



Eliciting national and subnational sets of disability weights in mainland China: Findings from the Chinese disability weight measurement study

Xiaoxue Liu,^a Fang Wang,^b Maigeng Zhou,^{d,*} Yong Yu,^e Jinlei Qi,^d Peng Yin,^d Shicheng Yu,^f Yuchang Zhou,^g Lin Lin,^d Yunying Liu,^d Qiqi Wang,^f Wenling Zhong,^h Shaofen Huang,^h Yanxia Li,ⁱ Li Liu,ⁱ Yuan Liu,^j Fang Ma,^k Yine Zhang,^k Yuan Tian,^k Qiuli Yu,^l Jing Zeng,^m Jingju Pan,ⁿ Mengge Zhou,ⁿ Weiwei Kang,^o Jin-Yi Zhou,^p Hao Yu,^p Yuehua Liu,^q Shaofang Li,^r Huiting Yu,^s Chunfang Wang,^s Tian Xia,^s Jinen Xi,^t Xiaolan Ren,^t Xiuya Xing,^u Qianyao Cheng,^u Fangrong Fei,^v Dezheng Wang,^w Shuang Zhang,^w Yuling He,^x Haoyu Wen,^a Yan Liu,^a Fang Shi,^a Yafeng Wang,^a Panglin Sun,^a Jianjun Bai,^a Xuyan Wang,^y Hui Shen,^a Yudiyang Ma,^a Donghui Yang,^a Sumaira Mubarik,^a Jinhong Cao,^a Runtang Meng,^z Yunquan Zhang,^{aa} Yan Guo,^{bb} Yaqiong Yan,^{bb} Wei Zhang,^{bb} Sisi Ke,^{bb} Runhua Zhang,^{cc} Dingyi Wang,^{dd} Tingting Zhang,^{ee} Shuhe Nomura,^{ff,gg} Simon I. Hay,^{hh} Joshua A. Salomon,ⁱⁱ Juanita A. Haagsma,^{jj} Christopher J.L. Murray,^{hh} Theo Vos,^{hh} and Chuanhua Yu^{a,c,*}

^aDepartment of Epidemiology and Biostatistics, School of Public Health, Wuhan University, No. 115 Donghu Road, Wuchang District, Wuhan 430071, Hubei Province, China

^bDepartment of Epidemiology and Biostatistics, School of Public Health, Xuzhou Medical University, Xuzhou 221004, China

^cGlobal Health Institute, Wuhan University, Wuhan 430072, China

^dNational Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, No. 27 Nanwei Road, Xicheng District, Beijing 100050, China

^eSchool of Public Health and Management, Hubei University of Medicine, Shiyan 442000, Hubei, China

^fChinese Center for Disease Control and Prevention, Beijing, China

^gDepartment of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong 250012, China

^hFujian Provincial Center for Disease Control and Prevention, No. 78 Jintai Road, Gulou District, Fuzhou City 350001, Fujian province, China

ⁱLiaoning Provincial Center for Disease Control and Prevention, No. 79 Jixian Street, Heping District, Shenyang City 110005, China

^jHunan Provincial Center for Disease Control and Prevention, No. 450 first section of Middle Furong Road, Changsha City 410005, Hunan Province, China

^kNingxia Center for Disease Control and Prevention, No. 528 Shengli Street, Xingqing District, Yinchuan City 750004, Ningxia, China

^lYunnan Center for Disease Control and Prevention, No. 158 Dongsu Street, Xishan District, Kunming City 650022, Yunnan Province, China

^mSichuan Center for Disease Control and Prevention, No. 6 Middle School Road, Wuhou District, Chengdu City 610041, Sichuan Province, China

ⁿHubei Provincial Center for Disease Control and Prevention, No. 6 Zhuodaoquan North Road, Hongshan District, Wuhan City 430079, Hubei Province, China

^oInner Mongolia Integrative Center for Disease Control and Prevention, No. 50 Ordos Street, Hohhot 010031, China

^pJiangsu Provincial Center for disease Control and Prevention, Public Health Research Institute of Jiangsu Province, Jiangsu Road No. 172, Gulou District, Nanjing city 210009, Jiangsu Province, China

^qHeilongjiang Provincial Center for Disease Control and Prevention, No. 40 Youfang Street, Xiangfang District, Harbin City 150030, China

^rHenan Provincial Center for Disease Control and Prevention, No. 105 Nongye South Street, Zhengdong New District, Zhengzhou City 450016, China

^sShanghai Municipal Center for Disease Control and Prevention, No. 1380 Zhongshan West Street, Changning District, Shanghai City 200051, China

^tGansu Provincial Center for Disease Control and Prevention, No. 230 Donggang West Street, Chengguan District, Lanzhou City 73000, China

^uAnhui Provincial Center for Disease Control and Prevention, No. 12560 Fanhua Avenue, Economic and Technological Development District, Hefei City 230601, China

^vZhejiang Provincial Center for Disease Control and Prevention, No. 3399 Binsheng Street, Binjiang District, Hangzhou City 310051, China

^wTianjin Centers for Disease Control and Prevention, No. 6 Huayue Street, Hedong District, Tianjin City 300011, China

*Corresponding authors.

E-mail addresses: yuchua@whu.edu.cn (C. Yu), zhoumaigeng@ncncd.chinacdc.cn (M. Zhou).

^xShanxi Center for Disease Control and Prevention, No. 6 Xiaonanguan Shuangta West Street, Yingze District, Taiyuan City 030012, China

^yUnion Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

^zDepartment of Preventive Medicine, School of Medicine, Hangzhou Normal University, Hangzhou 311121, Zhejiang, China

^{aa}Hubei Province Key Laboratory of Occupational Hazard Identification and Control, Wuhan University of Science and Technology, Wuhan 430065, China

^{bb}Wuhan Centers for Disease Control and Prevention, Wuhan 430024, Hubei, China

^{cc}Beijing Tiantan Hospital, Capital Medical University Beijing, China

^{dd}Institute of Clinical Medical Sciences, China-Japan Friendship Hospital; National Center for Respiratory Medicine, Beijing, China; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing, China

^{ee}Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health (Peking University), National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Beijing 100083, China

^{ff}Department of Health Policy and Management, School of Medicine, Keio University, Tokyo, Japan

^{gg}Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Japan

^{hh}Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA

ⁱⁱDepartment of Medicine, School of Medicine, Stanford University, Stanford, CA, USA

^{jj}Department of Public Health, Erasmus Medical Center, Rotterdam, Netherlands

Summary

Background The disability weight (DW) quantifies the severity of health states from disease sequela and is a pivotal parameter for disease burden calculation. We conducted a national and subnational DW measurement in China.

Methods In 2020–2021, we conducted a web-based survey to assess DWs for 206 health states in 31 Chinese provinces targeting health workers via professional networks. We fielded questions of paired comparison (PC) and population health equivalence (PHE). The PC data were analysed by probit regression analysis, and the regression results were anchored by results from the PHE responses on the DW scale between 0 (no loss of health) and 1 (health loss equivalent to death).

Findings We used PC responses from 468,541 respondents to estimate DWs of health states. Eight of 11 domains of health had significantly negative coefficients in the regression of the difference between Chinese and Global Burden of Disease (GBD) DWs, suggesting lower DW values for health states with mention of these domains in their lay description. We noted considerable heterogeneity within domains, however. After applying these Chinese DWs to the 2019 GBD estimates for China, total years lived with disability (YLDs) increased by 14.9% to 177 million despite lower estimates for musculoskeletal disorders, cardiovascular diseases, mental disorders, diabetes and chronic kidney disease. The lower estimates of YLDs for these conditions were more than offset by higher estimates of common, low-severity conditions.

Interpretation The differences between the GBD and Chinese DWs suggest that there might be some contextual factors influencing the valuation of health states. While the reduced estimates for mental disorders, alcohol use disorder, and dementia could hint at a culturally different valuation of these conditions in China, the much greater shifts in YLDs from low-severity conditions more likely reflects methodological difficulty to distinguish between health states that vary a little in absolute DW value but a lot in relative terms.

Funding This work was supported by the National Natural Science Foundation of China [grant number 82173626], the National Key Research and Development Program of China [grant numbers 2018YFC1315302], Wuhan Medical Research Program of Joint Fund of Hubei Health Committee [grant number WJ2019H304], and Ningxia Natural Science Foundation Project [grant number 2020AAC03436].

Copyright © 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Disability weight; Severity; Health state; Paired comparison; Population health equivalence; Probit regression model

The Lancet Regional Health - Western Pacific 2022;26: 100520

Published online 26 July 2022

<https://doi.org/10.1016/j.lanwpc.2022.100520>

Research in context

Evidence before this study

The Global Burden of Disease (GBD) study for its 2010 update moved away from disability weights (DW) set by health experts to valuations conducted in population and internet surveys. The GBD DW set is based on pairwise comparison (PC) data from surveys in nine countries (Bangladesh, Peru, Indonesia, Tanzania, United States, the Netherlands, Sweden, Hungary and Italy) and an open access internet survey. Population health equivalence (PHE) questions which were used to anchor the PC results on a zero to one scale, were fielded in the internet survey and the European survey but failed to produce interpretable results in the latter. The most recent update of GBD DWs was incorporated into GBD 2013 and subsequent iterations. It was based on health state valuation data of 60,890 persons. There were almost no data included from East Asia, however. Moreover, a recent Japanese study indicated there might be contextual differences in assessing the severity of health states compared to the GBD surveys. The low proportion of respondents from China contributing to the GBD surveys motivated this work to derive valuations for a sample of the Chinese population.

Added value of this study

The GBD 2013 study pointed out that future local survey data are needed to advance DW research as it is limited by lack of geographic representation in the data.¹ The current analysis created a new set of DWs based on 468,541 respondents covering 31 provinces in mainland China and added a comparison of Chinese DW compared with the GBD 2013 DWs and Japanese DWs to explore contextual differences in valuations of health across countries. We found high correlations of probit coefficients of the PC questions between provinces and by characteristics such as age, gender, profession, educational level and income, suggesting that despite the purposive sampling strategy these valuations broadly reflect opinions of the general public in China. We did find that the cognitively more demanding PHE questions were better suited to respondents with higher education.

Implications of all the available evidence

This study elicited national and subnational DWs for diseases and injuries based on methods from the GBD. In common with a recent Japanese study we found lower valuations of the severity of mental disorders, alcohol use disorders and dementia compared to the GBD DW values that were generated from surveys with low participation from East Asia. However, there were also considerable differences with the Japanese study that are less suggestive of a common East Asian pattern. For instance, while the Japanese study found higher DWs for health states with pain, in this study, pain was associated with generally lower DW but a mixed impact on osteoarthritis knee [much

higher years lived with disability (YLDs)], neck pain (much lower YLDs) or low back pain (similar YLDs). In fact, by far the largest difference between Chinese and GBD DWs concerned highly prevalent but low-severity conditions suggesting that further work needs to be done to explore methodological factors that may explain the large proportional difference in the DWs for these conditions. We propose adding the Japanese and Chinese data to a common analysis with all previous DW surveys.

Introduction

In burden of disease studies, the disability weight (DW) aims to consistently enumerate the severity of health states across all diseases and injuries. The DW is measured on a scale from 0 to 1, with 0 indicating a state that is equivalent to full health and 1 a state equivalent to death.^{2,3} Using DW results, the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) provides a systematic scientific assessment of published, publicly available, and contributed data to estimate years lived with disability (YLDs) for a collectively exhaustive list of diseases and injuries for every year.^{2,4–9} Combining YLDs with years of life lost (YLLs) due to premature mortality, the disability-adjusted life years (DALYs) metric, allows comparisons between the population health impact of death and the severity of disabling disease and injury outcomes.^{2,4,8,10}

The GBD 2010 study elicited DWs for 220 health states⁹ in five country surveys and an open-access internet survey and expanded to evaluating 235 unique health states for GBD 2013 and subsequent GBD cycles, including GBD 2019 by adding results from a European DW measurement survey in four countries.^{2,3} These GBD DWs have been criticized for overstating the between-country agreement.^{11,12} Previous studies indicated contextual and intercultural differences should be taken into account for DW measurement.^{13–16} The internet DW survey for GBD 2010 sampled 16,328 respondents around the world but under-representing respondents from China. Some studies suggested that the DW in East Asia may differ from that of Western countries.^{15,16} In 2019, a national web-based survey conducted in Japan with 37,318 respondents used the methods of the GBD DW measurement study and estimated DWs for 231 health states.¹⁴ Several studies have demonstrated that factors such as nationality, age, sex, disease experience status, income and educational levels^{17–21} or profession²² may have an impact on valuations of health. However, over the past decades, only a European study¹⁷ evaluated the implications of such factors on the valuation of health state severity, and suggested a more systematic approach in future studies, also in non-European countries. Considering potential contextual differences, a rigorous assessment of the severity of health

states relevant to the Chinese population is important for the further development of disease burden research.

In this study, we aimed to elicit national DWs for mainland China. Based on the resulting estimates of DWs, we compared these with previous DW surveys, and evaluated the effect of participant characteristics and province of residence on their estimation. This was with a view to increasing the fidelity of disease burden estimation in China and improving the representativeness of the DWs use in the GBD study for future iterations of the GBD study.

Methods

Study design and participants

We conducted a web-based survey between May 12 and July 22, 2020 using the same pair-wise comparison (PC) and population health equivalence (PHE) methods as were used by the GBD² and European surveys.³ Each PC question presented a random pair of health states with a brief description of the main features in lay language and asked respondents which person they think is the healthier. The PHE questions asked respondents to compare the health benefits of two hypothetical life-saving or health-improving programs and choose which health program they thought produced the greater overall population health benefit. The PHE questions were phrased as follows: “the first health program prevented 1,000 people from getting an illness that causes rapid death; the second health program prevented x people from getting an illness that is nonfatal but causes the lifelong health problems of y ”, where x is a randomly assigned ‘bid’ value of 1500, 2000, 3000, 5000, or 10,000 and y a lay description of a randomly chosen health state. In this study, 206 health states were evaluated using the PC questions, and a subset of 28 states were evaluated by means of PHE questions based on GBD 2010 and European DW study.^{2,3} A second survey round, administered to a subset of participants who completed survey 1, was conducted from September 23 to November 12, 2021, to collect further PHE data as the answers to the PHE questions in the first round after analysis lacked a signal of discrimination by severity.

We retained the answers provided by resident participants aged 18–69 years. People aged over 70 were excluded because they were expected to be less familiar with the internet and find the survey too difficult. We recruited participants through professional networks of the study investigators and staff from the Center for Disease Control and Prevention (CDC) offices in the 31 provinces. The sampling methods as aiming to reach motivated respondents to improve the quality of answer. We established a national DW investigation group consisting of staff from these CDC offices. Each province’s CDC staff deployed surveys in each city, and then each city’s CDC staff deployed surveys in each district or town. At each investigation site, potential

respondents received a link (<https://mp.weixin.qq.com/s/ixc3gTC9ocDQ7o35x2jWFg>) to the questionnaire via a personal WeChat message (Company URL: <https://weixin.qq.com/>). The responses to the questionnaire were transmitted to a registered cloud data management storage server for quality control and real-time information feedback. The average time for finishing the survey was 14 min. We also announced the web-based survey on relevant websites, and allowed participants to recruit others via word of mouth.

Data quality control – inclusion criteria

In the initial survey round, we used the questionnaire with 16 PC and three PHE questions, and the questionnaire with three PC and four PHE questions in the second round of survey. In the first round of survey, we assigned the same pair of health states in the third, 10th and 16th PC questions to allow assessment of test-retest reliability. In the additional survey round we only used the questionnaire with three PC and four PHE questions. We repeated the same health state with the same bid in the first and second PHE questions to allow assessment of internal consistency of PHE responses. We also added two questions asking respondents to state if they remembered the repeat PHE question. In order to improve data quality, a series of measures were set up for quality control: (1) allowing a user to answer once only; (2) requiring a minimum survey completion time of 3 minutes; (3) excluding answers by persons who did not consistently answer the repeated PC or PHE questions; (4) excluding answers to the 16 PC questions are all A or B, and all answers alternating A and B. Each respondent was randomly assigned to receive a reward ranging from US\$ 0.30 to 15. This study was approved by the Ethics Committee of Medical Department of Wuhan University (2019YF2055), and a waiver of written informed consent obtained from participants prior to web-based survey participation was approved.

Health states and lay description

Both the PC method and the PHE method rely on a description of the functional loss or symptoms that characterize a health state using simple, non-clinical vocabulary. We included 206 health states: 172 from the original GBD 2010 DW study, 32 from the European DW study³, and two health states with slightly modified lay descriptions of original GBD health states (moderate and severe hearing loss) on the advice of disease experts. Native speakers with a medical background translated the health states and lay descriptions from the GBD and European DW studies into Chinese. Subsequently, back translation was verified independently by bilingual native speakers. The health states and lay descriptions in English (pp 7–34) and Chinese are presented in full in the appendix (pp 35–48).

Statistical analysis

Using the same method as in the GBD surveys, we ran probit regression analyses on the choice responses in PC data.² The probit regression yields estimated results for each health state that captures its relative severity compared to all other health states. A second analytic step is needed to anchor the resulting estimates of the probit regression on the 0-1 DW scale. We ran an interval, linear regression on the probit coefficients of the disability weight estimates implied by the PHE responses: first, we obtained the logit-transformed disability weight estimates derived from interval regression of the population health equivalence responses; second, we conducted a linear regression of the probit coefficients from analysis of paired comparisons on the resulting estimates from interval regression of PHE. We then estimated mean values for DW using numerical integration. A bootstrapping approach with 1,000 replicate samples was used to estimate 95% uncertainty intervals (UI) around the mean disability weights. More detailed methods are provided in the appendix and publications of the GBD and European surveys.^{2,3} We performed all statistical analyses with R (version 4.0.2) and Stata/MP (version 15). We evaluated variation in probit regression coefficients by age, sex, disease experience status, educational and income level, medical background, profession and province, using the Pearson correlation coefficient.

We compared the estimated Chinese DWs for 206 comparable health states with the GBD 2013 and Japanese DWs to assess what symptoms mentioned in the lay descriptions of health state were associated with the difference in the DWs. Eleven identified symptom categories based on the lay descriptions are presented in the appendix (pp 7–34). We performed a linear regression of the proportional differences between DWs in the Chinese, Japanese and GBD 2013 studies against these 11 domains of health loss. A *p*-value of < 0.05 was considered statistically significant.

Finally, we recomputed the YLD estimates for China from the GBD 2019 study using the Chinese DWs from this study and computed the difference in number of YLDs and the ranking of the top-25 conditions against those of GBD 2019. After finding the largest differences between YLD results using Chinese or GBD DWs was for low severity health states we also computed a version using Chinese DWs but replacing any health state with a GBD DW less than 0.03 with the GBD value.

Role of the funding source

The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Characteristics of respondents

Table 1 shows the characteristics of the respondents who did and did not meet quality criteria in this study. A total of 1,087,319 respondents participated in the first survey round, and 1,025,930 (94.4%) completed the questionnaires. The completion rate of questionnaire was 94.4%. 468,541 respondents met all inclusion criteria, and the valid rate of questionnaire was 43.1%. As it was conducted on a convenience sample, we could not estimate a response rate.

Compared to the distribution of Chinese general population, those who participated in the web-based survey tended to be younger, more likely to be female, and have a high level of education. The percentage of college graduates and above among respondents (67%) was much larger than the whole population (19%). Most (66%) of participants had medium and high-income level, and 62% people had a non-manual profession. The information of the participants' socio-demographic characteristics from other countries is provided in the appendix (pp 49). The geographical distribution of participants by province are presented in the appendix (p74).

Paired comparison and population health equivalence responses

There were 1,087,319 participants in the first survey round and an additional 17,593 in the second round. Of 638,431 participants in the first survey who met all inclusion criteria, 468,541 (73%) of whom provided consistent answers to the test-retest PC questions and were included in the probit analysis.

Overall consistency was examined with heat maps in which the horizontal axis represents the first health state of a pair, and the vertical axis represents the second. Each cell in the heat map indicates the response probability for one pair of health states. The colours of the heat map correspond to the probability that the first health state in a PC is chosen as the healthier outcome. A smooth transition in colors from high to low probabilities between the upper left and lower right corner was observed, indicating high internal consistency (Figure 1). In comparison, the heatmap of respondents who gave inconsistent test re-test answers shows diminished internal consistency. Heat maps for 13 provinces with the greatest sample size are provided in the Appendix (p75).

In the initial survey round 272,206 respondents were given the PHE questions. However, discrimination between health states was poor (Appendix p79). We conducted a second survey round with 17,593 respondents 4925 of whom answered all PHE questions; consistently answered the repeat question; and correctly identified the repeat question when asked after completion of the questionnaire. The answers by these 4925

	Chinese population (%) ^a	PC responses			PHE responses		
		Respondents with eligible PC answers (n = 468,541)	Respondents who failed on test re-test PC questions (n=169,890) ^b	Other excluded respondents (n = 448,888) ^b	Respondents with eligible PHE answers (n = 4,925)	Respondents who failed on repeated PHE questions (n = 3,575) ^c	Respondents who did not correctly identify repeated questions at end of PHE questions (n = 5,085) ^c
Age (years)							
18–29	13.8	112,007 (23.9%)	58,544 (34.5%)	207,413 (46.2%)	1492 (30.3%)	1785 (49.9%)	1337 (26.3%)
30–49	30.5	287,577 (61.4%)	74,780 (44.1%)	133,182 (29.7%)	2807 (57.0%)	969 (27.1%)	2925 (57.5%)
50–69	26.2	68,957 (14.7%)	36,416 (21.4%)	44,459 (9.9%)	605 (12.3%)	114 (3.2%)	800 (15.7%)
≥70	8.3	0	0	2822 (0.6%)	20 (0.4%)	12 (0.3%)	23 (0.5%)
Sex							
Men	51.2	156,907 (33.5%)	58,596 (34.5%)	177,819 (39.6%)	1325 (26.9%)	2085 (58.3%)	1569 (30.9%)
Women	48.8	311,634 (66.6%)	111,294 (65.5%)	271,069 (60.4%)	3600 (73.1%)	1490 (41.7%)	3516 (69.2%)
Education level							
None	3.6	0	2 (<1%)	3,213 (0.7%)	2 (<1%)	57 (1.6%)	2 (<1%)
Primary	24.6	0	32 (<1%)	46,022 (10.3%)	26 (0.5%)	142 (4.0%)	29 (0.6%)
Secondary	49.6	152,895 (32.6%)	60,963 (35.9%)	176,364 (39.3%)	543 (11.0%)	861 (24.1%)	921 (18.1%)
Higher	15.4	315,646 (67.4%)	108,892(64.1%)	223,289 (49.7%)	4354 (88.4%)	2515 (70.3%)	4133 (81.3%)
Income level ^d							
Low	NA	157,593 (33.6%)	68,082 (40.1%)	196,867 (43.8%)	1128 (22.9%)	1022 (28.6%)	1435 (28.2%)
Medium		275,773 (58.9%)	91,793 (54.0%)	217,637 (48.5%)	3211 (65.2%)	2027 (56.7%)	3188 (62.6%)
High		35,180 (7.5%)	10,015 (5.9%)	34,384 (7.7%)	586 (11.9%)	526 (14.7%)	462 (9.1%)
Profession							
Profession-Non-manual							
Non-manual	NA	291,303 (62.2%)	103,932 (61.2%)	167,463 (37.3%)	3315 (67.3%)	1430 (40.0%)	3297 (64.8%)
Manual		89,349 (19.1%)	33,025 (19.4%)	107,526 (24.0%)	768 (15.6%)	1410 (39.4%)	926 (18.2%)
Others ^e		87,889 (18.8%)	32,933 (19.4%)	173,899 (38.7%)	842 (17.1%)	735 (20.6%)	826 (17.0%)

Table 1: Characteristics of the study population.

a: According to China Statistical Yearbook compiled by National Bureau of Statistics of China: <http://www-stats-gov-cn.proxy.www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm> (2020 population).

b: These participants (n=169,890) failed on test re-test PC questions and other respondents didn't meet the rest of other quality control criteria (n=448,888), both of which would be excluded from the total sample in the first survey round.

c: These participants (n=3,575) failed on repeated PHE questions and these respondents (n=5,085) did not correctly identify the recalling questions at end of PHE question, both of them would be excluded from PHE respondents in the second survey round.

d: The income level is divided into 3 groups: high (>US\$ 30,769), middle (US\$ 7,692-30,769) and low (<US\$ 7,692). We revised the annual household income from China won to US\$ at the exchange rate of 6.5 RMB won to the US\$.

e: Other occupation including housewife and student, soldier, out of work, retired, and other laborers.

NA: not available in China Statistical Yearbook; NR: not reported; GBD: Global Burden of Disease.

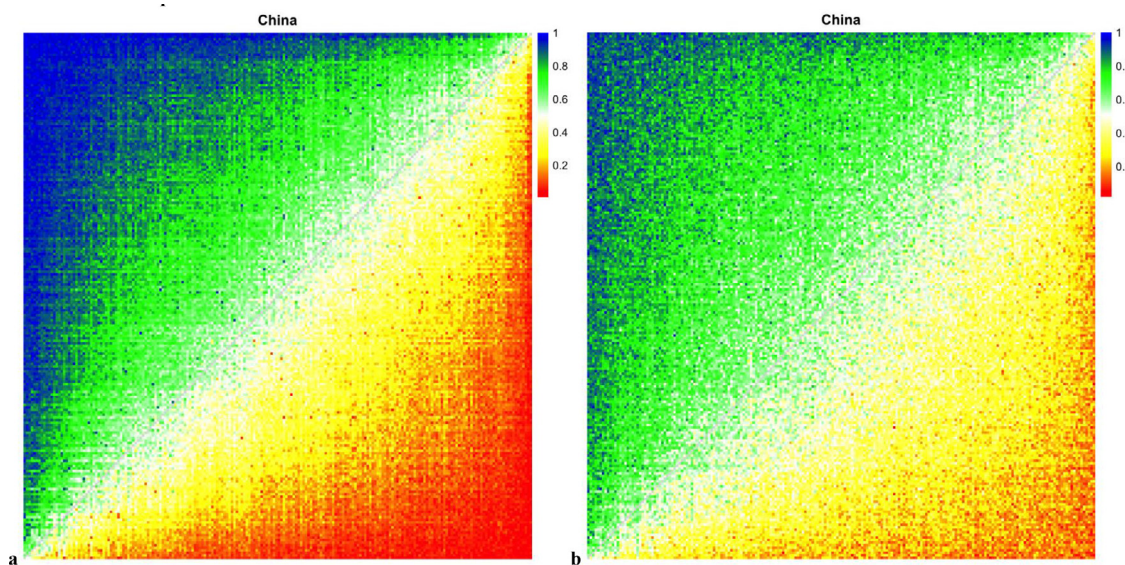


Figure 1. Response probabilities for paired comparisons in Chinese web survey for respondents included in analysis and respondents who gave inconsistent answers to the test re-test questions. Colors on the heat maps correspond to the probability that the first health state in a paired comparison was chosen as the healthier outcome. Variation in the amount of measurement error across surveys is reflected in the varying degrees to which response probabilities follow an orderly transition from high to low between the upper left and the lower right corners in each heat map. A heat map with no measurement error and high internal consistency indicated by a smooth color transition from blue to red along the diagonal, whereas a heat map with a high amount of error would have a random assortment of colored squares. a. Heatmap of PC answers by 468,541 respondents who consistently answered the test re-test questions; b. Heatmap of PC answers by 169,890 respondents who consistently answered the test re-test questions inconsistent answers.

respondents were included in the interval analysis for PHE data and used to rescale DWs from PC data. [Figure 2](#) shows the probabilities of PHE responses when choosing the second program at each bid value for the 28 selected health states. We ranked the 28 health states used in PHE questions based on the probabilities of choosing the second program as the one that prevented a specified number of cases of each health outcome. The five different bids are shown from left to right for each of the 28 health states. The y-axis shows the probability of choosing the second health program. As the bid increased from 1,500 to 10,000, the probability of choosing the second health program was generally higher. There also was discrimination by health state (general responsiveness to variation in the severity of the different outcomes). The Chinese PHE responses had a more noticeable increasing trend in probabilities with the severity of health states than observed in Japan but less than observed in the GBD 2010 web survey.^{3,14}

Disability weights

[Table 2](#) shows the estimated DW and 95% UI for 206 health states. Distance vision mild impairment and mild anemia shared the lowest DW (0.009) and severe heroin and other opioid dependence had the highest

DW (0.752). All correlation of the probit coefficients was high between age groups (range Pearson correlation r_s : 0.992–0.998), educational levels (r_s : 0.991), income levels (range r_s : 0.991–0.998), sex (r_s = 0.997), disease experience status (r_s = 0.998), medical background (r_s = 0.985) and profession (range r_s = 0.987–0.998) (appendix p65). We also found high correlation of the probit coefficients between provinces (range r_s : 0.897–0.997) (appendix p50). Within provinces, we observed a relatively lower, but still high, correlation of the coefficients by low and high income level (range r_s : 0.847–0.985) than by age group, educational level, sex, disease experience status, medical background and profession.

Compared to the Japanese study, this study found lower values of DWs for health states mentioning loss of mobility, pain, mental symptoms, sensory loss, infection or diarrhoea, substance use, impaired activities of daily living and cognitive symptoms. Mention of fatigue, disfigurement and a rest category of other physical symptoms was not associated with a significant difference in DWs. In comparison to GBD 2013, the Chinese study found significantly lower DWs for health states mentioning mobility, pain, mental symptoms, fatigue, sensory symptom, infection or diarrhoea, ADL and cognitive symptoms. Mention of disfigurement,

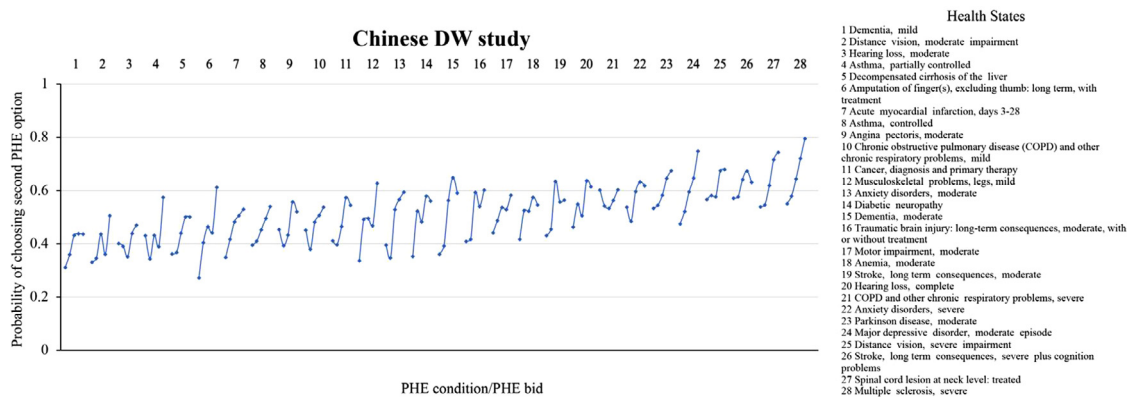


Figure 2. Probability of population health equivalence (PHE) responses of choosing the second program at each bid value for each 28 health states. Each line represents one health state and each dot represents a bid (of 1500, 2000, 3000, 5000, or 10,000 persons for the second health intervention in the PHE question) within one health state. Results are based on 4,925 respondents who answered all PHE questions; consistently answered the repeat question; and correctly identified the repeat question when asked after completion of the questionnaire.

sensory symptoms, infections/diarrhoea, substance use and impaired activities of daily living was not associated with a significant proportional difference in DWs (Table 3).

Implications for years lived with disability in China

Replacing the GBD DWs by the DWs from this study for the 2019 estimates of YLDs for China kept the ranking of top three conditions (age-related and other hearing loss, low back pain and migraine) constant (appendix p77). The YLDs for age-related and other hearing loss, however, increased by 69.1%, largely caused by much higher values of the Chinese DWs for mild (0.031 vs 0.009) and moderate hearing loss (0.027 vs 0.068). Other conditions among the top 25 of YLDs increasing with more than 50% (chronic periodontal disease, other gynecological diseases, other skin and subcutaneous diseases, scabies, atopic dermatitis, other oral disorders, near vision loss, dietary iron deficiency, gastro-oesophageal reflux disease, and gallbladder and biliary diseases) also have in common that the increases are completely or mostly explained by large relative increases in low severity disability weights. Causes with the largest decreases included neck pain, alcohol use disorders, Alzheimer's disease and other dementias and dysthymia. The latter three conditions dropped out of the top 25 when using the China DWs. The DWs were lower in the China study for each level of severity of dementia. The health state for dysthymia is the same as for mild depression and had a much lower China DW (0.053 vs 0.145). The change in major depressive disorders was not as great (-14.9%) as the China DWs for moderate and severe depression were higher than in GBD.

Total YLDs in China in 2019 were 14.9% higher if using the Chinese DWs (176,723,978 compared to 153,797,875 as reported in GBD 2019 using the GBD

DWs). The impact of higher DW values in the Chinese study for low severity health states was very substantial. After replacing the Chinese DWs by the GBD DWs for health states with a GBD DW value of less than 0.03, total YLDs reduced to 145,125,142 or 5.6% lower than in GBD. Higher estimates of these common, low severity conditions were characterized by mild anemia, mild infectious disease, mild motor impairment or mild disfigurement.

Discussion

This study presents valuations of health states by a large number of respondents across China. The range of DW values from this study (from 0.009 for distance vision mild impairment to 0.752 for severe heroin and other opioid dependence) was comparable to the range of values in a recent Japanese study (from 0.004 for distance vision mild impairment and mild anemia to 0.707 for untreated spinal cord lesion at neck level), and with the range of values in the GBD and European disability weight studies (from 0.003 for distance vision mild impairment to 0.778 for psychotic state of schizophrenia).^{2,3}

The relation between results of probit coefficients by respondent characteristics and province

Other studies have found that socioeconomic and demographic variables may influence how people value different health conditions.¹⁷⁻²² For instance, a study using PC questions in a low socio-economic setting in South Africa, showed low correlation (Pearson's r 0.44) with the DWs of the GBD 2010 study.¹⁹ Haagsma et al. suggested that educational level had influence on evaluation of injury health states in the Netherlands.²³ Ranking of health conditions estimated by Zimbabwean health professionals were very different from results by the lay public.²⁴ In contrast, we found a very high

Disability weight (95% UI)

		Disability weight (95% UI)
Infectious disease		
1	Acute episode, mild	0.027 (0.002-0.113)
2	Acute episode, moderate	0.137 (0.054-0.273)
3	Acute episode, severe	0.152 (0.066-0.288)
4	Post-acute consequences (fatigue, emotional lability, and insomnia)	0.080 (0.020-0.210)
Diarrhoea		
5	Mild	0.072 (0.016-0.198)
6	Moderate	0.151 (0.064-0.286)
7	Severe	0.259 (0.170-0.368)
8	Epididymo-orchitis	0.122 (0.044-0.259)
9	Herpes zoster	0.045 (0.006-0.154)
10	HIV: symptomatic, pre-AIDS	0.252 (0.163-0.363)
11	HIV/AIDS: receiving antiretroviral (ARV) treatment	0.074 (0.017-0.201)
12	AIDS: not receiving antiretroviral (ARV) treatment	0.367 (0.309-0.432)
13	Intestinal nematode infections: symptomatic	0.112 (0.037-0.248)
14	Lymphatic filariasis: symptomatic	0.153 (0.067-0.289)
15	Ear pain	0.052 (0.009-0.166)
Tuberculosis		
16	Not HIV infected	0.375 (0.319-0.435)
17	HIV infected	0.297 (0.216-0.390)
Cancer		
18	Diagnosis and primary treatment	0.264 (0.176-0.371)
19	Metastatic	0.324 (0.251-0.407)
Terminal phase		
20	With medication (for cancers and end-stage kidney or liver disease)	0.515 (0.505-0.527)
21	Without medication (for cancers and end-stage kidney or liver disease)	0.411 (0.369-0.454)
Cardiovascular and circulatory disease		
Acute myocardial infarction		
22	Days 1–2	0.261 (0.173-0.368)
23	Days 3–28	0.075 (0.018-0.202)
Angina pectoris		
24	Mild	0.037 (0.004-0.135)
25	Moderate	0.055 (0.010-0.173)
26	Severe	0.163 (0.074-0.297)
27	Cardiac conduction disorders and cardiac dysrhythmias	0.256 (0.165-0.364)
28	Claudication	0.021 (0.001-0.095)
Heart failure		
29	Mild	0.057 (0.010-0.175)
30	Moderate	0.089 (0.024-0.221)
31	Severe	0.162 (0.073-0.295)
Stroke		
32	Long-term consequences, mild	0.053 (0.009-0.169)
33	Long-term consequences, moderate	0.057 (0.010-0.175)
34	Long-term consequences, moderate plus cognition problems	0.111 (0.036-0.247)
35	Long-term consequences, severe	0.355 (0.290-0.423)
36	Long-term consequences, severe plus cognition problems	0.400 (0.353-0.449)
Diabetes and digestive and genitourinary disease		
37	Diabetic neuropathy	0.094 (0.027-0.227)
38	Chronic kidney disease (stage IV)	0.080 (0.020-0.210)
39	With kidney transplantation	0.071 (0.016-0.197)
40	On dialysis	0.672 (0.594-0.745)
41	Decompensated cirrhosis of the liver	0.108 (0.035-0.244)
42	Gastric bleeding	0.390 (0.338-0.443)

Table 2 (Continued)

		Disability weight (95% UI)
43	Crohn's disease or ulcerative colitis	0.135 (0.053-0.271)
44	Benign prostatic hypertrophy: symptomatic	0.088 (0.023-0.221)
45	Urinary incontinence	0.234 (0.143-0.351)
46	Stress incontinence	0.044 (0.006-0.151)
47	Impotence	0.058 (0.011-0.177)
	Infertility	
48	Primary	0.026 (0.002-0.111)
49	Secondary	0.018 (0.001-0.088)
50	Often has a burning sensation in the back of the chest after eating	0.088 (0.024-0.219)
Chronic respiratory disease		
	Asthma	
51	Controlled	0.032 (0.003-0.126)
52	Partially controlled	0.081 (0.020-0.211)
53	Uncontrolled	0.227 (0.136-0.346)
	Chronic obstructive pulmonary disease (COPD) and other chronic respiratory diseases	
54	Mild	0.022 (0.002-0.101)
55	Moderate	0.176 (0.086-0.308)
56	Severe	0.284 (0.200-0.383)
Neurological disorders		
	Dementia	
57	Mild	0.047 (0.007-0.157)
58	Moderate	0.167 (0.077-0.300)
59	Severe	0.276 (0.191-0.378)
	Headache	
60	Migraine	0.416 (0.375-0.457)
61	Tension-type	0.153 (0.065-0.288)
62	Medication overuse	0.184 (0.093-0.314)
	Multiple sclerosis	
63	Mild	0.129 (0.048-0.266)
64	Moderate	0.389 (0.340-0.442)
65	Severe	0.661 (0.586-0.731)
	Epilepsy	
66	Severe (seizures >= once a month)	0.462 (0.441-0.481)
67	Less severe (seizures 1-11 per year)	0.415 (0.376-0.457)
	Parkinson's disease	
68	Mild	0.033 (0.003-0.129)
69	Moderate	0.318 (0.241-0.403)
70	Severe	0.475 (0.458-0.488)
Mental, behavioural, and substance use disorders		
	Alcohol use disorder	
71	Very mild	0.060 (0.011-0.180)
72	Mild	0.144 (0.058-0.280)
73	Moderate	0.225 (0.133-0.344)
74	Severe	0.289 (0.202-0.386)
	Fetal alcohol syndrome	
75	Mild	0.028 (0.003-0.115)
76	Moderate	0.109 (0.036-0.245)
77	Severe	0.201 (0.109-0.326)
	Cannabis dependence	
78	Mild	0.38 (0.327-0.438)
79	Severe	0.597 (0.55-0.645)
	Amphetamine dependence	

Table 2 (Continued)

Disability weight (95% UI)

80	Mild	0.318 (0.241-0.404)
81	Severe	0.622 (0.565-0.679)
	Cocaine dependence	
82	Mild	0.287 (0.205-0.385)
83	Severe	0.497 (0.486-0.506)
	Heroin and other opioid dependence	
84	Mild	0.494 (0.483-0.502)
85	Severe	0.752 (0.640-0.841)
	Anxiety disorders	
86	Mild	0.029 (0.003-0.118)
87	Moderate	0.115 (0.040-0.253)
88	Severe	0.556 (0.528-0.588)
	Major depressive disorder	
89	Mild episode	0.053 (0.009-0.168)
90	Moderate episode	0.509 (0.500-0.520)
91	Severe episode	0.699 (0.608-0.777)
	Bipolar disorder	
92	Manic episode	0.501 (0.492-0.510)
93	Residual state	0.060 (0.012-0.181)
	Schizophrenia	
94	Acute state	0.711 (0.616-0.797)
95	Residual state	0.590 (0.548-0.633)
96	Anorexia nervosa	0.093 (0.026-0.226)
97	Bulimia nervosa	0.070 (0.015-0.195)
98	Attention deficit hyperactivity disorder	0.028 (0.002-0.116)
99	Conduct disorder	0.241 (0.149-0.354)
100	Borderline intellectual functioning	0.015 (0.001-0.078)
	Intellectual disability/mental retardation	
101	Mild	0.075 (0.018-0.202)
102	Moderate	0.096 (0.028-0.229)
103	Severe	0.115 (0.039-0.252)
104	Profound	0.228 (0.136-0.348)

Hearing and vision loss

	Hearing loss	
105	Mild	0.031 (0.003-0.122)
106	Moderate	0.068 (0.014-0.192)
107	Severe	0.246 (0.156-0.358)
108	Profound	0.200 (0.109-0.327)
109	Complete	0.151 (0.065-0.287)
110	Mild, with ringing	0.042 (0.006-0.146)
111	Moderate, with ringing	0.076 (0.017-0.204)
112	Severe, with ringing	0.233 (0.140-0.351)
113	Profound, with ringing	0.183 (0.093-0.313)
114	Complete, with ringing	0.236 (0.145-0.353)
	Distance vision	
115	Mild impairment	0.009 (0.0003-0.057)
116	Moderate impairment	0.040 (0.005-0.144)
117	Severe impairment	0.219 (0.127-0.340)
118	Blindness	0.221 (0.129-0.341)
119	Monocular	0.052 (0.008-0.165)
120	Presbyopia	0.016 (0.001-0.081)

Musculoskeletal disorders

Low back pain

Table 2 (Continued)

		Disability weight (95% UI)
121	Mild	0.046 (0.007-0.156)
122	Moderate	0.115 (0.039-0.250)
123	Severe, without leg pain	0.164 (0.076-0.298)
124	Severe, with leg pain	0.180 (0.089-0.311)
125	Most severe, without leg pain	0.180 (0.090-0.312)
126	Most severe, with leg pain	0.188 (0.095-0.317)
	Neck pain	
127	Mild	0.047 (0.007-0.159)
128	Moderate	0.070 (0.015-0.196)
129	Severe	0.126 (0.045-0.263)
130	Most severe	0.106 (0.033-0.242)
	Musculoskeletal problems	
131	Legs, mild	0.043 (0.006-0.149)
132	Legs, moderate	0.122 (0.043-0.259)
133	Legs, severe	0.162 (0.073-0.294)
134	Arms, mild	0.047 (0.007-0.157)
135	Arms, moderate	0.123 (0.044-0.260)
136	Generalized, moderate	0.138 (0.055-0.275)
137	Generalized, severe	0.306 (0.228-0.396)
138	Gout, acute	0.212 (0.119-0.335)
	Injury	
139	Amputation of one upper limb (long term, without treatment)	0.117 (0.040-0.253)
140	Concussion (short term)	0.083 (0.021-0.213)
141	Spinal cord lesion, below neck level (treated)	0.359 (0.297-0.425)
	Other	
	Abdominopelvic problem	
142	Mild	0.039 (0.005-0.143)
143	Moderate	0.175 (0.084-0.307)
144	Severe	0.420 (0.382-0.459)
	Anaemia	
145	Mild	0.009 (0.0003-0.057)
146	Moderate	0.094 (0.026-0.226)
147	Severe	0.179 (0.088-0.310)
148	Periodontitis	0.019 (0.001-0.091)
149	Dental caries: symptomatic	0.018 (0.001-0.087)
150	Severe tooth loss	0.046 (0.007-0.156)
	Disfigurement	
151	Level 1	0.048 (0.007-0.159)
152	Level 2	0.154 (0.067-0.289)
153	Level 3	0.572 (0.537-0.608)
154	Level 1, with itch or pain	0.063 (0.013-0.186)
155	Level 2, with itch or pain	0.192 (0.100-0.321)
156	Level 3, with itch or pain	0.625 (0.567-0.683)
	Generic uncomplicated disease	
157	Worry and daily medication	0.045 (0.006-0.153)
158	Anxiety about diagnosis	0.018 (0.001-0.087)
159	Kwashiorkor	0.079 (0.019-0.207)
160	Severe wasting	0.154 (0.067-0.287)
161	Speech problems	0.041 (0.005-0.146)
	Motor impairment	
162	Mild	0.029 (0.003-0.119)
163	Moderate	0.044 (0.006-0.151)
164	Severe	0.262 (0.174-0.370)

Table 2 (Continued)

Disability weight (95% UI)

		Disability weight (95% UI)
	Motor plus cognitive impairments	
165	Mild	0.036 (0.004-0.135)
166	Moderate	0.075 (0.017-0.202)
167	Severe	0.271 (0.184-0.373)
168	Rectovaginal fistula	0.340 (0.271-0.416)
169	Vesicovaginal fistula	0.229 (0.136-0.347)
170	Thrombocytopenic purpura	0.131 (0.05-0.268)
171	Hypothyroidism	0.030 (0.003-0.120)
172	Hyperthyroidism	0.117 (0.040-0.253)
173	Neck pain, moderate	0.103 (0.032-0.240)
174	Osteomyelitis	0.107 (0.034-0.242)
175	Shoulder lesions	0.031 (0.003-0.123)
176	Heart burn & reflux "GERD"	0.092 (0.025-0.225)
177	Constipation	0.061 (0.012-0.181)
178	Vaginal discharge	0.028 (0.002-0.115)
179	Dyspareunia	0.044 (0.006-0.151)
180	Stress incontinence	0.046 (0.006-0.155)
181	Irritable bowel syndrome	0.045 (0.006-0.151)
182	Somatoform disorder	0.097 (0.028-0.231)
183	Borderline personality disorder	0.196 (0.103-0.322)
184	Harmful alcohol use	0.098 (0.029-0.233)
185	Vertigo and balance disorder (Menière, labyrinthitis)	0.065 (0.013-0.186)
186	Trigeminal neuralgia	0.088 (0.024-0.22)
187	Encephalopathy - moderate	0.175 (0.085-0.306)
188	Encephalopathy - severe	0.285 (0.200-0.383)
189	Thrombocytopenic purpura	0.141 (0.056-0.277)
190	Lymphogranuloma Venereum - local infection	0.104 (0.032-0.238)
191	Subacute sclerosing panencephalitis - phase 1	0.056 (0.010-0.173)
192	Subacute sclerosing panencephalitis - phase 2	0.151 (0.065-0.286)
194	Haemorrhoids	0.091 (0.025-0.224)
195	Anal fissure/abcess/fistula	0.069 (0.015-0.194)
196	Hyperthyroidism	0.073 (0.016-0.198)
197	Allergic rhinitis (hay fever)	0.015 (0.001-0.076)
198	Varicose veins	0.039 (0.005-0.141)
199	Carpal tunnel syndrome	0.028 (0.003-0.117)
200	Intensive care unit admission	0.669 (0.589-0.741)
201	Invasive device/drain	0.175 (0.084-0.307)
202	Insomnia	0.040 (0.005-0.144)
203	Sleep apnoea	0.080 (0.020-0.209)
204	Hypothyroidism	0.042 (0.006-0.146)
205	Hearing loss, moderate (Modified)*	0.068 (0.014-0.193)
206	Hearing loss, severe (Modified)*	0.222 (0.131-0.342)

Table 2: Disability weights with 95% UI for 206 health states in mainland China.

Health state 'Subacute sclerosing panencephalitis - phase 2' numbered 193 was omitted in probit analysis due to its collinearity. Thus, we didn't report DW value of this health state finally. * We modified the two lay descriptions of moderate and severe hearing loss health states in our study.

degree of correlation between the probit coefficients by provinces and by medical background, profession, income levels, age groups, educational levels, sex and disease experience status. These findings were consistent with the European DW study which reported high correlations of the probit coefficients between country, age, sex, disease experience status, income and educational levels.¹⁷

Paired comparison and population health equivalence responses

Through a number of indicators, we believe we have collected information that is comparable or better in quality compared to other disability weight surveys. First, the high correlation of results by provinces and respondent characteristics is indicative of internal consistency. Second, on the test-retest PC questions we found that

Symptom (number of lay descriptions ^a)	China-GBD 2013			China-Japan			Japan-GBD2013		
	Coefficient	95% CI	p-value	Coefficient	95% CI	p-value	Coefficient	95% CI	p-value
Mobility (26)	-46.24	-86.87 to -5.60	0.03	-42.77	-79.12 to -6.41	0.02	-7.13	-49.70 to -35.45	0.74
Pain (72)	-40.14	-71.52 to -8.76	0.01	-64.43	-91.32 to -37.52	0.000*	40.95	9.45 to 72.45	0.01
Mental symptom (58)	-73.95	-105.16 to -42.74	0.000*	-30.52	-58.50 to -2.54	0.03	-49.14	-81.91 to -16.38	0.004
Fatigue (50)	-43.66	-75.80 to -11.52	0.01	2.01	-24.86 to 28.88	0.88	-29.03	-60.50 to 2.44	0.07
Disfigurement (9)	53.41	-13.01 to 119.83	0.11	-40.01	-103.38 to 23.36	0.21	71.07	-3.13 to 145.28	0.06
Sensory symptom (21)	-14.97	-65.23 to 35.28	0.56	-71.96	-114.92 to -29.00	0.001	56.27	5.96 to 106.58	0.03
Infection/diarrhoea (16)	-43.88	-95.43 to 7.67	0.10	-66.47	-113.63 to -19.33	0.01	47.28	-7.94 to 102.49	0.09
Substance use (13)	35.02	-21.38 to 91.41	0.22	-83.37	-160.14 to -6.60	0.03	-39.36	-129.26 to 50.53	0.39
ADL (83)	-20.70	-49.08 to 7.69	0.15	-27.82	-52.45 to -3.20	0.03	15.53	-13.31 to 44.37	0.29
Cognitive symptom (30)	-68.33	-110.90 to -25.75	0.002	-47.10	-84.35 to -9.84	0.01	-20.35	-63.97 to 23.28	0.36
Others (75)	-37.66	-67.02 to -8.29	0.000*	-9.89	-35.86 to 16.08	0.45	-7.60	-38.01 to 22.81	0.62

Table 3: Regression analysis results for proportional differences in the Chinese DWs compared with GBD 2013 study and Japanese DWs for 206 comparable health states.

^a The number of lay descriptions add to more than the total because lay descriptions often combine several symptom categories.

DW: disability weight; ADL: activities of daily living; Others: other physical symptoms, including dyspnoea, nausea, palpitations, reduced appetite, sleeping problems; CI: confidence intervals.

* Correlation coefficients significantly different from zero ($p < 0.05$).

73% of respondents provided the same answer, similar to the 71% reported by the GBD and European survey.^{3,9} Third, health states with a logical succession of severity level (e.g. mild, moderate and severe) disability weights followed an increasing pattern. There were certain exceptions to this rule. The DW for complete hearing loss was lower than that for profound and severe hearing loss; the DW for profound hearing loss with ringing was lower than the DWs for severe and complete hearing loss with ringing; and the DW for most severe neck pain was lower than that for severe neck pain. In all these cases the wide uncertainty intervals show considerable overlap. A possible explanation for this may be that brief lay descriptions were used to describe the major functional consequences and symptoms associated with the health state and that the disease label, indicating the cause of the health state, was removed from the description.^{2,3} Fourth, while the European and Japanese surveys concluded that they could not use the PHE questions because of an inadequate signal among the measurement noise, we found a reasonable differentiation between the 28 PHE questions, though not as much as in the internet survey that was part of the original GBD disability weight surveys.^{3,14} Removal of responses that did not meet pre-set quality criteria may have contributed to better results from the PHE questions, even if that meant dropping as much as 71% of responses. Another factor that may have contributed to better PHE responses is the relatively high level of education that is equivalent to universities and colleges in our study (67% higher education) relative to the Japanese (46%) and European surveys (60%), though not as high as reported for the respondents of the GBD internet survey (93%). The PHE questions are more cognitively demanding than the PC questions and may be more suited to well-educated respondents improving the signal-to-noise ratio in the responses.³ In future disability weight studies, it seems advisable to adopt similar quality control criteria and exclusions, and to select respondents with higher education only to answer the PHE questions.

Differences of estimated Chinese DW from DW in previous studies

Many previous studies have shown that contextual differences may impact the ways people perceive health problems and how such problems affect their lives.^{16,19,25-29} The positive coefficients from liner regression for substance use and disfigurement had very wide confidence intervals that spanned unity, which refers to China DW being higher than GBD DW. In the comparison between GBD and the Japanese study only three domains had significant coefficients: higher Japanese values for pain and sensory symptoms but lower values for mental symptoms. While this analysis by domain allows a more detailed comparison and

suggests some pattern to the differences observed, we found considerable heterogeneity in the differences between DWs within similar domains of health. For instance, the Chinese study assigned considerably higher DW values to all drug use disorder health states but lower values to alcohol use disorder and that led to a non-significant coefficient on the aggregate domain of substance use. A previous study using a different study design of ranking health conditions in 14 countries also noted a higher ranking of drug dependence in China compared to other countries.¹⁶ DWs for drug use disorders were also very high in a study in South Korea.¹⁵ In contrast, the Japanese study found lower DW values for substance use disorder health states. A common finding between the Chinese and Japanese surveys was a generally lower valuation of health states with mental symptoms.

The third and most revealing comparison was through a re-estimation of the GBD 2019 YLD estimates for China using this study's DWs. We found considerable variation in rank and estimated number of YLDs among the top-25 causes. The largest differences in YLDs were for common causes or sequelae of causes with low disability weights, including chronic periodontal diseases, mild disfigurement with pain for scabies and atopic dermatitis, mild and moderate anemia, mild and moderate hearing loss and near vision loss. A striking jump in ranking order was for chronic periodontal disease, from rank 29 in GBD (with DW 0.007) to rank 11 using the Chinese DW (0.019). However, while the relative difference is large (a factor 2.8) in absolute DW terms it is a rather small difference of 0.012 and well within the uncertainty bounds of the GBD and Chinese DWs. The DW values for mild disfigurement is closely to the values in Japan, the DWs of mild and moderate anemia of China and Japan are relatively close, and the values in both countries were higher than those in GBD (appendix p7).¹⁴ The more severe health states of neck pain had much lower Chinese DW values compared to GBD but this was not the case for low back pain. Mild and moderate osteoarthritis of the knee had much higher DWs in China but not the most severe health state. More generally, health states with pain as a key symptom had lower DWs in China compared to GBD while in Japan, these DWs were significantly higher. The Japanese study authors discussed that Japanese people may be more sensitive to pain-related as found in study comparing pain detection thresholds between ethnic groups.^{30,31} The lower values for pain-related health states in China, is possibly related to the large proportion of respondents with higher education. An Austrian study found a gradient in reporting disabling pain by socio-economic status.³² A Chinese study also demonstrated that socioeconomic variables such as education, occupation and health conditions are associated with both moderate severe pains.³³ These heterogeneous findings do not clearly point at domains of health

that explain the different valuations despite the regression finding lower coefficients for eight of the 11 domains examined as explanatory factors for the proportional difference between Chinese and GBD DWs. This makes it less likely that cultural factors play a major role unless we have missed defining an important domain of health. Particularly, the finding that the lowest level of severity health states in general were valued as more severe, though at small absolute differences, points more at a methodological difficulty to measure these low severity health states accurately. As many of these health states of minor severity are common, we saw large shifts in YLD estimates after replacing the GBD DWs by Chinese DWs. On the other hand, the lower values for mental symptoms in both the Chinese and Japanese surveys may indicate a cultural difference in East Asian countries how mental symptoms are valued. Japanese people believe that mental illness are curable,³⁴ and the loss of mental self-control is seen as basically an inability of the individual to exert will power. This has created a stigma towards people with mental illness and is one of the major social problems.^{34,35} In China, high prevalence of mental illness-related stigma was also generally existed.^{36,37} Fear of stigmatization might make Chinese individuals more likely to deny (or forget) transient psychological conditions.³⁸ This is possibly the reason why the DW of mental symptoms in both China and Japan is so low.

Limitations and strengths

This is by far the largest DW study ever conducted producing consistent estimates across 31 provinces of China. A rigorous application of quality criteria improved the signal to noise ratio, though at the expense of discarding more than half of respondent answers. Our quality criterion of test-retest reliability may have been too strict. Altering a valuation between the first and tenth or fifteenth pairs of health states may reflect a learning effect rather than being indicative of poor quality. Future disability weight surveys can explore this further as a research question.

There are a number of limitations in this study. First, due to our purposive sampling method respondents in this study were much better educated and of higher socio-economic status and younger than the general population. For logistical reasons, none of the previous DW surveys systematically sampled from national populations either. The GBD country surveys were conducted in small geographical areas. The European and Japanese surveys relied on marketing companies' internet panels and the GBD internet survey, similar to this study, relied on recruitment through professional networks. The very high correlation of results by income and educational level found in this study and the previous finding from the GBD surveys of high correlation between country surveys with very low education

attainment and the internet survey that had 93% higher-educated respondents suggests that these factors are not so influential. An important finding from this study is the strong relationship between usable answers to PHE questions and educational level suggesting that the complexity of these questions may require selection of more educated respondents. The PC questions are less cognitively demanding and experience from all DW surveys using these questions has been that sensible answers can be derived regardless of the level of education of respondents. Second, our study did not include participants older than 69 years. A previous study suggests that elderly individuals may value hypothetical health states differently than the general population.³⁹ Older raters in that study gave significantly more weight to functional limitations and social functioning and less to morbidities and pain experience, compared to younger raters. We found very high correlation between results for three age groups in this study. In future work, respondents over the age of 70 may be selected but may need to be interviewed in person rather than online. Thirdly, differences in estimated levels of DW can possibly be attributed to variation between countries in the interpretation of the translated lay descriptions for health states. For instance, the lay description for chronic periodontal disease, mentions occasional, minor bleeding of the gums and mild discomfort. Possibly, the translation into Chinese could be a reason for the more severe valuation of this health state. Besides, the ranking of certain conditions seems counterintuitive, which may be related to the formulating Chinese lay descriptions of health states, and we suggest these descriptions to be further improved by medical experts.

Conclusions

This study elicited national and subnational DWs of health states from disease based on the responses of almost 500,000 persons. With strict quality control criteria we derived a consistent set of DWs including usable answers to the anchoring PHE questions that had failed to provide estimates in the European and Japanese surveys. The relatively high level of education of Chinese respondents may be the reason. The considerable differences in ranking of the top 25 causes when using these new Chinese DWs, indicating that cultural or other contextual factors influencing results with lower valuations for mental disorders and alcohol use disorders, and higher DWs being assigned to drug use disorders in China. Other large changes in years lived with disability were due to methodological limitations of distinguishing between low severity DWs. Based on this study, China could consider applying these new weights to their burden of disease estimates. With the recent Japanese study and this large new study from China, it would be advisable to do a common analysis

with all previous DW study results as an update for the GBD to be used in future iterations.

Contributors

CY contributed to the conception and design of the study. XL implemented analysis of data, and wrote the first draft of the manuscript. CY, TV and FW contributed to interpretation of data, and revision of the manuscript. MZ, YY, CY, XL and FW contributed to implementation of data collection. TV, CJLM, JAH, JAS, SN and SIH contributed to guidance on data analysis, results interpretation. CY, MZ, FW, JQ, PY, SY, XL, HW, FS, YW and PS worked on development of health-state descriptions for the surveys. Other authors contributed to data collection and feedback on study results.

Data sharing statement

Data analyzed in this paper are publicly accessible and details of how to access these data are in the appendix.

Declaration of interests

All other authors declare no competing interests.

Acknowledgement

We thank all the survey respondents who participated in the study and the survey field staff who did the surveys. We acknowledge institutions that were pivotal in implementation of the survey in China, providing administrative and logistical support to data collection. We thank the many individuals who have contributed to Chinese DW Measurement Study in various capacities, and particularly recognize the important contributions to this work from many CDC staff members. We are grateful for guidance and feedback from participants at the Expert Consultation on DW.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.lanwpc.2022.100520](https://doi.org/10.1016/j.lanwpc.2022.100520).

References

- 1 Burstein R, Fleming T, Haagsma JA, Salomon JA, Vos T, Murray CJ. Estimating distributions of health state severity for the global burden of disease study. *Popul Health Metr*. 2015;13:31.
- 2 Salomon JA, Haagsma JA, Davis A, et al. Disability weights for the global burden of disease 2013 study. *Lancet Glob Health*. 2015;3(11):e712–e723.
- 3 Haagsma JA, Maertens de Noordhout C, Polinder S, et al. Assessing disability weights based on the responses of 30,660 people from four European countries. *Popul Health Metr*. 2015;13:10.
- 4 Murray CJ. Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bull World Health Organ*. 1994;72(3):429–445.
- 5 Murray CJ, Lopez AD. Quantifying disability: data, methods and results. *Bull World Health Organ*. 1994;72(3):481–494.

- 6 Murray CJ, Lopez AD, Jamison DT. The global burden of disease in 1990: summary results, sensitivity analysis and future directions. *Bull World Health Organ.* 1994;72(3):495–509.
- 7 Salomon JA, Murray CJ, Ustun B, Chatterji S. Health state valuations in summary measures of population health. *Value Health.* 2002;5(6):474.
- 8 Murray CJ, Acharya AK. Understanding DALYs (disability-adjusted life years). *J Health Econ.* 1997;16(6):703–730.
- 9 Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2129–2143.
- 10 GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2018;392(10159):1859–1922.
- 11 Ustun TB, Saxena S, Rehm J, Bickenbach J. Are disability weights universal? WHO/NIH Joint Project CAR Study Group. *Lancet.* 1999;354(9186):1306.
- 12 Voigt K, King NB. Disability weights in the global burden of disease 2010 study: two steps forward, one step back? *Bull World Health Organ.* 2014;92(3):226–228.
- 13 Haagsma JA, Polinder S, Cassini A, Colzani E, Havelaar AH. Review of disability weight studies: comparison of methodological choices and values. *Popul Health Metr.* 2014;12:20.
- 14 Nomura S, Yamamoto Y, Yoneoka D, et al. How do Japanese rate the severity of different diseases and injuries?—an assessment of disability weights for 231 health states by 37,318 Japanese respondents. *Popul Health Metr.* 2021;19(1):21.
- 15 Ock M, Ahn J, Yoon SJ, Jo MW. Estimation of disability weights in the general population of South Korea using a paired comparison. *PLoS One.* 2016;11(9):e0162478.
- 16 Ustun TB, Rehm J, Chatterji S, et al. Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. WHO/NIH joint project CAR study group. *Lancet.* 1999;354(9173):111–115.
- 17 de Noordhout CM, Devleeschauwer B, Salomon JA, et al. Disability weights for infectious diseases in four European countries: comparison between countries and across respondent characteristics. *Eur J Public Health.* 2018;28(1):124–133.
- 18 Hofman CS, Makai P, Boter H, et al. The influence of age on health valuations: the older olds prefer functional independence while the younger olds prefer less morbidity. *Clin Interv Aging.* 2015;10:1131–1139.
- 19 Neethling I, Jelsma J, Ramma L, Schneider H, Bradshaw D. Disability weights from a household survey in a low socio-economic setting: how does it compare to the global burden of disease 2010 study? *Global Health Action.* 2016;9:31754.
- 20 Dyer MTD, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health Qual Life Out.* 2010;8:13.
- 21 Fang MY, Oremus M, Tarride JE, Raina P. Study CW-t-p. A comparison of health utility scores calculated using United Kingdom and Canadian preference weights in persons with Alzheimer's disease and their caregivers. *Health Qual Life Out.* 2016;14(1):105.
- 22 Korpi T. Is utility related to employment status? Employment, unemployment, labor market policies and subjective well-being among Swedish youth. *Labour Econ.* 1997;4(2):125–147.
- 23 Haagsma JA, van Beeck EF, Polinder S, Hoeymans N, Mulder S, Bonsel GJ. Novel empirical disability weights to assess the burden of non-fatal injury. *Inj Prev.* 2008;14(1):5–10.
- 24 Jelsma J, Chivaura VG, Mhundwa K, De Weerd W, de Cock P. The global burden of disease disability weights. *Lancet.* 2000;355(9220):2079–2080.
- 25 Manchaiah V, Danermark B, Ahmadi T, et al. Social representation of “hearing loss”: cross-cultural exploratory study in India, Iran, Portugal, and the UK. *Clin Interv Aging.* 2015;10:1857–1872.
- 26 Comman JC, Gleit D, Rodriguez G, Goldman N, Hurng BS, Weinstein M. Demographic and socioeconomic status differences in perceptions of difficulty with mobility in late life. *J Gerontol B Psychol Sci Soc Sci.* 2011;66(2):237–248.
- 27 Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Med Decis Making.* 2014;34(1):8–20.
- 28 Diener E, Diener M. Cross-cultural correlates of life satisfaction and self-esteem. *J Pers Soc Psychol.* 1995;68(4):653–663.
- 29 Yiu EM, Ho EM, Ma EP, et al. Possible cross-cultural differences in the perception of impact of voice disorders. *J Voice.* 2011;25(3):348–353.
- 30 Komiyama O, Kawara M, De Laat A. Ethnic differences regarding tactile and pain thresholds in the trigeminal region. *J Pain.* 2007;8(4):363–369.
- 31 Komiyama O, Wang K, Svensson P, Arendt-Nielsen L, Kawara M, De Laat A. Ethnic differences regarding sensory, pain, and reflex responses in the trigeminal region. *Clin Neurophysiol.* 2009;120(2):384–389.
- 32 Dorner TE, Muckenhuber J, Stronegger WJ, Rasky E, Gustorff B, Freidl W. The impact of socio-economic status on pain and the perception of disability due to pain. *Eur J Pain.* 2011;15(1):103–109.
- 33 Yu T, Ma J, Jiang Y, et al. Assessing pain among Chinese Elderly-Chinese health and retirement longitudinal study. *Iran J Public Health.* 2018;47(4):553–560.
- 34 Kasahara-Kiritani M, Matoba T, Kikuzawa S, et al. Public perceptions toward mental illness in Japan. *Asian J Psychiatry.* 2018;35:55–60.
- 35 Desapriya EB, Nobutada I. Stigma of mental illness in Japan. *Lancet.* 2002;359(9320):1866.
- 36 Kudva KG, El Hayek S, Gupta AK, et al. Stigma in mental illness: perspective from eight Asian nations. *Asia Pac Psychiatry.* 2020;12(2):e12380.
- 37 Yin H, Wardenar KJ, Xu G, Tian H, Schoevers RA. Mental health stigma and mental health knowledge in Chinese population: a cross-sectional study. *BMC Psychiatry.* 2020;20(1):323.
- 38 Phillips MR, Zhang J, Shi Q, et al. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. *Lancet.* 2009;373(9680):2041–2053.
- 39 Botes R, Vermeulen KM, Gerber AM, Ranchor AV, Buskens E. Health-related quality of life and well-being health state values among Dutch oldest old. *Patient Prefer Adher.* 2019;13:721–728.