Disability Weights Measurement for 17 Diseases in Japan: A Survey Based on Medical Professionals

PIAO, Xiangdan and TSUGAWA, Shuichi and TAKEMURA, Yukie and ICHIKAWA, Naoko and KIDA, Ryohei and KUNIE, Keiko and MANAGI, Shunsuke

1 February 2021

Disability Weights Measurement for 17 Diseases in Japan: A Survey Based on Medical Professionals

Xiangdan PIAO*, Shuichi TSUGAWA**, Yukie TAKEMURA***, Naoko ICHIKAWA***, Ryohei KIDA***, Keiko KUNIE****, Shunsuke MANAGI*

Abstract

When judging a population's health to determine disability-adjusted life years, disability weight is a tool for measuring the severity of disability caused by a disease. However, previous studies have pointed out that surveys targeting ordinary citizens produce unclear disability weight values. Therefore, in an attempt to obtain clearer estimations, we conduct a paper-based questionnaire survey of medical professionals—nurses with over ten years of experience—believed to have extensive knowledge of diseases and experience in patient care. We find that disability weight estimations based on the survey of medical professionals presents higher values than those based on a survey of ordinary citizens using the same estimation approach, especially for non-terminal-stage diseases. This suggests that medical-professionals-based surveys may correct the underestimated disability weights of non-terminal diseases (e.g., early stage of cancers and mellitus) found through ordinary-citizens-based surveys. Moreover, we illustrate that depressive disorder and early-stage cancers have almost the same health loss since their disability weights are similar. While regulating policy, it is recommended that more attention be paid to non-terminal diseases and depression.

Key words: Disability weights; Japan; Medical professionals

* Urban Institute & Department of Civil Engineering, School of Engineering, Kyushu University 744 Motooka Nishi-ku Fukuoka,819-0395, Japan; ** Department of Law, Economics and Management, Okinawa University; 555 Kokuba Naha-shi Okinawa, 902-8521; *** Department of Nursing Administration, Division of Health Sciences and Nursing, Graduate; School of Medicine, the University of Tokyo; **** Department of Nursing Administration, School of Nursing, Tokyo Women's Medical University

Corresponding author: Shunsuke MANAGI. Email: managi@doc.kyushu-u.ac.jp

1. Introduction

When measuring a population's health status, the concept of disability-adjusted life years utilizing disability weights is adopted by global health organizations worldwide (Murray, Lopez and World Health Organization, 1996; Murray et al., 2012; Salomon et al., 2012). An example of this application is the Global Burden of Disease (GBD) project, which provides a list of health loss severities due to different diseases, injuries, and other risk factors worldwide, with disability weights as the key factors (Murray et al., 2012; Salomon et al., 2012). A disability weight expresses the severity of disability caused by a disease scaled between 0 and 1, where 0 equals perfect health and 1 is equivalent to death.

Previous studies have established an empirical evaluation of disability weights based on a worldwide survey (Murray et al., 1996; Jelsma et al., 2000; Brennan et al., 2007; Yoon et al., 2007; Basiri et al., 2008; Hong and Saver, 2009; Lai, Habicht and Kiivet, 2009; Lyons et al., 2011; Van Spijker et al., 2011; Salomon et al., 2015). Murray et al. (1996) and Salomon et al. (2012, 2015) focus on a global evaluation of disability weights, whereas other studies focus on specific countries and areas. Disability weight evaluations for developed countries include those of Brennan et al. (2007) for Australia, Yoon et al. (2007) for South Korea, Lyons et al. (2011) for the United Kingdom, and Van Spijker et al. (2011) for the Netherlands. Aside from those of developed countries, studies focusing on developing countries have accumulated in recent years. These include Basiri et al. (2008) for Iran and Jelsma et al. (2000) who compare global disability weights to those weights from Zimbabwe participations.

For the disability evaluation method, existing literature uses the following main approaches to measure individual or social preferences. The approaches include paired comparison (Salomon et al. 2012, 2015), visual analog scale (van Spijker et al., 2011; Baltussen et al. 2002; Murray et al., 1996), and person-trade-off (Murray et al., 1996; Yoon et al., 2007; Salomon et al., 2012, 2015). The paired comparison method entails comparing the severity of health loss among multiple paired diseases. It requires the respondents to choose the healthier individual between two hypothetical individuals with different health states (Salomon et al., 2012). The visual analog scale consists of respondents evaluating health states by assigning a value to each health state, where 0 represents the worst state of health and 10 or 100 represents the best state of health (Baltussen et al., 2002). The person-trade-off method consists of asking the respondents how many outcomes of a particular state of health they consider equivalent to outcomes of another state of health (Nord, 1995). It is believed that the person-trade-off approach is preferable because it represents the social preferences better than individual preferences (Van Spijker et al., 2011).

Samolom et al. (2012) combined paired comparison approach with health population equivalence, which is based on the person-trade-off approach. Please see Haagsma et al. (2014) for a detailed review.

Among the disability weight studies, the methodology adopted in the GBD 2010 project is representative and this valuation method and the results thereof were published by Salomon et al. (2012). For example, in Japan, Nomura et al. (2017) evaluate overall population health based on the disability weights reported in Salomon et al. (2015), using the same methodology used in Salomon et al. (2012). However, the validity of the disability weights is argued by later scholars (Taylor et al., 2013; Nord, 2013). Nord (2013) argues that the measurement of disability weights based on the method proposed in Salomon et al. (2012) has pointed out that the disability weights of some diseases (e.g., deafness) are relatively small, which might be caused by the survey sample containing participants that likely consider the deafness patients to be living with mild disability. When the targeted participants who thought the diseases were not in poor health, then the disability weight value may become small, and vice versa.

Further, disability weight studies have found that surveys based on the general public and those based on medical professionals produce different disability weight values (Jelsma et al., 2000; Baltussen et al., 2002). Baltussen et al. (2002) investigate disability weights of nine diseases between lay people and medical professionals in rural Burkina Faso in Africa through a survey of 37 lay people and 17 medical professionals. It was found that the disability weights from medical professionals were lower than those from lay people, establishing the disability weight from medical professionals as the proxy of general public. Similarly, Jelsma et al. (2000) compared the disability weights from medical professionals, non-professionals, and global burden diseases. They concluded that the weights of diseases' severity are highly correlated with the global burden diseases. However, medical professionals record much lower weights than non-professional Zimbabweans in a survey of 12 medical professionals and 58 non-professionals. The reason the disability weight from medical professionals is higher than that from non-professionals could be that the medical care system in rural Africa is underdeveloped or that the local residents or lay people are concerned about the limited access to medical services. However, existing literature has so far focused on developing nations and studies focused on Japan are scarce. Therefore, a survey targeting medical professionals and results comparison may provide insightful evidence to clarify this phenomenon. Moreover, well-experienced nurses might have a neutral judgment of health loss caused by diseases than doctors and the public. Nursing is defined as "the protection, promotion, and optimization of health and abilities, prevention of illness and injury, facilitation of healing, alleviation of suffering through the diagnosis and treatment of human response, and advocacy in the care of individuals, families, groups, communities, and populations" (American Nurses Association, 2015, p.1). Therefore, nurses can be considered appropriate for assessing the impact of illness on people's lives.

In this study, we aim to estimate the disability weights of 17 diseases in Japan based on the original survey conducted for this study in 2018 using a methodology adopted by the World Health Organization and published in the Lancet (Salomon et al., 2012). The targeted participants of the survey are medical professionals (nurses) who work in national or public university hospitals and have more than 10 years of nursing experience. As argued above, the estimation of disease disability weights based on a survey of ordinary citizens may include unclear values due to participants' knowledge gap of the diseases. Thus, the results derived in this study are expected to provide clearer disability weight estimates, insightful evidence for measuring health loss more accurately, and the implications for related policies.

The outline of this paper is as follows: Sections 2.1 and 2.2 provide the materials and methodology, respectively. The empirical results are summarized in Section 4, and Section 5 presents the conclusion.

2. Methodology

2.1 Materials and Design

To investigate the disability weights of diseases, we conducted a paper-based survey in Japan in December 2018. In a bid to overcome the unclear results derived from surveys based on medical specialists and ordinary citizens, the participants of the current survey included (randomly selected) nurses only. A purposive sampling technique was utilized to select targeted respondents with extensive knowledge of the diseases. More precisely, the questionnaires were sent to the nursing directors of all 42 national university hospitals and all 8 public university hospitals. Each nursing director circulated the questionnaires to nurses with more than 10 years working experience (30 nurses for each national university hospital and 20 nurses for public university hospitals). In total, 1,420 questionnaires were mailed to the targeted hospitals and 294 responses were collected, representing a response rate of 21%.

The respondents were informed about the study's purpose, method, voluntary participation, and anonymity of information and the study was approved by the Research Ethics Committee of the Graduate School of Medicine, The University of Tokyo (No. 2018068NI). In the questionnaire, the respondents were asked the following question: "After reading the instruction of this survey, are you still willing to participate in this research analysis?" to which the responses were either 1, which equaled "agree", or 2, "disagree". Of the 294 respondents, eight declined participation in the measurement of disability weight research or were missing, while the rest consented to having their answers included in the analysis. Therefore, 286 observations are included in our analysis.

When measuring the disability weights, we followed the extensive methodology proposed by Salomon et al. (2012) to design the questionnaire for collecting information on paired comparisons and population health equivalence. The survey included seventeen representative diseases like cancer, stroke, and lifestyle diseases. These entailed six types of 'cancers' (stomach cancer, three forms of bowel cancer, and two of lung cancer), four types of adult disease (including two forms of diabetes, high blood pressure, and hyperlipidemia), four types of stroke, two types of mental disease that disturb daily life (Alzheimer's and dementia, and depression), and chronic kidney disease (requiring dialysis). We focused only on these important diseases to keep the respondents' burden of answering the survey at a minimum.

The key aspects of the questionnaire used for measuring the disability weight of each disease, paired comparison and population health equivalence, were designed as follows. For paired comparisons, the nurses were asked 17 questions, the first of which, for example, was, "Compared to early stomach cancer, which diseases would you rather suffer from? Please select from the list below," followed by a list of the other 16 diseases[1]. It should be noted that respondents who were missing and there were

[1] Early colorectal cancer; Terminal colorectal cancer; Early lung cancer; Terminal lung cancer; Acute myocardial infarction; Stroke: long-term consequences, agnosia (or aphasia); Stroke: long-term consequences, hemiplegia; Stroke: long-term consequences, persistent consciousness disorder;

no choices in the choice list might be the same. To distinguish between the two situations, we added an option in the list that states that “none of the above 16 diseases are preferable”. In total, 171 paired comparisons were obtained.

Regarding population health equivalence, for each disease, the respondents were asked to select the overall health benefit that they believe to be equivalent to the presented hypothetical health policy. The questions were posed, for example, as follows: “Policy A, which prevented 1000 people from getting an illness that causes rapid death, has the same population health benefit as___”, then the respondents were asked to select either “(1) less than 1000; (2) 1500; (3) 2000; (4) 3000; (5) 5000; or (6) more than 10000 people prevented from getting Early stomach cancer.” The population range, from less than 1000 to more than 10000 people, represented the options of the overall health benefit equivalent to Policy A (see Figure 1).

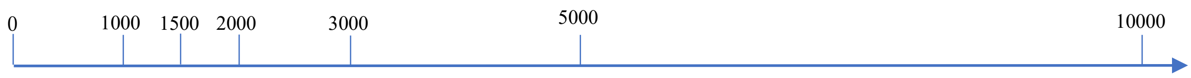


Figure 1. Number of population on the number line of population health equivalence.

After deleting the observation of missing key information, the total number of observations obtained were 286 for paired comparison and 253 for population health equivalence. Moreover, the demographic characteristics of the respondents were included in the survey. These included age, gender, type of qualification, length of experience, position, marital status, number of children, highest education level achieved by the respondent, respondent’s (and their parents and partner’s) yearly income, household income, volunteer experience within one year, health status, consumption on health

Diabetes mellitus: with insulin injection; Diabetes mellitus: without insulin injection; High blood pressure; Dyslipidemia; Alzheimer dementia; Depressive disorder; Chronic renal failure: dialysis.

status, periodic medical examination, and other medical/health-related questions.

2.2 Empirical strategy

For each participant, the disability weights of the diseases are computed according to the paired disease comparison and population health equivalence categories following Salomon et al. (2012). In the case of the paired comparisons, we apply a probit regression that was utilized in the analysis of discrete choice experimental data in Salomon et al. (2012), whereas for the population health equivalence, the interval regression is adopted[2].

For the paired comparisons, we regard the selected choice between the diseases as the dependent variable. Furthermore, we treat the 17 diseases as independent variables and create dummy variables from them. The binary choice variable $Y = 1$ when the comparison health states[3] are considered healthier compared to the primary health state. Thus,

$$P(Y = 1|X) = \Phi(X'\beta), \quad (1)$$

where Φ is the cumulative distribution function of the standard normal distribution, P denotes the probability, and the matrix X is the disease indicator, and β is a set of parameters estimated from the maximum likelihood. Following Salomon et al. (2012), the disease indicators are selected as independent variables and the respondents' socioeconomic factors are *not* included in the probit regression.

The population health equivalence questions provide insight on tradeoffs between mortality and nonfatal outcomes, which is needed to anchor the results from the probit regression analysis (Eq.1) of the paired comparisons onto the disability weight scale (0,1). Because the population health equivalence questions are framed in terms of a binary comparison between an intervention that averted 1000 fatalities and another intervention that averted some number of nonfatal

---

[2] It is thought that the interval regression is appropriate in this study. The population health equivalence portion of the paper-based survey is structures as a categorized choice, which is left-censored at 1250; 1250-1750; 1750-2500; 2500-4000; 4000-7500; and right-censored at a value of 7500. On the contrary, in Salomon et al. (2012), the population health equivalence portion of the survey consists of a randomly chosen value for the participants.

[3] The paired comparison occurs with one of the 16 diseases when one of the diseases is selected as the base disease.

outcomes, responses provide interval-censored information rather than exact values[4]. Therefore,

$$L = Z'\theta + \varepsilon, \tag{2}$$

where L is a logit-transformed variable of the interval population health equivalence of diseases following Salomon et al. (2012). L is logit transformed unobserved continuous outcomes is in the observed interval from population health equivalence. That is, it is left-censored at a value of 1250 in the population health equivalence measurement—1250-1750, 1750-2500, 2500-4000, 4000-7500—and right-censored at a value of 7500[5]. A detailed description of the transformation is shown in the Appendix. The matrix Z represents the dummy variables of the diseases. The parameters of diseases θ, estimated by maximum likelihood, represent the logit-transformed disability weights. Thus,

$$\hat{\beta} = \alpha_0 + \alpha\hat{\theta} + \mu, \tag{3}$$

where $\hat{\beta}$ is the parameter of the 17 diseases from Eq. (1) estimated by the maximum likelihood, whereas $\hat{\theta}$ represents the parameters estimated by Eq. (2) using an interval regression. α_0 and α are the parameters estimated by ordinary least squares. μ is the error term.

The regression results in an estimated slope and intercept, which are then used for predicting the probit coefficients and transforming them onto the (logit) disability-weight scale (Eq.3), as well as for the numerical integration of the obtained disability weights estimates on the natural 0-to-1 scale (Eq.4). First, we simulate the normal random variates on the logit scale with the means of each disease predicted from Eq. (3) and the variance defined by the overall standard deviation of these coefficients across survey-specific estimates from Eq. (2). Next, we transform each of these simulated values through an inverse-logit function (Eq.4). Finally, we compute the mean across the resulting values for each disease. Therefore, we simulate the disability weight w of the 17 diseases by predicting the dependent variable $\hat{\beta}$ in Eq. (3), which we rescale from 0 to 1, as in Eq. (4). Therefore,

$$w = \frac{e^K}{1+e^K}, \tag{4}$$

where w denotes the disability weight of diseases, K is the prediction of the dependent variable $\hat{\beta}$ based on Eq. (3), and e^K is the natural exponential function. All statistical analyses are performed using Stata MP 16.0 software (StataCorp, College Station, TX, USA).

---

[4] Furthermore, due to this study using paper-based questions rather than web-based one, the current analysis method is different from that of Salomon et al. (2012).

[5] For example, the independent variables with logit-transformed values of 1250, 1750, 2500, 4000, 7500 are 1.386; 0.288; -0.405; -1.099; -1.872, respectively.

3 Results

Table 1 displays the characteristics of the respondents, who are skilled nurses with over 10 years of working experience in hospitals. As expected, 89% of the nurses are women, whereas 7% of the nurses are men. Because our survey targets experienced nurses, only 42% of the nurses are younger than 45 years. Regarding appointment, 9% and 41% of the respondents hold the position of director or nurse manager, respectively. The other nurses include assistant nurse manager/charge nurse; staff and other positions share around 50% of the sample.

Table 1. Characteristics of the medical professionals

Variables	Obs.	Percent
Gender		
Male	20	7%
Female	255	89%
unknown	11	4%
Age		
31–35	29	10%
36–40	47	16%
41–45	46	16%
46–50	59	21%
51–55	53	19%
56–60	49	17%
unknown	3	1%
Clinical experience years		
–10	4	1%
11–15	61	21%
16–20	45	16%
21–25	36	13%
26–30	62	22%
31–35	55	19%
36–	20	7%
unknown	3	1%
Appointment		
Director	27	9%

Table 1. continued.		
Nurse manager	116	41%
Assistant nurse manager/charge nurse	80	28%
Staff	59	21%
The other positions	2	1%
unknown	2	1%
Total	286	

Note: Observations of paired comparison: 286; observations of population health equivalence: 253.

Table 2 displays the estimation results of the disability weights of the 17 diseases. The estimated disability weights of the diseases range between 0.346 and 0.605, which indicates that the nurses believe that these diseases have a strong influence on health. The lowest estimated disability weight is that of *high blood pressure*, sitting at 0.346, whereas the highest disability weight is that of *terminal lung cancer*, sitting at 0.605. Intuitively, in terms of disability-adjusted life years, *terminal lung cancer* indicates a high level of health loss, meaning patients may be disabled to the extent that they cannot live without various assistants. Similarly, such high levels of disability are more likely to result from other terminal cancers, stroke with serious sequelae, stroke with long-term consequences, and chronic renal failure with dialysis. On the contrary, low disability weights appear for *high blood pressure*, *diabetes mellitus with/without insulin injection*, and *dyslipidemia*. However, these diseases are thought to have negative influences on lives as the weight ranges between 0.346 and 0.453.

The measurement of disability weight provides an illustration of health loss due to diseases and the comparisons of disease severities. For example, based on the estimation, *depressive disorder* has a disability weight of 0.498 and ranks at 13 as shown in Table 2. Similarly, health losses from early stomach, lung, and colorectal cancers are 0.500, 0.518, and 0.520, respectively. This result illustrates that depressive disorder and early cancers have almost the same health loss. This may provide insightful evidence for using the comparison between the early stages of stomach, colorectal, and lung cancers to understand patients' health losses due to depressive disorder. The results suggest that patients with depressive disorders might need more attention or aid from the government or health organizations regarding their resultant disabilities.

Table 3 presents comparisons between the disability weight results of this study (derived from Table 2) and those of a representative previous study; that is, Salomon et al. (2015), which has been adopted by the World Health Organization and Japan (Nomura et al., 2017). This aims to show the differences in the disability weights found when the participants are medical professionals rather than ordinary citizens so as to correct any unclear values resulting from ordinary citizens' lack of disease knowledge or experience. Where the categories of the diseases are slightly different, we select the closest categories from Salomon et al. (2015) based on the

Table 2: Estimation results of the disability weights of the 17 diseases

Health state	Disability weight	95% Interval		Rank
		Lower bound	Upper bound	
Early stomach cancer	0.500	0.148	0.852	12
Terminal stomach cancer	0.577	0.230	0.924	4
Early colorectal cancer	0.520	0.166	0.875	10
Terminal colorectal cancer	0.599	0.250	0.949	2
Early lung cancer	0.518	0.164	0.872	11
Terminal lung cancer	0.605	0.268	0.943	1
Acute myocardial infarction	0.546	0.200	0.892	8
Stroke: long-term consequences, agnosia (or aphasia)	0.550	0.207	0.893	7
Stroke: long-term consequences, hemiplegia	0.573	0.228	0.917	5
Stroke: long-term consequences, persistent consciousness disorder	0.597	0.251	0.943	3
Diabetes mellitus: with insulin injection	0.453	0.104	0.801	14
Diabetes mellitus: without insulin injection	0.385	0.052	0.717	15
High blood pressure	0.346	0.019	0.672	17
Dyslipidemia	0.367	0.037	0.696	16
Alzheimer's and dementia	0.536	0.189	0.882	9
Depressive disorder	0.498	0.146	0.851	13
Chronic renal failure: dialysis	0.570	0.228	0.912	6

Note: Observations of paired comparison: 286; observations of population health equivalence: 253.

Table 3: Comparison of disability weights between current estimation and previous study

Health state	Disability weight	Disability weight (Salomon et al. 2015)
Early stomach cancer	0.500	
Early colorectal cancer	0.520	0.288
Early lung cancer	0.518	
Terminal stomach cancer	0.577	
Terminal colorectal cancer	0.599	0.54
Terminal lung cancer	0.605	
Acute myocardial infarction	0.546	0.422
Stroke: long-term consequences, hemiplegia	0.573	0.070
Stroke: long-term consequences, agnosia (or aphasia)	0.550	0.316
Stroke: long-term consequences, persistent consciousness disorder	0.597	0.588
Diabetes mellitus: with insulin injection	0.453	0.020 or 0.133
Diabetes mellitus: without insulin injection	0.385	0.020 or 0.133
Alzheimer's and dementia	0.536	0.377
Depressive disorder	0.498	0.396
Chronic renal failure: dialysis	0.570	0.571

Note: In Salomon et al. (2015) Table 2, for *acute myocardial infarction* for days 1-2, the disability weight is 0.432, whereas for days 3-28, it is 0.074. For stroke, the weight of *long-term consequences, mild case* is 0.019; *long-term consequences, moderate case* is 0.070; *long-term consequences, moderate, plus cognition problems* is 0.316; *long-term consequences severe* is 0.552; and *long-term consequences, severe, plus cognition problems* is 0.588. *Diabetic foot* is 0.020, and *diabetic neuropathy* is 0.133. *Major depressive disorder (moderate episode)* is 0.396. *End-stage renal disease on dialysis* is 0.571. *Neurological disorders of dementia (moderate)* is 0.377. *Cancer (diagnosis and primary treatment)* is 0.288; *Metastatic* is 0.451 and *Terminal phase with medication (for cancers and end-stage kidney or liver disease)* is 0.540.

disease description and consequences. Column 1 displays the disability weights derived from Table 2, whereas Column 2 summarizes the disability weights from Salomon et al. (2015).

The results are as follows: we find that the estimated disability weights based on a survey of well-experienced nurses vary from those in Salomon et al. (2015), which are based on a survey of ordinary citizens, despite using the same estimation approach. With regard to the 15 corresponding diseases covered in both studies, the overall results of the current study show higher disability weights than those estimated in previous studies (see Table 3). These results suggest that ordinary citizens underestimate the health loss caused by the diseases compared to nurses. This may be because ordinary citizens have less knowledge about or experience with diseases compared to well-experienced medical professionals.

Regarding serious diseases, we find slight differences and even consistencies between the disability weights found in the current estimations and those found in the previous study. For example, the disability weight of *chronic renal failure: dialysis* in this study is 0.570, whereas in previous studies, the disability weight of this condition was 0.571. Similarly, for *stroke: long-term consequences, persistent consciousness disorder*, we found a slight difference. We estimated the disability weight of this disease at 0.597, whereas the corresponding disability weight in the previous study was 0.588. These relatively small differences are also found among the terminal cancers.

In contrast, the estimated disability weights of early- and middle-stage diseases based on the current survey are higher than those of the previous survey. For example, we find the disability weights of *early stomach cancer*, *early colorectal cancer*, and *early lung cancer* to be 0.500, 0.520, and 0.518, respectively, whereas their corresponding disability weight in the previous study is estimated as 0.288. We also confirm inconsistent results among *stroke: long-term consequences, hemiplegia*; *stroke: long-term consequences, agnosia (or aphasia)*, *Alzheimer's and dementia*, and *depressive disorder*. This suggests that disability weight differences are more likely to appear in the early and middle stages of diseases. The reason for this may be that, because in the beginning stages, some patients with these diseases are able to live independently, some even having the ability to work, ordinary citizens may be less likely to perceive the long-term health loss severities of these diseases. In contrast, for the same non-severe patients, nurses can evaluate both the medical aspect and the impact on life in general, while imagining not only the patient's condition at that time, but also the deterioration in quality of life during the treatment period, the response rate of the treatment, and the long-term course, thus leading to higher values. Therefore, the results of this study are likely the more appropriate assessment of the effects of diseases.

Table 4 compares this study with previous representative studies by the regions they cover in developed and developing countries. The disability weight of disease from Japan is from the current study (column 1), whereas the disability weight of disease from Europe, Africa, the Netherlands, and Globally have been published in previous studies (Schwarzinger et al., 2003; Baltussen et al., 2002; van Spijker et al., 2011; Salomon et al., 2015).

The results are as follows. First, when the disability weight is compared by different countries and regions, it highlights the severe loss of health caused by acute depression. For a severe case of depressive disorder, the disability weight is reported as 0.78 in Europe and 0.74 in the

Netherlands, which shows a higher loss of health than other diseases such as diabetes. In rural Africa, medical professionals also evaluate the disability weight of acute depression as high. In line with previous studies, our results also suggest developing policies that promote psychological well-being while preventing mental and depressive disorders and providing more aid and public attention to patients with such disorders.

Second, the regional evaluation of health loss by different diseases in Europe, Japan, and Africa tend to be higher than the global burden disease Salomon et al. (2015). For example, the disability weight of diabetes is estimated at 0.385-0.452 in Japan; in Europe, it is at 0.34; in Africa, it is at 0.34, whereas the corresponding disease is 0.02 to 0.133 (Schwarzinger et al., 2003; Baltussen et al., 2002; Van Spijker et al., 2011; Salomon et al., 2015). A similar tendency is found in dementia and strokes with long-term consequences. These differences are thought to mainly exist due to the respondents' knowledge of the diseases, the capacity of the medical care system, the respondents' preferences, and the estimation techniques used to measure the disability weight.

Table 5 displays the disability weight of diseases by different health professional groups based on the estimation by Salomon et al. (2015). To show heterogeneity among the medical professionals, age and duties of experienced nurses, which represent major demographic and social characters, are selected. Young or old age groups and nurse manager or non-nurse manager groups are established for comparison. The young group are nurses younger than or 45 years old whereas the old group of the respondents is older than 45 years. Columns 1, 2, 5, and 6 summarize the disability weights of diseases categorized by medical professionals aged 45 years or younger, older than 45 years, nurse manager (includes director and nurse manager), and non-nurse manager (includes assistant nurse manager/charge nurse, other staff, and other positions), respectively. Difference of disability weight between young age group and old age group is displayed in column 3 with the p-value in column 4, whereas the difference between chief and non-nurse manager group is shown in column 7 with p-value in column 8.

The results are summarized as follows. We found heterogenous tendency in 12 out of 17 diseases regarding the severity of diseases between young and old health professional groups. The difference is obvious for non-terminal diseases. For example, the disability weight of early stomach cancer in the young respondent group is 0.519, whereas the corresponding weight by old respondent group is 0.467 (see column 1 and 2). The difference between the two groups is 0.052 and statistically significant at 1%. It indicates that the evaluation of loss of health due to early stage of stomach cancer is more severe in the young respondent group. Consistent results are found in early colorectal cancer, early lung cancer, acute myocardial infarction, diabetes mellitus: without insulin injection, and Alzheimer dementia.

This may because the older nurses have more experience in various departments or that they are more experienced in diseases that cause suffering. Based on these experiences, older nurses may think that patients (e.g., early stomach cancer) with a severe diagnosis may be able to continue social life without obstacles given proper treatment in a medical center and practicing self-care.

On the contrary, regarding terminal diseases, the disability weight difference between young and old respondent groups tends to be insignificant. Terminal diseases include stroke with long-term consequences, agnosia (or aphasia), terminal lung cancer. Finally, the older medical professionals tend to evaluate severe health loss among 6 out of 17 diseases and the difference between the young and old groups are statistically significant at 1%. The 6 diseases are depressive disorder, chronic renal failure (dialysis), Dyslipidemia, high blood pressure, Diabetes mellitus with insulin injection, stroke with long-term consequences, and persistent consciousness disorder. In sum, the older medical professionals have more experience in those diseases, therefore, the disability weights estimated by the older group suggests a stronger policy implication for the government.

When comparing the disability weight of diseases between the nurse manager and non-nurse manager groups we found that disability weights are significantly different among 9 out of 17 diseases with statistical significance at 1%. On one hand, it is worthy to note that the nurse managers are more likely evaluate health loss caused by mild diseases and terminal diseases with higher severity. For example, the mild diseases are Dyslipidemia and high blood pressure and the terminal diseases are terminal colorectal cancer and terminal lung cancer. On the other hand, nurse managers tend to estimate the loss of health by diseases between above mild and terminal diseases with less severity. These diseases include early stomach cancer, terminal stomach cancer, early colorectal cancer, Diabetes mellitus with insulin injection, and Alzheimer dementia. Additionally, insignificant difference between chief and non-nurse managers is observed for other diseases, such as early lung cancer, acute myocardial infarction, stroke with long-term consequences, and agnosia (or aphasia). The difference between the two groups may be because the nurse-manager group includes directors and nurse managers who tend to have longer nursing experiences than other nurses in hospitals. Moreover, in hospitals, directors are connected with other departments or provide counseling services to discharged patients and their families. On the contrary, nurse managers are supposed deepen the regular communication with patients. Therefore, directors or nurse managers are more likely to have a comprehensive understanding of the impact of end-stage cancer on patients and their families.

Table 4: Disability weight of diseases by different region

	Japan (1)	Europe (2)	Africa (3)	The Netherlands (4)	Global (5)
Major depression	0.498	0.78 (severe)	0.66	0.74 (severe)	0.396 (moderate); 0.658 (severe)
Diabetes mellitus	0.453 (insulin); 0.385	0.34	0.34	-	0.020 (foot) or 0.133 (neuropathy)
Dementia	0.536	0.46 (mild)	-	-	0.30
Stroke: long-term consequences	hemiplegia: 0.573; agnosia (or aphasia): 0.550; persistent consciousness disorder; 0.597	0.68 (moderate)	0.55 (paraplegia)	-	0.070 (moderate); 0.316 (moderate with cognition problem); 0.588 (severe with cognition problems)
Resource	current study	Schwarzinger et al. (2003)	Baltussen et al. (2002)	van Spijker et al. (2011)	Salomon et al. (2015)

Table 5: Disability weights for diseases by different health professional groups

	Age≤45	Age>45	Difference	p-value	Nurse manager	Other nurse	Difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Early stomach cancer	0.519	0.467	0.052	<0.001	0.470	0.522	-0.053	<0.001
Terminal stomach cancer	0.588	0.576	0.012	0.022	0.570	0.599	-0.029	<0.001
Early colorectal cancer	0.529	0.503	0.026	<0.001	0.510	0.531	-0.021	<0.001
Terminal colorectal cancer	0.594	0.601	-0.007	0.179	0.605	0.583	0.022	<0.001
Early lung cancer	0.531	0.517	0.015	0.008	0.515	0.520	-0.005	0.376
Terminal lung cancer	0.593	0.595	-0.002	0.723	0.613	0.581	0.031	<0.001
Acute myocardial infarction	0.553	0.533	0.021	<0.001	0.540	0.545	-0.006	0.310
Stroke: long-term consequences, agnosia (or aphasia)	0.562	0.558	0.003	0.555	0.561	0.549	0.012	0.041
Stroke: long-term consequences, hemiplegia	0.578	0.571	0.008	0.158	0.577	0.568	0.010	0.079
Stroke: long-term consequences, persistent consciousness disorder	0.577	0.595	-0.018	0.002	0.598	0.591	0.007	0.189
Diabetes mellitus: with insulin injection	0.448	0.463	-0.015	0.006	0.443	0.461	-0.018	0.001
Diabetes mellitus: without insulin injection	0.406	0.383	0.022	<0.001	0.388	0.395	-0.007	0.197
High blood pressure	0.341	0.361	-0.020	<0.001	0.360	0.335	0.025	<0.001
Dyslipidemia	0.356	0.371	-0.016	0.004	0.379	0.362	0.017	0.002
Alzheimer dementia	0.550	0.517	0.034	<0.001	0.525	0.550	-0.025	<0.001
Depressive disorder	0.479	0.535	-0.056	<0.001	0.513	0.505	0.008	0.160
Chronic renal failure: dialysis	0.550	0.588	-0.038	<0.001	0.567	0.571	-0.004	0.447

Note: nurse manager includes direct and nurse manager and non-nurse manager includes assistant nurse manager/charge nurse and other staff and other positions.

4 Conclusion

This study aims to measure the disability weights of 17 diseases in Japan using a paper-based questionnaire survey of medical professionals. Disability weight is a measurement of the severity of disability caused by a disease, where 0 equals perfect health and 1 equals death. These weights provide insightful evidence for policy makers, particularly regarding determining the population's disability-adjusted life years to improve public welfare and health. However, previous studies have pointed out that the measurement of diseases' disability weights done through surveys that target ordinary citizens include unclear values because ordinary citizens lack disease knowledge. Thus, using the same estimation approach applied in a previous ordinary-citizens-based research, this study entails a survey targeting veteran nurses—those with over 10 years working experiences and extensive knowledge—to identify the differences between the results of the two types of surveys. The main results of this study are summarized as follows:

First, the disability weights found through our survey are higher than those from the survey targeting ordinal citizens. Importantly, the disability weights are much higher among the early- and middle-stage diseases in this study than the previous one. This may be because patients with early- or middle-stage diseases (e.g., depressive disorder, early-stage cancers) have the ability to live independently or even work, causing ordinary citizens to rank these diseases low on health loss since compared to nurses, who have extensive knowledge on the effects of these diseases across a wide population and over the long-term. This suggests that disability weight estimations based on surveys of skilled nurses may correct the underestimations of early- and middle-stage diseases resulting from surveys of ordinary citizens. Further, the results suggest that when the government measures the population disability-adjusted life years, more attention should be paid to non-terminal patients.

Second, the disability weight of early-stage cancers (e.g., lung or colorectal cancer) and depressive disorder are almost the same, which indicates that the severity of health loss from depressive disorder is the same as that of early-stage cancers. This comparison with cancers suggests that the patients with disabilities emanating from depressive disorder require more attention or aid from health organizations or the public.

Here, we summarize the policy implications based on the estimation results. First, our results show that the disability weight of early-stage of cancer is similar to the weight of depression, which illustrates related loss of health from acuter depression as well as cancer. Therefore, when the number of patients with acute depression increase, policies guiding hospitals and regions to provide more aid and attention to psychological well-being should be implemented. Second, the disability weight of non-terminal diseases evaluated by skilled nurses is higher than the weight from the general public. It indicates that the government may over-evaluate the population health status, such as disability-adjusted life years. We recommend evaluating the status of population health based on the disability weight from experienced medical professionals.

This study has two limitations. First, in this survey 21% responding rate is recorded, which may have caused a sample selection bias in estimating disability weight. To address this potential sample selection bias, future research should either increase the response rate or collect detailed data including the demographic or socioeconomic background of all targeted nurses. Second, because of the data limitation, the disability weights of seventeen representative diseases are estimated. This may not comprehensively cover the overall population health status (e.g., deaf). Future studies should try to evaluate more diseases to broaden the knowledge on disability weights.

Declaration of competing interest

The authors declare that there are no known competing financial or personal interests that could have appeared to influence the content of this article.

This study was approved by the Research Ethics Committee of the Graduate School of Medicine, The University of Tokyo (No. 2018068NI).

Acknowledgement: This work was supported by JSPS KAKENHI Grant Number JP20H00648; This research was supported by the 4th Environmental Economics Research Fund of the Ministry of the Environment, Japan. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the agencies. We thank Hiromi Sanada at University of Tokyo for her support. We thank the editor and the reviewers for your thoughtful suggestions and insights.

References

American Nurses Association. (2015). Nursing: Scope and Standards of Practice, 3rd Ed.

Basiri, A., Mousavi, S. M., Naghavi, M., Araghi, I. A., & Namini, S. A. (2008). Urologic diseases in the Islamic Republic of Iran: What are the public health priorities? *Eastern Mediterranean health journal*, *14*(6), 1338–1348.

Baltussen, R. M. P. M., Sanon, M., Sommerfeld, J., & Würthwein, R. (2002). Obtaining disability weights in rural Burkina Faso using a culturally adapted visual analogue scale. *Health economics*, *11*(2), 155–163. https://doi.org/10.1002/hec.658

Brennan, D. S., Spencer, A. J., & Roberts-Thomson, K. F. (2007). Quality of life and disability weights associated with periodontal disease. *Journal of dental research*, *86*(8), 713–717. https://doi.org/10.1177/154405910708600805

Haagsma, J. A., Polinder, S., Cassini, A., Colzani, E., & Havelaar, A. H. (2014). Review of disability weight studies: Comparison of methodological choices and values. *Population health metrics*, *12*(1), 20. https://doi.org/10.1186/s12963-014-0020-2

Hong, K-S., & Saver, J. L. (2009). Quantifying the value of stroke disability outcomes: WHO global burden of disease project disability weights for each level of the modified Rankin Scale. *Stroke*, *40*(12), 3828–3833. https://doi.org/10.1161/STROKEAHA.109.561365

Jelsma, J., Chivaura, V. G., Mhundwa, K., De Weerdt, W., & De Cock, P. (2000). The global burden of disease disability weights. *The Lancet*, *355*(9220), 2079–2080. https://doi.org/10.1016/S0140-6736(05)73538-8

Kaji, R. (2019). Global burden of neurological diseases highlights stroke. *Nature Reviews Neurology*, *15*(7), 371–372. https://doi.org/10.1038/s41582-019-0208-y

Lai, T., Habicht, J., & Kiivet, R. A. (2009). Measuring burden of disease in Estonia to support public health policy. *The European Journal of Public Health*, *19*(5), 541–547. https://doi.org/10.1093/eurpub/ckp038

Lyons, R. A., Kendrick, D., Towner, E. M., Christie, N., Macey, S., Coupland, C., Gabbe, B. (2011). Measuring the population burden of injuries—implications for global and national estimates: A multi-centre prospective UK longitudinal study. *PLoS Med*, *8*(12 Art. No. e1001140), 1–13. https://doi.org/10.1371/journal.pmed.1001140

Murray, C. J., Lopez, A. D., & World Health Organization. (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries,* *and risk factors in 1990 and projected to 2020: Summary*. World Health Organization.

Murray, C. J., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., ... & Lopez, A. D. (2012). GBD 2010: Design, definitions, and metrics. *The Lancet*, *380*(9859), 2063–2066. https://doi.org/10.1016/S0140-6736(12)61899-6.

Nomura, S., Sakamoto, H., Glenn, S., Tsugawa, Y., Abe, S. K., Rahman, M. M., ... & Shibuya, K. (2017). Population health and regional variations of disease burden in Japan, 1990–2015: A systematic subnational analysis for the Global Burden of Disease Study 2015. *The Lancet*, *390*(10101), 1521–1538. https://doi.org/10.1016/S0140-6736(17)31544-1

Nord, E. (1995). The person-trade-off approach to valuing health care programs. *Medical decision making*, *15*(3), 201-208. https://doi.org/10.1177/0272989X9501500302

Nord, E. (2013). Disability weights in the Global Burden of Disease 2010: Unclear meaning and overstatement of international agreement. *Health policy*, *111*(1), 99–104. https://doi.org/10.1016/j.healthpol.2013.03.019

Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., ... & Vos, T. (2015). Disability weights for the Global Burden of Disease 2013 study. *The Lancet Global Health*, *3*(11), e712–e723. https://doi.org/10.1016/S2214-109X(15)00069-8

Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A., ... & Murray, C. J. L. (2012). Common values in assessing health outcomes from disease and injury: Disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2129–2143. https://doi.org/10.1016/S0140-6736(12)61680-8

Schwarzinger, M., Stouthard, M. E., Burström, K., & Nord, E. (2003). Cross-national agreement on disability weights: the European Disability Weights Project. *Population Health Metrics*, *1*(1), 1-9. https://doi.org/10.1186/1478-7954-1-9

Taylor, H. R., Jonas, J. B., Keeffe, J., Leasher, J., Naidoo, K., Pesudovs, K., & Resnikoff, S. (2013). Disability weights for vision disorders in Global Burden of Disease study. *The Lancet*, *381*(9860), 23. https://doi.org/10.1016/S0140-6736(12)62081-9

Whiteford, H., Ferrari, A., & Degenhardt, L. (2016). Global burden of disease studies: Implications for mental and substance use disorders. *Health Affairs*, *35*(6), 1114–1120. https://doi.org/10.1377/hlthaff.2016.0082.

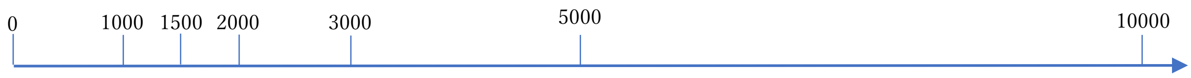
Yoon, S-J., Bae, S-C., Lee, S-I., Chang, H., Jo, H. S., Sung, J-H., ... & Shin, Y. (2007). Measuring the burden of disease in Korea. *Journal of Korean medical science*, *22*(3), 518–523. https://doi.org/10.3346/jkms.2007.22.3.518

van Spijker, B. A., van Straten, A., Kerkhof, A. J., Hoeymans, N., & Smit, F. (2011). Disability weights for suicidal thoughts and non-fatal suicide attempts. *Journal of affective disorders*, *134*(1–3), 341–347. https://doi.org/10.1016/j.jad.2011.05.020

Appendix

The disability weights of the diseases for each participant were computed in pair comparisons (PC) and population health equivalence measures (PHE). In the case of the PC, we applied the same probit regression that was utilized in the analysis of discrete choice experimental data in Salomon et al. (2012). We regarded the stated choice between two diseases in a PC as the dependent variable. Furthermore, we treated the 17 diseases as independent variables and created them as dummy variables. We computed the predicted probabilities of each disease from the coefficient estimates of the probit regression.

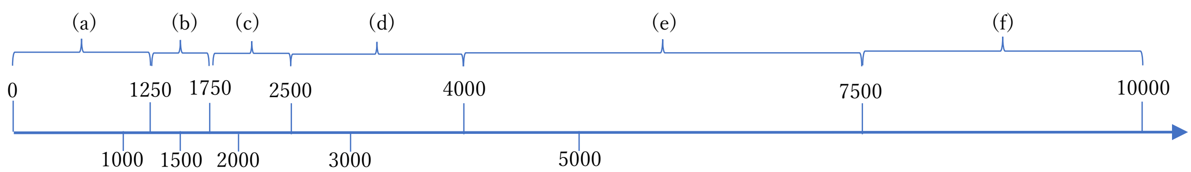
The PHE questions provide information on tradeoffs between mortality and nonfatal outcomes, which is needed to anchor the results from the probit regression analysis of paired comparisons on the (0,1) disability weight scale. Because the population health equivalence questions are framed in terms of a binary comparison between an intervention that averted 1000 fatalities and another intervention that averted some number of nonfatal outcomes, and due to paper-based questions rather than web-based ones, the responses provide interval-censored information rather than exact values. Therefore, we changed the analysis method of population health equivalence from Salomon et al. (2012). First, we presented the scenario as "Program A, which prevented 1000 people from getting an illness that causes rapid death, has the same population health benefit as___", then the respondents were asked to select either "(1) less than 1000; (2) 1500; (3) 2000; (4) 3000; (5) 5000; and (6) more than 10000" people from getting each disease. We draw the number of populations on the number line as follows:



Next, we presented that Program B prevented people from being afflicted with a certain disease. Here, we correspond each answer in the following manner: the participants answer (B) for "less than 1250 persons" if they answered "less than 1000"; they answer (B) for "between 1250 persons and 1750 persons" when they answered 1500. Based on the formula, we make each response fit the PHE number. The list is as follows:

Survey response	Program A as quasi-choice	Program B as quasi-choice	Interval
Less than 1000	None	1250	(a)
1500	1250	1750	(b)
2000	1750	2500	(c)
3000	2500	4000	(d)
5000	4000	7500	(e)
More than 10000	7500	none	(f)

The key idea is that we want to capture the questionees' interval" opinions; in other words, our question about equivalent or approximate numbers reveal their interval of population health equivalence. In detail, we take the midpoint of each interval and show them as follows:



In the PHE measure, the disability weights of the diseases for each participant were estimated using the following formula: 1,000/final number of people in program B. For example, if a respondent indicates that Program (A) that averted 1000 fatalities produced equivalent or the closest health benefit to a Program (B) that averted 3000 cases of chronic kidney disease, then this response is taken to indicate that the respondent attaches a disability weight to chronic kidney disease that lies somewhere in the interval between 0.25 and 0.4.

In sum, the table below shows each response corresponding to the quasi-choices:

Survey response	Upper bound of the PHE-interval	Lower bound of the PHE-interval	Logit-transformed value for upper-bound	Logit-transformed value for lower-bound
Less than 1000	none	$\frac{4}{5}$	none	1.38629436112
1500	$\frac{4}{5}$	$\frac{4}{7}$	1.38629436112	0.287682072452
2000	$\frac{4}{7}$	$\frac{2}{5}$	0.287682072452	-0.405465108108
3000	$\frac{2}{5}$	$\frac{1}{4}$	-0.405465108108	-1.098612288668
5000	$\frac{1}{4}$	$\frac{2}{15}$	-1.098612288668	-1.871802176902
More than 10000	$\frac{2}{15}$	none	-1.871802176902	none

After the PC and PHE statistical analyses, we rescale the probit coefficients obtained by PC since the probit regression of PC responses yields values that are on an arbitrary scale with constant variance across health states. For this, we completed two steps. The first step was to run a linear regression of the probit coefficients from the pooled analysis of the disability weight estimates (these were derived from a maximum likelihood estimation of the PHE responses). The regression resulted in an estimated slope and intercept for a linear transformation of the probit coefficients onto the (logit) disability-weight scale. The second step was to use numerical integration to obtain mean estimates of disability weights on the natural 0-to-1scale. First, we simulated normal random variates onto the logit scale with means defined by the rescaled probit coefficients and variance defined by the overall standard deviation of these coefficients across survey-specific estimates. Next, we transformed each of these simulated values through an inverse-logit function. Finally, we computed the mean across the resulting values for each health state.