

# BMJ Open Developing and testing culturally relevant bolt-on items for EQ-5D-5L in Chinese populations: a mixed-methods study protocol

Zhuxin Mao <sup>1</sup>, Jingzhi Fan,<sup>2</sup> Fanni Rencz,<sup>3</sup> Zhihao Yang,<sup>4</sup> Nan Luo,<sup>5</sup> Pei Wang<sup>2</sup>

**To cite:** Mao Z, Fan J, Rencz F, *et al.* Developing and testing culturally relevant bolt-on items for EQ-5D-5L in Chinese populations: a mixed-methods study protocol. *BMJ Open* 2024;**14**:e081140. doi:10.1136/bmjopen-2023-081140

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-081140>).

ZM and JF contributed equally.

Received 19 October 2023  
Accepted 04 January 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>University of Antwerp, Antwerpen, Belgium

<sup>2</sup>Fudan University, Shanghai, China

<sup>3</sup>Corvinus University of Budapest, Budapest, Hungary

<sup>4</sup>Health Services Management Department, Guizhou Medical University, Guiyang, China

<sup>5</sup>National University Singapore Saw Swee Hock School of Public Health, Singapore

## Correspondence to

Dr Pei Wang;  
[wang\\_p@fudan.edu.cn](mailto:wang_p@fudan.edu.cn)

## ABSTRACT

**Introduction** EQ-5D is one of the most frequently used health-related quality of life (HRQoL) measures but has been found to be insensitive in detecting differences in health status in some general populations and disease groups. For example, the appropriateness of applying EQ-5D in the Chinese cultural setting has been widely discussed. Adding additional HRQoL dimensions (bolt-on items) can be a solution to both retain the original descriptive system of EQ-5D, while enhancing its sensitivity to the local context. To date, no studies have proposed culturally relevant bolt-ons for China or examined the psychometric properties of such bolt-on items. This protocol documents the identification, development, selection and psychometric testing of culturally relevant bolt-on items for EQ-5D-5L in China.

**Methods and analysis** We will identify and develop candidate bolt-on items that are most relevant in the Chinese culture, through former literature reviews on health concepts important for the Chinese population, conducting expert consultations and qualitative interviews. We will quantitatively test the acceptability and measurement properties (including distributional characteristics and construct validity) of the candidate items in both general and disease populations in a cross-sectional setting. The patient group will be followed up to collect two-time-point data to assess test-retest reliability of the candidate items. Bolt-on item selection will consider both the qualitative and quantitative evidence gathered. This protocol outlines a comprehensive mixed-methods process for identifying, developing, selecting and testing bolt-on items that are relevant and culturally appropriate in China. This study may serve as a guide for similar initiatives in other cultural contexts.

**Ethics and dissemination** This study received ethics approval from the Institutional Review Board of School of Public Health, Fudan University (IRB number: 2022-TYSQ-03-154). Study findings will be disseminated through international peer-reviewed journal articles as well as public, academic presentations at national and international conferences.

## BACKGROUND

EQ-5D is one of the most frequently used health-related quality of life (HRQoL) measures.<sup>1–3</sup> It has been widely used as

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ The first to use both qualitative and quantitative methods to systematically compare the EQ-5D-5L plus culturally relevant bolt-on(s) with the standard EQ-5D-5L. The protocol adopts a comprehensive and evidence-based approach to bolt-on item development.
- ⇒ The whole bolt-on item development process will follow EuroQol recommended criteria ensuring scientific standards and maintaining consistency with the EQ-5D-5L descriptive system.
- ⇒ The empirical data to be collected from patients and the general population may not be representative of the whole Chinese population.
- ⇒ While this study will not assess the impact of the developed bolt-on items on the valuation of health states, a follow-up study may be planned for this purpose.

a preference-based measure to calculate quality-adjusted life years (QALY) that can be used in economic evaluations to guide health resource allocation strategies.<sup>4–5</sup> EQ-5D is composed of two components: the descriptive system and the visual analogue scale (EQ VAS). The descriptive system contains five dimensions including mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Despite its common use, it has been argued that the five dimensions of the EQ-5D might be insensitive to some cultural populations or groups.<sup>6–7</sup>

In China, EQ-5D has been widely used in general and patient populations, since the official Chinese version was approved in 2003.<sup>8–9</sup> It is a recommended HRQoL tool for use in economic evaluations by the Chinese pharmacoeconomic evaluation guidelines.<sup>10</sup> However, a high overall ceiling effect was found in 87% of respondents in a national population study that collected EQ-5D-3L responses in China.<sup>11–12</sup> This rate is considerably higher than those observed in other

countries.<sup>13</sup> When EQ-5D-5L was used in China, higher ceiling effects (54%,<sup>14</sup> 50.8%<sup>15</sup>) were also observed, compared with other countries (35.2% in Belgium,<sup>16</sup> 33.3% in Portugal,<sup>17</sup> 31.2% in the USA<sup>18</sup>). EQ-5D was also found to be insensitive in detecting differences in health status, for example, between groups of respondents with different levels of subhealth.<sup>19 20</sup> Specifically, it has been found that the utility value did not significantly distinguish the 'illness' and 'severe sub-health' groups nor the 'health' and 'mild sub-health' groups. Some scholars have argued that due to cultural differences between China and the West, health can be conceptualised differently in China, therefore, EQ-5D may miss some relevant and important aspects of health for Chinese populations. For example, a scoping review has shown that health concepts including 'emotion control', 'weather adaption' and 'social adaption' were exclusively introduced by HRQoL measures developed in China.<sup>6</sup> An empirical study showed that sleep quality, body constitution and spiritual appearance were considered to be most important in a Chinese sample but none of them are included in EQ-5D.<sup>7</sup> As a result, Chinese scholars have developed locally relevant HRQoL measures, such as China Health-Related Outcomes Measure<sup>21</sup> and Chinese Medicine Quality of Life-11 Dimensions (CQ-11D),<sup>22</sup> introducing specific dimensions such as appetite, sleep quality and social interactions that are not covered in EQ-5D. However, these measures are not internationally recognised and, for example, CQ-11D is designed specifically for assessing disease impacts and health outcomes associated with traditional Chinese medicine (TCM) interventions.<sup>20</sup> Therefore, their usefulness in cross-country studies may be limited. The absence of value sets for most of these HRQoL measures also restricts their use, while both 3L and 5L have developed national value sets for use in China.<sup>23 24</sup>

Adding health dimensions, also known as bolt-on items, can be a solution to retain the original descriptive system of EQ-5D, while enhancing its sensitivity to the local context (hereafter referred to as 'culturally relevant bolt-ons').<sup>25</sup> While possible bolt-ons have been shown added value in terms of improvement in content validity, evidence is not sufficient on the impact of bolt-on dimensions on responsiveness, reliability and validity, and the evidence of the impact of bolting items on valuation was also mixed: some reported significant differences in utility values while some revealed otherwise.<sup>26 27</sup> Some suggested bolt-on items in patient groups, such as cognitive impairment,<sup>28 29</sup> psoriasis,<sup>30 31</sup> hearing impairments<sup>32</sup> or under cataract surgery.<sup>33</sup> Some studies tested the performance of EQ-5D bolt-ons among a general population and most of these studies were conducted in Western countries including the UK,<sup>34</sup> Switzerland<sup>35</sup> and the Netherlands.<sup>36</sup> A few studies have explored cultural bolt-ons in non-Western populations, such as in Korea, Thailand and Malaysia.<sup>37–39</sup> A Thai study showed that adding culturally relevant bolt-on items could improve the discriminative capacity of standard EQ-5D,<sup>37</sup> and a

Korean study found that additional bolt-on items could more effectively explain variation in HRQoL.<sup>38</sup> But to date, no studies have thoroughly detailed the process of identifying culturally relevant bolt-on items and/or examined the a comprehensive set of psychometric properties of these bolt-ons (eg, test-retest reliability has not been assessed in these studies). Additionally, no comprehensive studies have developed culturally relevant bolt-ons for the Chinese population, leaving a notable research gap.

The primary objective of this study is to systematically identify culturally relevant and important HRQoL dimensions for the Chinese population. Using these identified dimensions, the study aims to develop bolt-on items to the EQ-5D-5L. We aim to select the best performing dimensions for use as culturally relevant bolt-ons to the EQ-5D-5L in China based on a combination of qualitative and quantitative (acceptability and psychometric) evidence.

## METHODS

### Study design

The EuroQol Research Foundation provided permission to conduct this research. We will identify, develop, assess and select bolt-on items following predefined criteria proposed by Mulhern and colleagues,<sup>25</sup> which focus on issues relating to structure, language and consistency with the existing EQ-5D dimension structure. We will implement a mixed-methods design, combining both qualitative and quantitative research techniques. The first-stage qualitative data collection started on 1 November 2023, after we submitted the protocol to the journal. We aim to complete the data collection by June 2024.

### Patient and public involvement

Our mixed-methods approach will efficiently facilitate the involvement of patients and general public in designing and conducting this study, as our surveying questions in the quantitative phase will be largely shaped by the input from patients and members of the general population through the qualitative interviews.

### Literature reviews and qualitative interviews to identify a list of potential bolt-on items

Previous (systematic) reviews<sup>6 40</sup> and qualitative work<sup>7</sup> have identified various health dimensions reflecting Chinese cultural characteristics and could be important in health measurement in China. We therefore will not replicate the previous work by conducting similar literature reviews, but will build on the existing recent work.

Based on the evidence collected in the existing reviews and the above-mentioned qualitative study, we will conduct internal discussions to select potential candidate dimensions, which are widely mentioned or used in health description and measurement in China, for use in subsequent qualitative interviews with HRQoL experts and lay people. Because the TCM concepts reflect the

understanding of health in the Chinese culture,<sup>20</sup> they will also be considered as potential bolt-on items.

Examples of potential candidate dimensions may include sleep, tiredness/lack of strength, appetite, climate adaption, emotional control, social adaption and social support. Some of these candidate dimensions already have been proposed as bolt-ons in other languages,<sup>39 41</sup> such as sleep, tiredness, social support and appetite, and we may use their phrases as a starting point to translate them to Mandarin.

We will then conduct semistructured, face-to-face, one-to-one qualitative interviews with HRQoL experts, members of the general population as well as patients with chronic diseases to evaluate the content validity, including relevance, comprehensiveness and comprehensibility of each candidate dimension and its levels. Consolidated Criteria for Reporting Qualitative Research checklist<sup>42</sup> will be followed to ensure the rigour of the qualitative data collection and analysis.

### Respondents

The sample size for each group of respondents (experts, healthy people and patients) will meet the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) for assessing the content validity of HRQoL measures, which is at least seven respondents per group for qualitative studies.<sup>43</sup> When determining the exact number of respondents, we will take into account the concept of 'saturation'. The data collection will only end once saturation is achieved, namely the point at which additional data collection will not generate new themes or information. To attain this goal, after every three interviews, we will conduct preliminary data analysis, which will allow us to categorise and compare across collected data, refine interview questions and determine subsequent participant recruitment. This iterative process will end once saturation is achieved.

We will reach and contact at least 10 clinicians and HRQoL researchers who are Chinese, familiar with the Chinese context, with a minimum of 5-year clinical/research experience and have conducted/published health outcomes research in China.

In addition, a sample of respondents with and without health conditions will also be recruited to evaluate the content validity of the candidate dimensions. We will use purposive sampling to recruit a group of healthy general public (without any physical functioning or long-term health conditions, at least 10) and a group of patients (patients with functioning problems or chronic diseases, at least 20). By defining the healthy general public group, we aim to distinguish each group from the other.

Informed consent will be obtained from each respondent before each interview starts.

### Data collection procedure

To ensure the privacy of the interviewees and to facilitate unrestricted communication, the interviews will be conducted in quiet and private places, such as an office

or a meeting room. We will follow two sets of predefined topic guide of the semistructured interviews for HRQoL experts and non-experts separately. The questions were designed based on criteria for developing, assessing and selecting candidate EQ-5D-5L bolt-ons<sup>25</sup> as well as COSMIN criteria for evaluating the content validity of patient-reported outcome measures.<sup>44</sup>

For experts, we will ask their opinions on EQ-5D-5L as well as a selected set of candidate bolt-on items (box 1). Experts will be asked to comment on the comprehensibility, relevance, comprehensiveness, alignment with standard EQ-5D-5L and descriptive validity (suitability for valuation tasks) of the items. For non-expert respondents, we will start with a few warm-up questions to ask the respondents to describe their own current health and HRQoL and what health means to them. Respondents will then be asked to complete the EQ-5D-5L. Cognitive interview procedures will be followed to assess understanding of the instrument's instructions and items as well as the relevance and comprehensiveness of the content. Respondents will then be provided with a list of preliminary candidate dimensions and comments on the comprehensibility, relevance, comprehensiveness and alignment with standard EQ-5D-5L of the candidate dimensions (box 2). Descriptive validity will not be asked among non-expert participants because we consider questions relating to hypothetical health state can be difficult for lay respondents to interpret and significantly add the respondent burden. Probing questions will be used to seek additional information from respondents: clarifying their responses, exemplifying their responses or reasoning their opinions.

Although we assumed that previous literature reviews and qualitative studies have provided us with sufficient materials to select potential candidate dimensions, it is possible that new health dimensions may emerge during our interviews. Therefore, we will adopt a systematic, cyclical and iterative procedure in terms of data collection and data analysis.<sup>45</sup> We will analyse the collected data in the process of data collection and have internal discussion to check if any health dimensions emerging in the collected data can be developed as candidate bolt-on items; if there are any, we will add them in the subsequent interviews.

We will select interviewers (n=3) from the master's students majoring in public health at Fudan University to conduct the interviews with both experts and lay people. They will be trained to be competent in undertaking qualitative interviewing and be familiar with the interviewing topic guide. Each interviewer will perform at least one pilot interview to test the topic guide (boxes 1 and 2) before the formal data collection.

At the end of each in-person interview, a gift valued at approximately RMB200 will be provided to the respondents as compensation for their time. Each interview will be audio recorded and transcribed in Mandarin.



## Box 1 Planned topic guide and interviewer script for health-related quality of life (HRQoL) experts

1. Introduce the background of the interview
  - ⇒ Introduce the interviewer.
  - ⇒ Introduce the purpose of this interview, that is, understand your views about health description, health measurement, EQ-5D-5L and EQ-5D-5L bolt-on items.
  - ⇒ Introduce the procedure of the interview, that is, (1) they will first be provided with EQ-5D-5L, then will be asked a few questions that are listed below; (2) they will be provided with the initial set of candidate bolt-on items, then will be asked a few probing questions.
  - ⇒ Introduce the length and features of the interview, for example, length about 60–90 min, semistructured interview, no right or wrong answers, etc.
2. Opinions on EQ-5D-5L
  - ⇒ What does HRQoL mean to you?
  - ⇒ Do you think the questionnaire can fully reflect one's HRQoL? Why or why not?
  - ⇒ Do you think there are any aspects of HRQoL that the questionnaire does not include but are important in describing HRQoL?
  - ⇒ If so, what additional questions should be asked? What would you do to improve the questionnaire to make it better reflect your HRQoL?
3. Opinions on the selected bolt-on dimensions
  - ⇒ Our project team conducted a literature review of HRQoL scales in the cultural background of China and carried out qualitative research in the early stage. We have preliminarily selected candidate dimensions. We would like to seek your professional evaluation and suggestions.
  - ⇒ For each bolt-on dimension, we will ask the following structured questions:
    - Content validity: comprehensibility
      1. What is your understanding of 'XXX'?
      2. Do you think the expression of 'XXX' can be accepted and understood by the general public? Is there a better way to express it?
      3. Do you think the title and content of the 'XXX' dimension are concise and clear?
      4. Do you think the descriptions of the five levels of 'XXX' can be accepted and understood by the general public and patients? Is there a better way to express them?
      5. Do you think it is easy or difficult to differentiate between the five response levels?
      6. Are the descriptions of the five levels of 'XXX' concise and clear?
    - Content validity: relevance
      1. Do you think 'XXX' is relevant to the concept of HRQoL? Why or why not?
      2. Do you think 'XXX' is relevant to the Chinese population? Why or why not?
      3. Do you think 'XXX' is relevant for use in measuring HRQoL in China? Why or why not?
      4. Do you think 'XXX' is suitable for describing different HRQoL?
      5. Can 'XXX' reflect differences in the HRQoL of different populations (eg, gender, age groups) or different diseases and their severity?
    - Alignment with standard EQ-5D-5L
      1. (When appropriate) Is the provided example explanation in brackets appropriate? Is there anything you would remove or add or revise?
      2. Do you think the structure and description of the 'XXX' dimension are consistent with the representation of the standard EQ-5D-5L?
    - Descriptive validity

Continued

## Box 1 Continued

1. Do you think 'XXX' is suitable for valuation of health?
2. Do you think it will be easy or difficult for members of the general population to imagine the response levels as statements in hypothetical health state descriptions?
3. Is there anything you would change to make the item or its response levels more suitable for valuation?
4. Additional bolt-on dimensions
  - ⇒ In your opinion, which candidate dimensions are the most important in reflecting the health status of the Chinese population?
  - ⇒ Do you think there is any overlap between these bolt-on dimensions and the standardised EQ-5D-5L?
  - ⇒ Would you like to propose any bolt-on dimensions in addition to the ones that we have discussed?

## Data analysis

We will conduct thematic analysis to analyse interview transcripts.<sup>46</sup> Content analysis will be used to analyse responses to closed-ended questions (specifically, yes or no questions on the content validity, alignment and language of bolt-on items). This technique focuses on the frequency of specific content or response options, enabling to categorise and quantify the responses based on predefined categories or response choices. Two senior scientists and two master's students, majoring in public health, will first read all the transcripts to be familiar with the content and develop the preliminary coding framework. The two students, both will be trained in qualitative data analysis, will be responsible to conduct the analysis. They will independently code five transcripts. A collective review and discussion of the codes for the initial transcript will be conducted to address and resolve any disagreements that may arise. After the research team reaches a consensus by discussing five transcripts, the two students will independently code the remaining transcripts. Any differences in coding will again be discussed and resolved by consensus in the presence of a senior researcher, ensuring the reliability and trustworthiness of the coding process.

Based on the opinions of experts, healthy general populations as well as patients, we will have internal discussions with the research team members to determine a list of candidate bolt-on items and refine their wording. We estimate to identify around 5–10 items. Subsequently, we will submit the list to the EuroQol Group and consult with the experts to further confirm the appropriateness of the candidate bolt-on items for use in quantitative psychometric tests.

## Quantitative psychometric tests to evaluate the acceptability and measurement properties

We will then evaluate the acceptability and measurement properties of the candidate bolt-on items in both the general population and patient groups. We will learn from the methods adopted in previous bolt-on studies in various general population and patient samples to test the measurement properties.<sup>25 26</sup>

## Box 2 Planned topic guide and interviewer script for non-experts

1. Introduce the background of the interview
  - ⇒ Introduce the interviewer.
  - ⇒ Introduce the purpose of this interview, that is, understand your views about health, the questionnaire (EQ-5D-5L) and potential culturally relevant items in addition to EQ-5D-5L.
  - ⇒ Introduce the procedure of the interview, that is, (1) they will first be asked a few open-ended questions; (2) they will be provided with EQ-5D-5L, then will be asked a few questions that are listed below; (3) they will be provided with the additional items, then will be asked a few questions.
  - ⇒ Introduce the length and features of the interview, for example, length about 60–90 min, semistructured interview, no right or wrong answers, etc.
2. Understanding of health
  - ⇒ Can you tell me how you would rate your own health on a scale of 1–10, with 1 and 10 representing the worst and best imaginable health, respectively? Why do you say that?
  - ⇒ Can you think of someone with poor health? What makes you think they have poor health?
  - ⇒ Can you think of someone with good health? What makes you think they have good health?
  - ⇒ What health issues do you think have the greatest impact on your quality of life?
3. Opinions on EQ-5D-5L after they complete the questionnaire
  - ⇒ Do you think the questionnaire can fully reflect your health? Why or why not?
  - ⇒ Do you think there are any aspects of health that the questionnaire does not include but are important in describing health?
  - ⇒ If so, what additional questions should be asked? What would you do to improve the questionnaire to make it better reflect your health?
4. Opinions on the selected additional dimensions
  - ⇒ Our project team has initially selected a few candidate dimensions through preliminary preparatory work. We would like to know your views on these dimensions.
  - ⇒ For each additional dimension, we will ask the following questions:
 

Content validity: comprehensibility

    1. What is your understanding of 'XXX'? Using your own words, how would you explain what this question means?
    2. Do you think the expression of 'XXX' can be understood by you? Is there a better way to express it?
    3. Do you think the title and content of the 'XXX' dimension are concise and clear?
    4. What caused you to choose this response? Would you ever choose level N (N=1–5)? Why or why not? Can you describe an experience where you might choose level N?
    5. Do you think the descriptions of the five levels of 'XXX' can be understood by you? Is there a better way to express them?
    6. Are the descriptions of the five levels of 'XXX' concise and clear?

Content validity: relevance

    1. Do you think 'XXX' is closely related to your health status\*? Why or why not?
    2. Do you think 'XXX' is closely related to the health status of the people around you?
    3. Do you think 'XXX' is closely related to the health status of a healthy/unhealthy person?
    4. Do you think 'XXX' is suitable for describing different health status?

Continued

## Box 2 Continued

5. Can 'XXX' reflect differences in the health status of different populations or different diseases and their severity?

Alignment with standard EQ-5D-5L

1. (When appropriate) Is the provided example explanation in brackets appropriate? Is there anything you would remove or add or revise?
2. Do you think the structure and description of the 'XXX' dimension are consistent with the representation of the standard EQ-5D-5L?
5. Additional dimensions
  - ⇒ In your opinion, which candidate dimensions are more important in reflecting the health status of the Chinese population?
  - ⇒ Do you think there is any overlap between these health dimensions and the standardised EQ-5D-5L?
  - ⇒ Do you think there are any other important dimensions to consider?

\*For ease of understanding by non-experts, we use the term 'health status' instead of 'HRQoL'.

## Sampling

We will recruit two groups of respondents. First, we will recruit a representative sample of the general Chinese population in terms of age, gender, education level and urban-rural residence. Considering the greater heterogeneity of the population distribution in China, and based on the experience of previous studies conducted in similar general population settings,<sup>37 38</sup> we plan to recruit a representative sample of 1000 members of the general population from the eastern, central and western cities of China, such as Tianjin (north), Wuhan (central), Shanghai (east), Guangzhou (south), Harbin (north-east), Chengdu and Guiyang (southwest), as well as the surrounding rural areas of these cities, using a quota sampling method based on the population structure of the seventh national census.<sup>47</sup>

We will also recruit patients with diabetes, cancer and cardiovascular disease, which are three common chronic diseases in China,<sup>48</sup> in two or three hospitals in Shanghai. The reason for selecting patients with chronic diseases is because QALY is commonly used to measure health outcomes in these populations. More specifically, we will recruit patients with type 2 diabetes, as China is the country with the highest number of diabetics worldwide<sup>49</sup>; we will recruit patients with hypertension and at least 50% of them are expected to have moderate to severe cardiovascular comorbidities, and this is because hypertension is the most common risk factor for developing cardiovascular conditions and there are 245 million patients with hypertension in China as reported in 2020<sup>50</sup>; we will also recruit patients with cancer (including patients with lymphoma and myeloma) due to cancer's substantial impact on patients' HRQoL as well as its increasing prevalence and disease burden. For the three disease populations (diabetes, cancer, cardiovascular disease), based on the previous sample size calculation and empirical research evidence,<sup>51</sup> approximately 200 patients with each disease will be recruited. We aim to have a balanced

sample of patients with each disease in terms of severity (mild, moderate and severe).

The inclusion criteria for the general population and patients are: (1) aged 18 years or older; (2) able to understand Mandarin; (3) cognitively able to understand the question; (4) willing to provide informed consent; and (5) the diagnosis of a relevant chronic disease by a physician for the patient sample. The exclusion criteria are: (1) illiterate; (2) poor vision or hearing, since we will use a self-complete questionnaire survey and communication with data collectors is necessary; (3) unable to understand the survey questions; and (4) unable or unwilling to provide written consent.

For the recruitment of patients, we will contact staff in two or three tertiary hospitals in Shanghai to access groups of potential patients who meet inclusion criteria. For recruiting people from the general public, we will work with local collaborators in the above-mentioned cities. We will recruit individuals using our own social network to reach out potential respondents and we will ensure that data collectors do not know the respondents directly. At the end of each data collection, a gift valued at approximately RMB100 will be provided to the respondents as compensation for their time. Informed consent will be obtained from each respondent before each survey starts.

### Data collection

We will collect cross-sectional data using a self-complete paper survey for general population. The patient group will be followed up to collect two-time-point data using a self-complete paper survey.

Each respondent will self-complete the EQ-5D-5L, the questions on each candidate bolt-on item as well as a general health rating question, providing responses that span from 'excellent' to 'poor', while data collectors will be present in case the respondents have any questions about the survey. SF-6Dv2, which is another recommended QALY elicitation tool for use in economic evaluations in China,<sup>52 53</sup> will also be included in the survey to test the psychometric properties of candidate bolt-on items. In addition, patients will be asked to complete different disease-specific scales depending on their disease, for example, the EORTC QLQ-C30<sup>54</sup> will be used in patients with cancer; Audit of Diabetes-Dependent Quality of Life<sup>55</sup> will be used in patients with diabetes; and Hypertension Quality of Life Questionnaire (MINICHAL)<sup>56</sup> will be used in patients with hypertension. A health and well-being measure, the EuroQol Health and Wellbeing instrument (EQ-HWB), will be used among respondents from both the patient group and the general population.<sup>57</sup> The patient group will be followed up to complete the survey within a timeframe of 1–2 weeks after the initial interview to check test–retest reliability.

After completing the survey, the respondents will be asked questions on the acceptability of the candidate bolt-on items. We will ask them to respond to the following features, 'The item is clearly understandable', 'The item is relevant to my health condition', 'The item is easy to

complete', by indicating strongly disagree, disagree, neutral, agree or strongly agree. They will also be asked to provide sociodemographic information such as age, gender, education level and income. For the patients, clinical information including disease severity, disease duration, presence and types of comorbidities, current treatment, outpatient and hospitalisation information will be collected.

### Analysis

We will analyse data from the general population and patient population separately. The distribution of socio-demographic and clinical information will be presented. For testing candidate bolt-on items, we will start from testing each bolt-on item separately before testing combinations of bolt-on items.

Descriptive analyses will be conducted on the distribution of responses on the standard EQ-5D-5L dimensions and the candidate bolt-on items. Ceiling effect (ie, the proportion of respondents who report no problems across dimensions) will be compared across the EQ-5D-5L and EQ-5D-5L+bolt-ons. Additionally, we will examine the distribution of responses to the candidate bolt-on items among those who report no problems across all five standard EQ-5D-5L dimensions ('11111'). We will also obtain the proportion of respondents who report not full health according to EQ VAS score (<100), despite reporting no problems in each of the five EQ-5D-5L dimensions and candidate bolt-on items.

We will then compare all dimensions' classification efficiency in differentiating the participants using the Shannon index ( $H'$ ),<sup>58</sup> which is defined as:

$$H' = - \sum_{i=1}^L p_i \log_2 p_i$$

where  $L$  is the number of descriptive levels within an item and  $p_i$  is the percentage of individuals reporting the  $i$ th level ( $i=1 \dots L$ ). Higher  $H'$  values indicate better classification efficiency. In the case of individuals being evenly distributed among all levels,  $H'$  reaches its maximum. Shannon evenness index ( $J'$ ), which is defined as  $J'=H'/H'_{\max}$ , will also be used to show the relative informativity of each dimension. We will compare the  $H'$  and  $J'$  values across the five core EQ-5D-5L dimensions and the candidate bolt-on items. To compare the informativity between the original EQ-5D-5L instrument with and without candidate bolt-on items, we will follow two approaches. We will calculate the  $H'$  and  $J'$  values by the instrument as a whole (ie, treating each health state as a unique category) and also the average  $H'$  and  $J'$  values of individual EQ-5D-5L dimensions, with and without candidate bolt-on items. The descriptive system with candidate bolt-on items would have higher  $H'$  and  $J'$  values if the candidate bolt-on items improve the informativity.

### Acceptability of candidate bolt-on items

We will obtain the percentage of respondents reporting difficulty in understanding the bolt-on questions to assess



the acceptability of candidate bolt-on items. We will also calculate the average importance rating score for each EQ-5D-5L item and candidate bolt-on item.

### Test–retest reliability

To assess the test–retest reliability, we will first obtain an intraclass correlation coefficient of the level sum score, where a value less than 0.50 indicates poor reliability, a value between 0.50 and 0.74 indicates moderate reliability, between 0.75 and 0.90 indicates good reliability and a value greater than 0.90 indicates excellent reliability, based on the data collected from patients at two different time points.<sup>59</sup> We will apply the commonly used indicator weighted kappa ( $\kappa$ ) to determine the test–retest reliability for each EQ-5D-5L and candidate bolt-on items, which are categorical variables. We will rely on the guidance to define kappa: 0–0.40 suggests poor to fair, 0.41–0.60 moderate, 0.61–0.80 substantial, 0.81–1.00 almost perfect.<sup>60</sup>

### Construct validity

We will use the EQ VAS score as the dependent variable and the standard EQ-5D-5L dimensions and candidate bolt-on items as independent variables to explore which items improve the explanatory power in EQ VAS scores, controlling sociodemographic characteristics.

We will also investigate convergent validity, which focuses on the convergence of two items that are intended to measure the similar or related constructs. The convergent validity of candidate bolt-on items will be established if the Spearman's  $r$  correlation coefficients between candidate bolt-on items and existing EQ-5D-5L as well as the dimensions of SF-6Dv2 and other condition-specific measures (where the overlap is expected) show at least moderate correlation ( $>0.3$ ). On the other hand, the divergence between candidate bolt-on items and the existing EQ-5D-5L items will be assessed, where two items are expected to measure different constructs (a low correlation  $<0.4$ ).<sup>25</sup>

Known-groups validity, which refers to the ability to distinguish people with different HRQoL, will also be assessed in the study. It will be tested based on comparing the scaled level sum scores (transformed to a 0–100 scale) of the standard EQ-5D-5L and EQ-5D-5L plus candidate bolt-on items across known groups. We will compare the level sum score across groups of patients based on relevant clinical characteristics, for example, with different degrees of illness severity, as well as patients with and without comorbidities. For the general population, we will group people based on their age, their self-rating of the general health status and whether they have any chronic condition. It is assumed that candidate bolt-on items will improve the ability to distinguish between the known groups and it will be tested by the F-statistic from the analysis of variance.

### Bolt-on selection

Having discussed internally based on all evidence collected, we will select bolt-on items relying on the combination of both the qualitative and quantitative evidence gathered. Among the candidate items, bolt-ons that perform well in terms of content validity, acceptability and psychometric performance in improving the standard EQ-5D-5L can be selected and used for patient and population health measurement. In a follow-up project, a few of these candidate items may be selected for a future potential valuation experiment, which would aim to assess the impact of adding selected bolt-on items on the valuation of EQ-5D-5L health states. The selected bolt-on items will be submitted to the EuroQol Group and reviewed by the experts for their suggestions.

## DISCUSSION

Recent studies have argued that the descriptive system of EQ-5D is not comprehensive enough to sufficiently describe and measure HRQoL in China.<sup>7 21</sup> This study proposes that, by adding culturally relevant bolt-on items, we can both retain the original descriptive system of EQ-5D and make EQ-5D more sensitive to the local context. This protocol outlines a comprehensive process for identifying, developing, testing and selecting bolt-on items that are relevant and culturally appropriate in China. It also has important methodological implications, as it may serve as a guide for developing and testing culturally relevant bolt-on items in other cultural contexts.

This study has the following advantages. First, the bolt-on item development process will follow EuroQol-recommended criteria,<sup>25</sup> which ensures that the bolt-on items can have the best acceptability and measurement performance. Second, the development process uses both qualitative and quantitative techniques and is based on a thorough review of the literature as well as expert opinions and empirical data from both the general and patient populations. By using evidence from both sources, the protocol provides a comprehensive and evidence-based approach to item development. This ensures that the selected items are appropriate and valid and helps minimise potential biases that may arise from relying on a single source of information. Third, this study will be the first to systematically compare the EQ-5D-5L plus culturally relevant bolt-on(s) with the standard EQ-5D-5L in terms of descriptive statistics, measurement properties, so as to assess whether and to what extent the bolt-on EQ-5D-5L can outperform the EQ-5D-5L, which is a significant research gap that has not been addressed in previous cultural bolt-on studies in Thai<sup>37</sup> or Korean<sup>38</sup> populations. By addressing this gap, this protocol can also provide a more complete understanding of how to measure health in diverse cultural contexts.

Despite its advantages, there are some potential limitations associated with the use of this protocol. One limitation is that some of the bolt-on development criteria, for example, the criteria on 'bolt-on descriptors should be

translatable' and 'the direction of dimension wording should be the same as the core EQ-5D-5L dimensions (ie, negative)',<sup>25</sup> are not fully applicable or relevant to the cultural bolt-on contexts. We, therefore, need to adapt the criteria to fit our study. Another limitation is that the empirical data collected from patients and the general population may not be representative of the whole Chinese population. Although we aim to recruit participants from different regions of China using a stratified quota sampling method, we may fail to recruit people from very remote areas and/or ethnic minority populations, which is also one of the limitations of the Chinese EQ-5D-5L valuation study. Additionally, since we will only obtain patient data from three health conditions, it remains unclear how the bolt-on items may perform in some other health conditions with specific symptoms. Moreover, the responsiveness of the bolt-on items will not be assessed due to the study design. In addition, this study will not assess the impact of the developed bolt-on items on the valuation of health states. This is planned to be addressed in a follow-up project. Despite the limitations, this study can provide valuable insights into how to develop culturally relevant bolt-on items before assessing the impact of adding bolt-on items to the valuation of EQ-5D in the future.

In summary, this protocol documents a systematic and thorough process in terms of identifying, developing, testing and selecting culturally relevant bolt-on items for EQ-5D-5L in China. As the future next step, we aim to examine the impact of developed bolt-on items on valuation and, in the end, develop an EQ-5D-5L+bolt-on(s) value set that may improve the HRQoL and QALY estimation in China.

## ETHICS AND DISSEMINATION

This study received ethics approval from the Institutional Review Board of School of Public Health, Fudan University (IRB number: 2022-TYSQ-03-154), and will be conducted according to the guidelines of the Declaration of Helsinki. Study findings will be disseminated through international peer-reviewed journal articles as well as public, academic presentations at national and international conferences.

**Contributors** All authors conceptualised the study design. ZM and JF wrote the protocol. PW, FR, ZY and NL critically reviewed and edited the protocol. All authors approved the final version of the protocol.

**Funding** The study has been sponsored by the National Natural Science Foundation of China (Project No 72274037) awarded to PW.

**Disclaimer** Views expressed in the article are those of the authors and are not necessarily those of the EuroQol Research Foundation.

**Competing interests** ZM, FR, ZY and NL are members of the EuroQol Group.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

Zhuxin Mao <http://orcid.org/0000-0002-7444-1584>

## REFERENCES

- 1 EuroQol Group. Euroqol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- 2 Brooks R. Euroqol: the current state of play. *Health Policy* 1996;37:53–72.
- 3 Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
- 4 Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ* 2020;21:1245–57.
- 5 Rencz F, Gulácsi L, Drummond M, et al. EQ-5D in central and Eastern Europe: 2000–2015. *Qual Life Res* 2016;25:2693–710.
- 6 Mao Z, Ahmed S, Graham C, et al. Similarities and differences in health-related quality of life concepts between the East and the West: a qualitative analysis of the content of health-related quality of life measures. *Value Health Reg Issues* 2021;24:96–106.
- 7 Mao Z, Ahmed S, Graham C, et al. Exploring subjective constructions of health in China: a Q-methodological investigation. *Health Qual Life Outcomes* 2020;18:165.
- 8 Zhou T, Guan H, Liu G, et al. Health-related quality of life for disease population in China based on EQ-5D: a systematic review. *Chinese J Evidence-Based Medicine* 2016;16:135–42.
- 9 Zhou T, Guan H, Liu G, et al. Health-related quality of life for the general population in China: a systematic review. *Chinese Health Service Management* 2016;33:621–3.
- 10 Liu G, Hu S, Wu J, et al. *China Guidelines for Pharmacoeconomic Evaluations*. Beijing: China Market, 2020.
- 11 Sun S, Chen J, Johannesson M, et al. Population health status in China: EQ-5D results, by age, sex and socio-economic status, from the national health services survey 2008. *Qual Life Res* 2011;20:309–20.
- 12 Tan Z, Liang Y, Liu S, et al. Health-related quality of life as measured with EQ-5D among populations with and without specific chronic conditions: a population-based survey in Shaanxi province, China. *PLoS One* 2013;8:e65958.
- 13 Janssen MF, Szende A, Cabases J, et al. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. *Eur J Health Econ* 2019;20:205–16.
- 14 Yang Z, Busschbach J, Liu G, et al. EQ-5D-5L norms for the urban Chinese population in China. *Health Qual Life Outcomes* 2018;16:210.
- 15 Xie S, Wu J, Xie F. Population norms for SF-6Dv2 and EQ-5D-5L in China. *Appl Health Econ Health Policy* 2022;20:573–85.
- 16 Van Wilder L, Charafeddine R, Beutels P, et al. Belgian population norms for the EQ-5D-5L, 2018. *Qual Life Res* 2022;31:527–37.
- 17 Ferreira PL, Pereira LN, Antunes P, et al. EQ-5D-5L Portuguese population norms. *Eur J Health Econ* 2023;24:1411–20.
- 18 Jiang R, Janssen MFB, Pickard AS. US population norms for the EQ-5D-5L and comparison of norms from face-to-face and online samples. *Qual Life Res* 2021;30:803–16.
- 19 Liu Y, Wan C, Xi X. Measurement properties of the EQ-5D-5L in sub-health: evidence based on primary health care workers in China. *Health Qual Life Outcomes* 2023;21:22.
- 20 Zhu W, Zhang M, Pan J, et al. Valuing Chinese medicine quality of Life-11 dimensions (CQ-11D) health states using a discrete choice experiment with survival duration (DCETTO). *Health Qual Life Outcomes* 2023;21:99.
- 21 Wu J, He X, Chen P, et al. China health related outcomes measures (CHROME): development of a new generic preference-based measure for the Chinese population. *Pharmacoeconomics* 2022;40:957–69.
- 22 Zhu W, Gao H, Zhang M, et al. Development of the Chinese medicine life quality evaluation scale (CQ-11D). *China J Pharmac Eco* 2022;17:10–5.
- 23 Liu GG, Wu H, Li M, et al. Chinese time trade-off values for EQ-5D health states. *Value Health* 2014;17:597–604.



- 24 Luo N, Liu G, Li M, *et al.* Estimating an EQ-5D-5L value set for China. *Value Health* 2017;20:662–9.
- 25 Mulhern BJ, Sampson C, Haywood P, *et al.* Criteria for developing, assessing and selecting candidate EQ-5D bolt-ons. *Qual Life Res* 2022;31:3041–8.
- 26 Geraerds A, Bonsel GJ, Janssen MF, *et al.* Methods used to identify, test, and assess impact on preferences of bolt-ons: a systematic review. *Value Health* 2021;24:901–16.
- 27 Kangwanrattanukul K, Phimarn W. A systematic review of the development and testing of additional dimensions for the EQ-5D descriptive system. *Exp Rev Pharmacoeconomi Out Res* 2019;19:431–43.
- 28 Wolfs CAG, Dirksen CD, Kessels A, *et al.* Performance of the EQ-5D and the EQ-5D+ C in elderly patients with cognitive impairments. *Health Qual Life Outcomes* 2007;5:1–10.
- 29 Krabbe PF, Stouthard ME, Essink-Bot ML, *et al.* The effect of adding a cognitive dimension to the Euroqol multiattribute health-status classification system. *J Clin Epidemiol* 1999;52:293–301.
- 30 Swinburn P, Lloyd A, Boye KS, *et al.* Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: lessons learned from a feasibility study in the UK. *Value Health* 2013;16:1156–62.
- 31 Rencz F, Mukuria C, Bató A, *et al.* A qualitative investigation of the relevance of skin irritation and self-confidence bolt-ons and their conceptual overlap with the EQ-5D in patients with psoriasis. *Qual Life Res* 2022;31:3049–60.
- 32 Wang P, Chong S-L, Tan R-Y, *et al.* A hearing bolt-on item increased the measurement properties of the EQ-5D-5L in a community-based hearing loss screening program. *Eur J Health Econ* 2023;24:393–8.
- 33 Gandhi M, Ang M, Teo K, *et al.* A vision 'bolt-on' increases the responsiveness of EQ-5D: preliminary evidence from a study of cataract surgery. *Eur J Health Econ* 2020;21:501–11.
- 34 Finch AP, Brazier J, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: testing the impact of hearing, sleep, cognition, energy, and relationships on preferences using pairwise choices. *Med Decis Making* 2021;41:89–99.
- 35 Perneger TV, Courvoisier DS. Exploration of health dimensions to be included in multi-attribute health-utility assessment. *Int J Qual Health Care* 2011;23:52–9.
- 36 Spronk I, Polinder S, Bonsel GJ, *et al.* Adding a fatigue item to the EQ-5D-5L improves its psychometric performance in the general population. *J Patient Rep Outcomes* 2022;6:1.
- 37 Kangwanrattanukul K, Gross CR, Sunantiwat M, *et al.* 'Exploration of a cultural-adaptation of the EQ-5D for Thai population: a 'bolt-on' experiment'. *Qual Life Res* 2019;28:1207–15.
- 38 Kim S-H, Jo M-W, Ock M, *et al.* Exploratory study of dimensions of health-related quality of life in the general population of South Korea. *J Prev Med Public Health* 2017;50:361–8.
- 39 Lim C, Shafie A, Thakumar A. Exploration of EQ-5D-5L bolt-on items among Malaysian population. *Malaysian J Pharm (MJP)* 2017;3:8–17.
- 40 Ding Y, Yang Z, Busschbach J, *et al.* Do EQ-5D dimensions reflect the Chinese perception of health-related quality of life? A systematic review. 3rd EuroQol Early Career Researcher (ECR) Meeting; Milan, Italy.2023
- 41 Yang Y, Brazier J, Tsuchiya A. 'Effect of adding a sleep dimension to the EQ-5D descriptive system: a 'bolt-on' experiment'. *Med Decis Making* 2014;34:42–53.
- 42 Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
- 43 Terwee CB, Prinsen C, Chiarotto A, *et al.* COSMIN Methodology for Assessing the Content Validity of PROMs–User Manual. Amsterdam: VU University Medical Center, 2018.
- 44 Terwee CB, Prinsen CAC, Chiarotto A, *et al.* COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
- 45 Creswell JW, Poth CN. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage publications, 2016.
- 46 Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psycho* 2006;3:77–101.
- 47 Office of the leading group of the state Council for the seventh national population census. *China Population Census Yearbook* 2020. 2020.
- 48 National Health Commission of the People's Republic of China. Report on Chinese residents' chronic diseases and nutrition 2020. Beijing; 2020.
- 49 Li Y, Teng D, Shi X, *et al.* Prevalence of diabetes recorded in Mainland China using 2018 diagnostic criteria from the American diabetes Association: national cross sectional study. *BMJ* 2020;369:m997.
- 50 National Center for Cardiovascular Diseases China. Annual report on cardiovascular health and diseases in China. Beijing; 2020.
- 51 Kangwanrattanukul K, Gross CR, Sunantiwat M, *et al.* Adding two culture-specific 'bolt-on' dimensions on the Thai version of EQ-5D-5L: an exploratory study in patients with diabetes. *Expert Rev Pharmacoecon Outcomes Res* 2019;19:321–9.
- 52 Wu J, Xie S, He X, *et al.* Valuation of SF-6Dv2 health States in China using time trade-off and discrete-choice experiment with a duration dimension. *Pharmacoeconomics* 2021;39:521–35.
- 53 Brazier JE, Mulhern BJ, Bjorner JB, *et al.* Developing a new version of the SF-6D health state classification system from the SF-36V2: SF-6Dv2. *Med Care* 2020;58:557–65.
- 54 Fayers P, Bottomley A, EQoL G. Quality of life research within the EORTC—the EORTC QLQ-C30. *Europ J Cancer* 2002;38:125–33.
- 55 Bradley C, Todd C, Gorton T, *et al.* The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the Addqol. *Qual Life Res* 1999;8:79–91.
- 56 Badia X, Roca-Cusachs A, Dalfó A, *et al.* Validation of the short form of the Spanish hypertension quality of life questionnaire (MINICHAL). *Clin Ther* 2002;24:2137–54.
- 57 Brazier J, Peasgood T, Mukuria C, *et al.* The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. *Value Health* 2022;25:482–91.
- 58 Janssen MF, Birnie E, Bonsel GJ. Evaluating the discriminatory power of EQ-5D, Hui2 and Hui3 in a US general population survey using Shannon's indices. *Qual Life Res* 2007;16:895–904.
- 59 Koo TK, Li MY. A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
- 60 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159.