards a Better Assessment of Pain and Suffering Damages. *Journal of European Tort Law* 1:48-74

Visscher, Louis and Vaia Karapanou. 2015. Non-Pecuniary Losses in the Economic Analysis of Torts: a Plea for Ex Ante Determined Damages. *The Chinese Journal of Comparative Law* 3(2):204-225.

King, Joseph H. Jr. 2003. Counting Angels and Weighing Anchors: Per Diem Arguments for Noneconomic Personal Injury Tort Damages. *Tennessee Law Review* 71:1-50.

Landis, J. Richard and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33:159-174.

Leebron, David W. 1989. Final Moments: Damages for Pain and Suffering Prior to Death. *New York University Law Review* 64:256-363.

Leiter, Andrea, Magdalena Thöni and Hannes Winner. 2012. Evaluating Human Life Using Court Decisions on Damages for Pain and Suffering. *International Review of Law and Economics* 32:119-128.

Loewenstein, George, and Peter A. Ubel. 2008. Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics* 92:1795-1810.

Miller, Ted R. 2000. Valuing Nonfatal Quality of Life Losses With Quality-Adjusted Life Years: The Health Economist's Meow. *Journal of Forensic Economics* 13(2):145-167.

Murray, Christopher J.L., Majid Ezzati, Abraham D. Flaxman, Stephen Lim, Rafael Lozano, Catherine Michaud, Mohsen Naghavi, Joshua A. Salomon, Kenji Shibuya, Theo Vos, and Alan D. Lopez. 2012. GBD 2010: A Multi-investigator Collaboration for Global Comparative Descriptive Epidemiology. *Lancet* 380(9859):2055-58.

Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-Effectiveness in Health and Medicine. 2nd Edition. Oxford University Press, 2016.

Ngui, Emmanuel M., Lincoln Khasakhala, David Ndeti and Laura W. Roberts. 2010. Mental Disorders, Health Inequalities and Ethics: A Global Perspective. *International Review of Psychiatry* 22(3):235-244.

Noah, Lars. 2008. Comfortably Numb: Medicalizing (and Mitigating) Pain-and-Suffering Damages. *University of Michigan Journal of Law Reform* 42:431.

Pennington, Mark., Rachel Baker, Werner Brouwer, Helen Mason, Dorte Gyrd Hansen, Angela Robinson, Cam Donaldson, and the EuroVaQ Team. 2015. Comparing WTP Values of Different Types of QALY Gain Elicited from the General Public. *Health Economics* 24(3): 280-293.

Pinto-Prades, Jose Luis., Graham Loomes, Raul Brey. 2009. Trying To Estimate a Monetary Value for the QALY. *Journal of Health Economics* 28(3):553–562.

Pustilnik, Amanda C. 2014. Painful Disparities, Painful Realities. University of Maryland Francis King Carey School of Law Legal Studies Research Paper No. 2014-18.

Pryor, Ellen S. 1990. Flawed Promises: A Critical Evaluation of the American Medical Association's Guides to the Evaluation of Permanent Impairment. *Harvard Law Review* 103: 964.

Robinson, Angela., Dorte Gyrd-Hansen; Philomena Bacon, Rachel Baker, Mark Pennington, Cam Donaldson, and the EuroVaQ Team. 2013. Estimating a WTP-based Value of a QALY: The 'Chained' Approach. *Social Science & Medicine* 92:92-104.

Robinson, James P., Dennis C. Turk, and John D. Loeser. 2004. Pain, Impairment, and Disability in the AMA Guides. *The Journal of Law, Medicine & Ethics* 32(2):315-326.

Ryen, Linda, and Mikael Svensson. 2015. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. *Health Economics* 24(10):1289-1301.

Safe Work Australia. 2016. Comparison of Workers' Compensation Arrangements in Australia and New Zealand. Canberra, ACT: SWA.

Salomon Joshua A., Theo Vos, Daniel R. Hogan, Michael Gagnon, Mohsen Naghavi, et al. 2012. Common Values in Assessing Health Outcomes from Disease and Injury: Disability Weights Measurement Study for the Global Burden of Disease Study 2010. *Lancet* 380(9859):2129-43.

Sanders, Joseph. 2008. Reforming General Damages: A Good Tort Reform. *Roger Williams University Law Review* 13(1):115, 123–28

Schkade, David, Cass R. Sunstein, Daniel Kahneman. 2000. Deliberating About Dollars: The Severity Shift. *Columbia Law Review* 100:1139–75.

Schultz, Izabela Z, and Catherine M. Chlebak. 2014. Disability and Impairment in Medicolegal Settings: Pain Disability Controversies. Pp. 251-272 in *Handbook of Musculoskeletal Pain and Disability Disorders in the Workplace*, edited by Robert J. Gatchel and Izabela Z. Schultz. New York: Springer-Verlag.

Schwartz, Alan. 1988. Proposals for Products Liability Reform: A Theoretical Synthesis. *Yale Law Journal* 97:353.

Seabury, Seth A., Frank Neuhauser, and Teryl Nuckols. 2013. American Medical Association Impairment Ratings and Earnings Losses Due to Disability. *Journal of Occupational and Environmental Medicine* 55(3):286-291.

Sengupta, Ishita., Marjorie L. Baldwin and Virginia P. Reno. 2014. Workers' Compensation: Benefits, Coverage, and Costs, 2012. National Academy of Social Insurance, Washington, DC.

Sharkey, Catherine M. 2005. Unintended Consequences of Medical Malpractice Damages Caps. *New York University Law Review* 80:391-512.

Sinclair, Sandra and John F. Burton. 1994. Measuring Non-economic Loss: Quality-of-Life Values Versus Impairment Ratings. *Workers' Compensation Monitor* (July-Aug):1-14.

Spieler, Emily A., Peter S. Barth, John F. Burton Jr, Jay Himmelstein, and Linda Rudolph. 2000. Recommendations to Guide Revision of the Guides to the Evaluation of Permanent Impairment. *JAMA* 283(4):519-523.

Sunstein Cass R. 2008. Illusory losses. *Journal of Legal Studies* 37:S157-S194.

Sunstein, Cass R., Daniel Kahneman, and David Schkade. 1998. Assessing Punitive Damages (with Notes on Cognition and Valuation in Law). *Yale Law Journal* 107:2071–2153.

Studdert, David M., Allen Kachalia, Joshua A. Salomon, and Michelle M. Mello. 2011. Rationalizing Noneconomic Damages: A Health-Utilities Approach. *Law & Contemporary Problems* 74:57.

Studdert, David M., Y. Tony Yang, and Michelle M. Mello. 2004. Are Damages Caps Regressive? A Study of Malpractice Jury Verdicts in California. *Health Affairs* 23(4):54-67.

Thurstone, Louis L. 1927. A Law of Comparative Judgment. *Psychology Review* 34:273–86.

Ubel, Peter A and George Loewenstein. 2008. Pain and Suffering Awards: They Shouldn't Be (Just) About Pain and Suffering. *Journal of Legal Studies* 37:S195.

United Kingdom Ministry of Justice. 2013. Reducing the Number and Costs of Whiplash Claims: A Government Response to Consultation on Arrangements Concerning Whiplash Injuries in England and Wales. London, UK: The Stationery Office Ltd.

Venkatachalam L. 2004. The Contingent Valuation Method: A Review. *Environmental Impact Assessment Review* 24(1):89-124.

Viscusi, W Kip. 2013. Empirical Analysis of Tort Damages. Pp. 460-485 in *Research Handbook on the Economics of Torts*, edited by Jennifer Arlen. Edward Elgar Publishing.

Vos, Theo., Abraham D. Flaxman, Mohsen Naghavi, Rafael Lozano, Catherine Michaud, et al. 2012. 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* 380:2163-96.

Wissler, Roselle., Michael J. Saks, and Allen J. Hart. 1999. Decisionmaking About General Damages: A Comparison of Jurors, Judges, and Lawyers, *Michigan Law Review* 98:751-826.

Table 1: Health States Assessed in the Study

Musculoskeletal problems
Low back pain, mild
Back pain, chronic, (with or without leg pain)
Neck pain, chronic (mild, severe)
Musculoskeletal problems, lower limbs, (mild, moderate, severe)
Musculoskeletal problems, upper limbs (mild, moderate)
Musculoskeletal problems, generalised (moderate, severe)
Pain and physical limitations
Mild pain (due to a broken bone in the past)
Pain and difficulty using a part of the body (due to broken bone in the past) (neck of femur; pelvis; foot bones; patella, tibia or fibula or ankle; radius or ulna)
Pain, swelling, tingling and limited feeling in part of the body (due to crush injury)
Numbness and difficulty moving a part of the body (due to nerve injury)
Amputation of finger(s), excluding thumb
Amputation of one limb (upper or lower)
Amputation of both lower limbs (less severe, severe)
Spinal cord lesion below neck level (less severe, severe)
Spinal cord lesion at neck level (less severe, severe)
Scarring (due to burns) (moderate, ² severe ³)
Traumatic brain injury (mild, moderate, severe)
Mental disorders
Anxiety disorders (mild, moderate, severe)
Major depressive disorder (mild, moderate, severe episodes)
Sensory disorders
Hearing loss (mild, moderate, complete, complete with ringing)
Distance vision impairment (severe)
Other
Disfigurement (mild, moderate, severe)

NOTES: See Appendix for the descriptions and wording of health states that were provided to study participants in the 2010 Global Burden of Disease project. Moderate scarring (due to burns) was defined as <20% total burned surface area or <10% total burned surface area if head/neck or hands/wrist involved. Severe scarring (due to burns) was defined as ≥20% total burned surface area or ≥10% total burned surface area if head/neck or hands/wrist involved

Table 2: Characteristics of Survey Participants, Compared to the Australian General Population

	Survey Participants (n = 4,129)	Australian Population
State or territory of residence		
New South Wales	31% (1,265)	32%
Victoria	27% (1,102)	25%
Queensland	19% (790)	20%
Western Australia	10% (415)	10%
South Australia	7% (295)	8%
Tasmania	2% (83)	2%
Australian Capital Territory	3% (143)	2%
Northern Territory	1% (36)	1%
Age group		
18-29 years	16% (651)	21%
30-39 years	18% (732)	18%
40-49 years	20% (839)	18%
50-59 years	20% (811)	17%
60-69 years	17% (714)	13%
70+ years	9% (382)	13%
Sex		
Male	49% (2,024)	49%
Female	51% (2,105)	51%
Country of birth		
Australia	78% (3,239)	65%
Other	22% (890)	35%
Highest level of education		
Secondary education or below	25% (1,307)	47%
Technical or further education	26% (1,060)	28%
University or tertiary education	49% (2,027)	25%
Household income		
Less than \$400 / week (<\$20,799 annually)	18% (730)	12%
\$400 to less than \$1,000 / week (\$20,800 – \$36,399 annually)	31% (1,272)	25%
\$1,000 to less than \$1,500 / week (\$52,000 – \$77,999 annually)	17% (711)	15%
\$1,500 to less than \$2,000 / week (\$78,000 – \$103,999 annually)	12% (506)	11%
\$2,000 or more / week (\$104,000 or more annually)	11% (449)	26%
Unknown	11% (461)	11%

NOTES: The counts and percentages reported are unweighted and relate to respondents to Survey Versions 1, 2, and 3 of the web survey. Estimates of the Australian population come from Australian Bureau of Statistics (ABS) 2011 National Census data and relate to adults 18 years or older. Education attainment data comes from the ABS Survey of Education and Work and relate to persons aged 15-64 years and excludes people who are permanently unable to work. The "Secondary education or below" category includes people with technical or further education who did not complete a high school education. Household income data comes from the ABS and are not restricted to adults 18 years and older.

Author Manuscript

Figure 1: Estimated severity weights and 95% uncertainty intervals

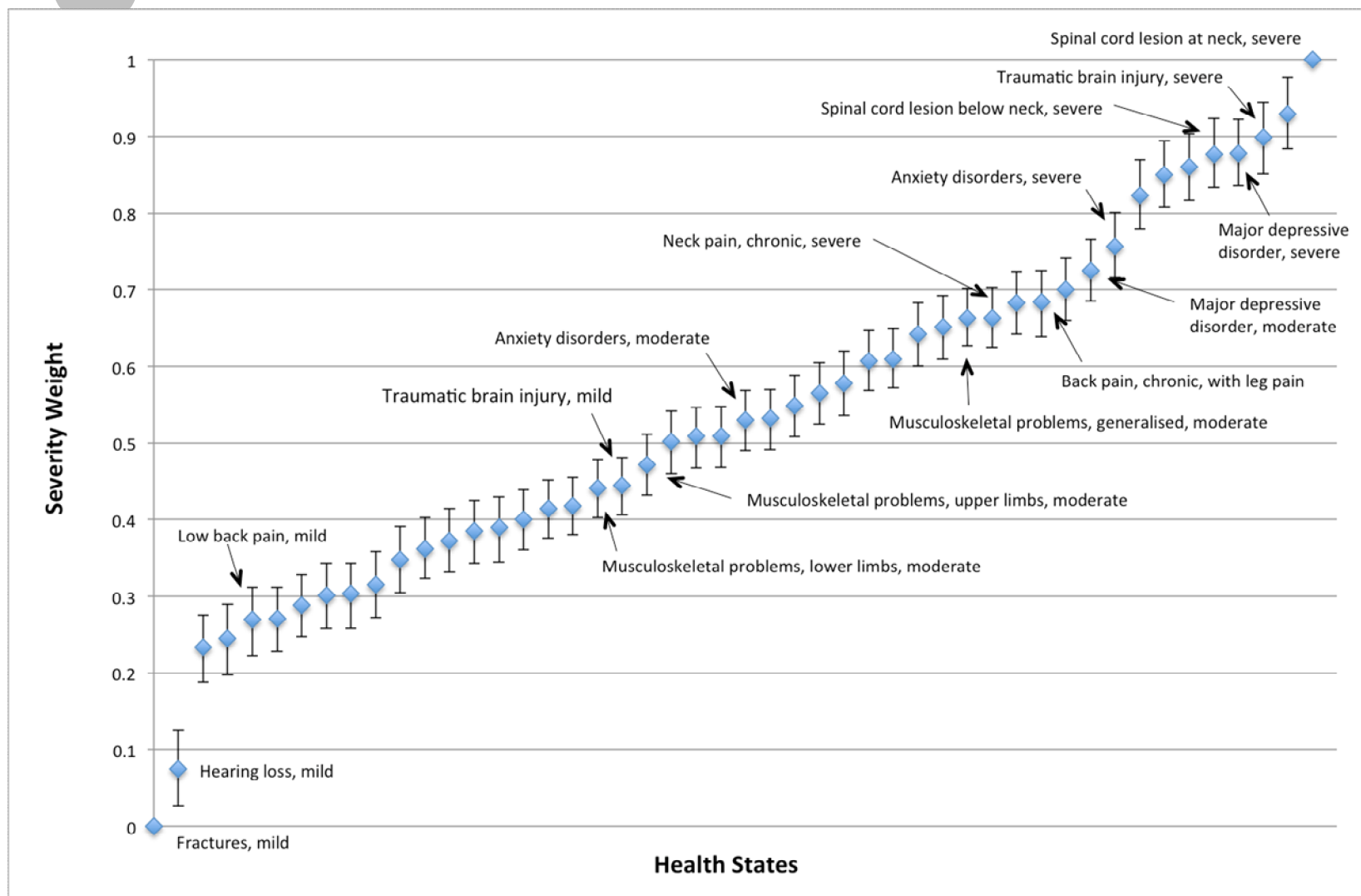
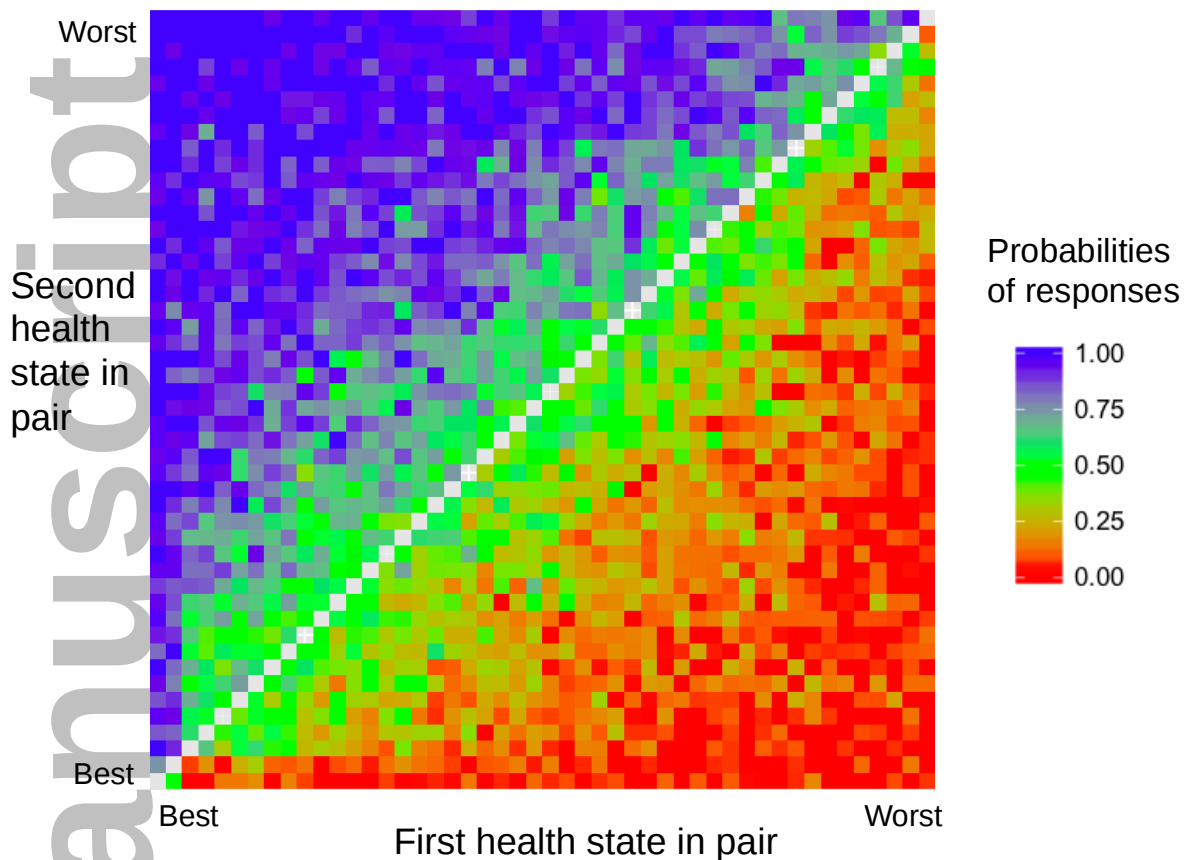
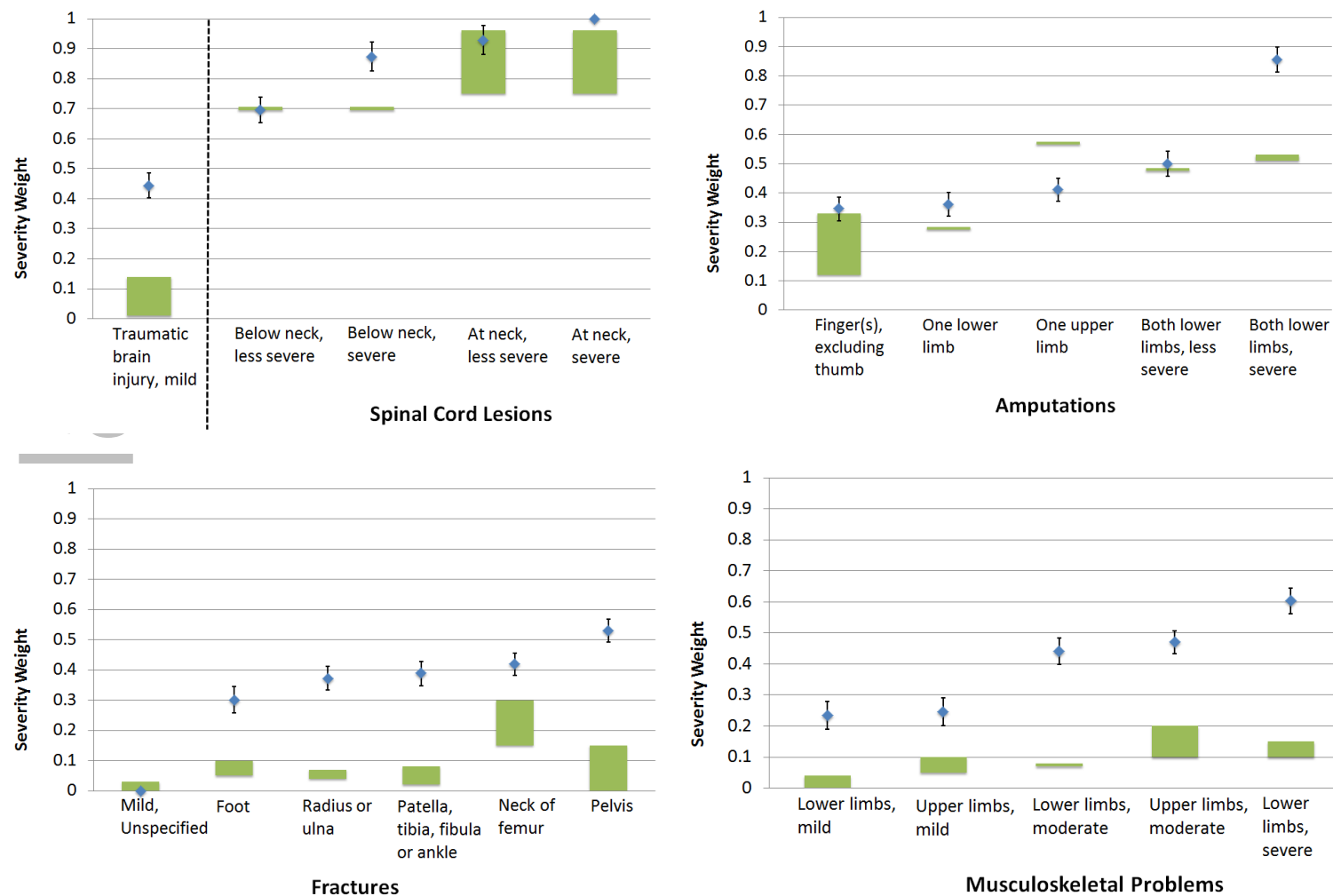


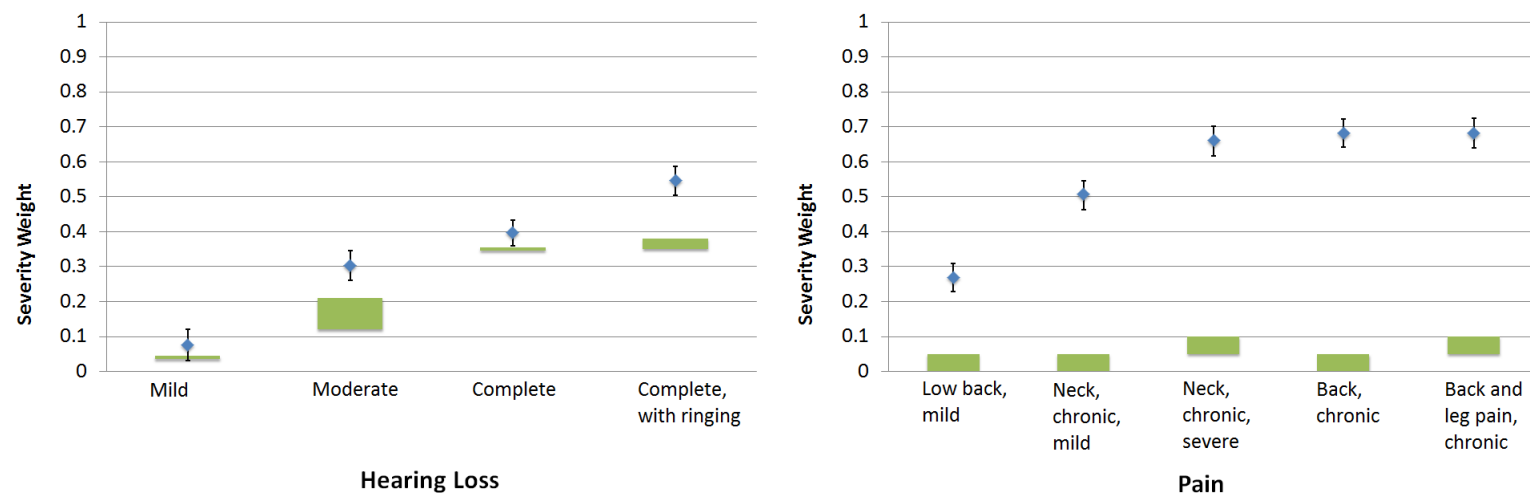
Figure 2: Heat map of response probabilities for all pairwise response questions



NOTES: The x and y axes reflect the first health state and second health state, respectively, described in the pairwise comparisons questions. Health states are ordered from lowest to highest severity weight (or best to worst quality of life) on both axes. The response probability corresponds to the probability that the first health state in a pairwise comparison was chosen as the preferred health state.

Figure 3: Comparison of estimated severity weights with AMA rating or interval





NOTES: The estimated severity weights (with 95% confidence intervals) are shown as dots and American Medical Association (AMA) ratings (with plausible ranges) are shown as bars. The figure presents comparisons for health states for which it was feasible to rate whole body impairment using the AMA Guides.