## Introduction

Arabic is the fifth most spoken language in the world, with an estimated 420 million native speakers across the world (~ 5% of the world's population), and is the official language in 27 countries spread throughout the Middle East and North Africa (MENA) region.[1,2] Spoken Arabic varies from one country to another in both dialect and vocabulary; however, standard Arabic (spoken and written Fus-ha) is the same across all Arabic-speaking countries.

Preference-based generic measures of health-related quality of life (HRQoL) are used to derive "utility" weights, which are essential building blocks in health economic evaluation studies that use quality-adjusted life years (QALYs) as a measure of outcomes, i.e. cost-utility analyses (CUAs).[3] Given the increased interest in the use of pharmacoeconomic evaluations in setting pricing, market access and reimbursement decisions in the MENA region, it has become essential to develop the "toolkits" needed to conduct such evaluations, including HRQoL measures.

The SF-6D (Short-Form Six-Dimension) is a generic, preference-based measure of HRQoL that has been originally developed in English from the SF-36, a much longer, widely used measure of health. The SF-36 has previously been translated into Arabic.[4-6] The SF-6D is derived from the SF-36 HRQoL measure, as a six-dimensional health state classification. The six dimensions are: physical functioning, role limitation, social functioning, bodily pain, mental health and vitality. Each of the six dimensions consists of a number of levels ranging from four to six levels. The SF-6D represents an "off-the-shelf" measure that comes with a readily-available set of preference values elicited from the general public, hence, can be used in CUA studies.[7] The preference elicitation technique used is the "Standard Gamble" (SG); which aims at measuring

the 'utility' of a health state described by the six dimensions by eliciting the willingness to accept a certain level of risk of death in order to avoid living in that state (and living in full health).

The SF-6D can also be used for analyzing existing SF-36 data when there is no other means of estimating the preference-based health utility values for generating QALYs. It also provides an alternative to existing preference-based measures for CUA such as the EQ-5D measure and the health utilities index (HUI).[8,9] The EQ-5D is currently the most commonly used preference-based measure of HRQoL and although it has been translated into Arabic with demonstrated validity and reliability;[10,11] SF-6D has been shown in many studies to be a tool with a lower ceiling effect and greater sensitivity.[12-14] Significant differences have been demonstrated between the two measures, and they can produce quite different utility scores from the same subject.[15-17]

Currently, there is no Arabic version of SF-6D available. Thus, the aim of this study was to translate, culturally adapt and validate the SF-6D for use in research studies across the Arabic-speaking countries in general.

## Methods

### *Translation process*

Translation of the questionnaire was performed according to international guidelines for translational studies; where the International Quality of Life Assessment (IQOLA) methodology was followed.[18,19] This involved: 1) Two forward translations, 2) A consensus version, 3) Committee review and validation of the forward translation, 4) One back translation and 5) Committee review of the back-translation. Forward translation of the original questionnaire from English to Arabic was carried out by two independent qualified linguistic translators (fluent in both Arabic and English). The two translations were then used to develop a consensus Arabic

version by agreement between the translators. Where disagreements occurred, these were resolved by discussion with the research team members. An advisory committee consisting of nine members including clinicians and health outcomes researchers fluent in both English and Arabic reviewed the consensus version and provided feedback on the accuracy of the translation, compatibility with Arabic-speaking countries' cultures and traditions and on the cognitive burden associated with the different items. Comments received from the committee members were categorized and discussed. Agreed changes were made and the forward translation was finalized. Backward translation of this consensus version from Arabic to English was then carried out by another qualified independent translator. The back-translated version was evaluated by the committee for similarity of the instructions, items and response format regarding wording, sentence structure, meaning, and relevance. The difficulties encountered by the translators and highlighted by the committee members during the translation and cultural adaptation process were categorized as grammatical, idiomatic, semantic/conceptual, and cultural.[20,21] This final version was also checked and approved by the developer (Professor John Brazier) through reviewing the overall process and the backwards translation.

***Validation study***

Participants from Arabic-speaking countries including Egypt, Jordan, UAE, Qatar and Palestine were invited via social media platforms including Facebook and WhatsApp to complete our Arabic-translated SF-6D online if they were older than 18 years and were literate (i.e. speak, read, and write in the Arabic language). Snowball sampling technique was used where two of the authors (DD, FE) invited all their social networking contacts to complete the questionnaire and share it each with their own contacts. Informed consent was obtained electronically from each participant prior to completing the questionnaire. In addition, participants were informed of the

research objectives, confidentiality of their responses, and their right to withdraw from the study. Based on the sample-size calculation from Krejcie and Morgan, a minimum of 385 respondents were required to achieve a 95% level of confidence and 5% significance level for up to 420 million population of Arabic speaking countries.[22] Respondents were also asked a question about their self-related health state, on a scale of 0 to 100. The data were collected over a period of one week in August 2018. The validation study was approved by Jordan University Science and Technology Institutional Review Board (Ref: 59/117/2018).

***Statistical analysis***

Demographic data were collected, including age, gender, education level and presence of chronic disease (as reported by the respondents), and descriptive statistics were presented to summarize the study variables of interest as counts and percentages for the categorical variables and as means and standard deviations for the continuous ones. Chi-square ($\chi 2$) and independent t-tests were used to chart comparisons between two categorical and continuous variables respectively.. Content validity tests were conducted, and Cronbach's alpha test of internal consistency was used for the measurement of reliability. Confirmatory factor analysis (CFA), including goodness-of-fit tests and different incremental fit measures, was also used to assess construct validity and test a pre-specified relationship between items. This is a statistical method that assesses the dimensionality of a latent construct. It can be also used to verify SF-6D validity by other criteria, in this case a question asking the respondent about his/her self-related health status, on a scale of 0 to 100. Loading values $\geq 0.35$ were considered acceptable, based on current literature.[23,24]

The discriminant validity of the translated SF-6D was also assessed by calculating the invariance in the construct structure of the study model between respondents who reported having chronic

disease and those who did not. Statistical analysis was conducted using PASW 18 (SPSS Inc, Chicago, USA) and IBM SPSS Amos 23 (IBM, New York, USA). All reported *p*-values were based on two-sided tests and were compared with a significance level of 5%.

## Results

### *Translation:*

Overall, no major problems were encountered during the forward translations. However, the two translators reported encountering minor difficulties with a few words and phrases. Some were related to the inherent differences between English and Arabic language in terms of sentence structure and grammar. Others were related to idiomatic and cultural issues.

The following were the main words/phrases that posed particular challenges; classified according to the pre-approved categories:

- *Semantic* (instances=1): "Role Limitations" where further explanation of what the word "role" means was required

- *Grammatical* (instances=1): "None of the time" where there is no Arabic equivalent of this phrase. A negative sentence structure was used instead to give the equivalent meaning

- *Cultural* (instances=2): "Emotional Problems" which when translated literally to Arabic would refer mainly to problems in romantic relationships and would be culturally inappropriate; "Mental Health" when translated literally to Arabic would be stigmatizing, hence, an equivalent Arabic term was used instead.

- *Idiomatic* (instances=1): "downhearted and low" as these idioms have no equivalent in Arabic and hence, alternative terms were used instead to convey the closest meaning.

There were few differences between the two forward translations, which were resolved by discussion and consensus was reached. The advisory committee members unanimously judged the translation to be semantically equivalent and culturally appropriate, following a number of key refinements and suggestions as follows.

*A. Translation issues:*

- Introductory instructions: simplifying the last sentence (instances=2)
- Physical Functioning: adding a verb to explain that the limitation is in "choosing" the "kind of work" (instances=1)
- Role-Limitations: translating the word "emotional" to Arabic is culturally challenging and a change to the word used was suggested, consistently with one of the forward translations (instances=3)
- Social Functioning: use of consistent words to describe the levels (a little, some, most) (instances=1)
- Pain: explaining that "pain" refers to "bodily" pain (instances=1)

*B. Questionnaire Structure:*

Item reduction may be needed due to the cognitive burden in the differentiation between response levels (instances=3)

*C. Cognitive burden*

The overall cognitive burden was assessed as relatively high for the average educational level among respondents in Arabic-speaking countries (instances=4)

- Three members were supportive of adding examples of activities; however, two of them recognized that this may influence respondents' perceptions of their own level of functioning. Another member was against adding these examples for the same reason
- The Role-Limitations domain is cognitively demanding due to combining the effect of the physical and mental health in some items (instances=1)
- Possible confusion between Physical Functioning and Role Limitation domains (instances=1) and between the activities covered by the Physical Functioning, Social Functioning and Vitality domains (instances=1)

The back translation did not reveal major problems in terms of cross-cultural equivalence to the original version. Further refinements were made following the back-translation including minor change to items in the "Role Limitations" (instances =1) and "Mental Health" (instances =1) domains.

Overall, the research team reached consensus on the semantic, grammatical, cultural and idiomatic equivalence of the Arabic translation and the English SF-6D. The resultant version is provided in the supplementary materials.

***Validation***:

A total of 470 participants completed the questionnaire, with almost equal distribution of males and females. The majority of the respondents had a university degree (n=401, 85.32%). Of the respondents, 17.7% had chronic disease(s) (n=83) (Table 1).

**Table 1. Demographic characteristics of participants in the Arabic-translated SF-6D validation study**

Insert Table 1 here

188 ***Construct validity***

189 An initial model that assumes uni-dimensionality of the SF-6D construct did not fit the data well
190 ($\chi 2$= 114.58; Degrees of Freedom = 9, P value ≤ 0.01), which confirmed the multi-dimensional
191 nature of SF-6D.[23] In the case of a model that lacks the acceptable fit indicators, modified
192 indicators should be used to inform whether covariance should be allowed among item errors.
193 Byrne suggested the use of a modified index that includes residuals between two items (± 2.58).[23]
194 In this study, four covariances were added between items that had highest residuals. Specifically,
195 one residual was added between the item errors (e1 – e2, e2 – e5, e4 – e5, e4 – e6) (see for
196 illustration Figure 1 below). After covariance was allowed among item errors, the model fit indices
197 were recalculated. While the new $\chi 2$=12.99 (DF=5, P≤0.05) meant that the measurement model
198 still did not fit the data, the more precise CMIN/DF indicator, recommended by Byrne, [23] showed
199 adequate model fit (CMIN/DF=2.59, which is <3). Additionally, incremental fit indices were
200 calculated for stepwise addition of error covariance. A value close to 1 indicates best model fit.
201 The result shows that the values were (CFI=0.991, GFI=0.992, NFI=0.987, RFI=0.962, IFI=0.992,
202 and TLI=0.976). All the incremental indices values are ≥0.90 and close to 1, which shows the
203 model has very good fit. Furthermore, the root mean square error (RMSEA) value is 0.058, which
204 is less than the cut-off value close to 0.06.[23] Finally, item loading values ranged from 0.52 to 0.87,
205 which are within the acceptable range of high loading values (all >0.50) (Table 2).

206 **Table 2. Loading value for the items on the latent SF-6D construct.**

207 Insert Table 2 here

208 For assessing the relationship between the latent SF-6D construct and self-reported health status,
209 a construct model was established by adding the observed variable for self-reported health status,

and goodness of fit was assessed. For goodness of fit, the value of CMIN/DF=1.97, which is <3, indicates good model fit. Additionally, the model incremental fit indices were calculated. The result showed that the values were (CFI = 0.998, GFI =0.992, NFI 0.985, RFI= 0.969, IFI = 0.993, and TLI = 0.984). All the incremental indices values are ≥0.90 and close to 1, which shows the model fits the data well. Furthermore, the RMSEA value =0.046 is less than the cut-off value close to 0.06.[24]

The Standardized beta coefficient for the association between SF-6D and self-reported health status is 0.74 (t-value 16.421, P≤0.01), which shows a strong association between SF-6D indictors and self-reported health status. Also, the percentage of variation in self-reported health status that is explained by variance on SF-6D is about 55%, which shows a relatively high value of using SF-6D to predict self-reported health status among patients.

**Figure 1. Confirmatory Factor Analysis (CFA) model for the Arabic version of SF-6D**

Insert Figure 1 here

***Discriminant Ability of the Arabic SF-6D Version***

The invariance in the construct structure of study model was assessed between respondents who reported chronic disease and those who did not. The comparison was run in AMOS 23 by creating two data sets to process the analysis. The $\chi 2$ for the measurement model was 42.56 ($P \geq 0.01$) which means that invariance between the two groups on measurement weight could not be rejected. Table 3 shows the loading values for both groups of responders on the measurement model.

**Table 3. Measurement model weights for respondents reporting or not reporting chronic disease**

Insert Table 3 here

As shown in Table 3, loading values (weights) for SF-6D items were greater for those with chronic health conditions (range: 0.68-0.91) than for those without chronic conditions (range: 0.42-0.73) Also, the weight of SF-6D on self-reported health status among respondents with chronic disease was 0.87, while among respondents reporting no chronic disease it was 0.61 (t-value for the difference: -8.93, $P \leq 0.01$). This result reflects the sensitivity of SF-6D to differences in quality of life and health status among patients with chronic health conditions vs. those with no chronic diseases.

**Discussion**

To our knowledge, this is the first Arabic translation of the SF-6D questionnaire. The translation and cultural adaptation of the SF-6D into standard Arabic resulted in a conceptually equivalent,

culturally appropriate, and psychometrically valid version that can be used in Arabic-speaking countries.

Translation, validation and/or valuation studies of the SF-6D have been undertaken in a number of countries including China, Portugal, Australia and Japan.[25-28] These studies have confirmed validity and feasibility of developing local valuation algorithms for the SF-6D. The translation and cultural adaptation to produce the Arabic SF-6D version followed the recommended translation standards used for HRQoL measures to ensure that an equivalently valid measure is obtained. The cognitive burden of this Arabic translation, however, has been reported to be relatively high for the average individual in Arabic-speaking countries. A previous validation study of the SF-6D undertaken in China has also reported that.[25] Thus, developing colloquial versions for interviewer-administration to illiterate individuals could facilitate its use.

The SF-6D is derived from the SF-36 measure, which has been previously translated to Arabic.[29] The availability of this Arabic version of the SF-36 will facilitate the adoption and use of our Arabic SF-6D measure in clinical studies; as it is generally recommended that both SF-36 and SF-6D are administered together as the valuation algorithms are based on the SF-6D derived from the SF-36 and there is no evidence whether this equals a directly administered SF-6D.[30]

In this validation study, the analysis assumed that the SF-6D factorial structure is based on a unidimensional construct, which means that all the indictors are related to each other. CFA was used to test the measurement model of the questionnaire. Based on our results, we suggest a model that interrelates the latent SF-6D construct with the observed self-reported health status. Additionally, in this study the invariance factorial structure was tested according to whether the

respondent has chronic disease or not. The results supported the discriminatory ability of this Arabic translation of SF-6D.

The original SF-6D valuation study applied an innovative standard gamble (SG) protocol with an analytical sample of 611 UK subjects.[4] The health state values differ between countries, so it is often important to have local valuations. The SF-6D preference scores have been derived for a number of countries.[25-28] More recent innovations in the valuation of SF-6D have used a pivoted paired comparison design where a base scenario (or pivot) is introduced and the respondents choose between 2 alternative changes in the base scenario (i.e., paired comparison). By varying only two contrasting attributes at a time, the pivot design reduces cognitive difficulty and response times, similar to partial profile designs.[31,32] Developing a value set/valuation algorithm to accompany this version, perhaps using a pivoted paired comparison design, would be a research priority in the future.

Our Arabic translation of the SF-6D has produced a culturally-adapted version that is psychometrically valid and can be used as HRQoL measure in Arabic-speaking countries, offering an alternative to the most commonly used generic preference-based measure, the EQ-5D. This is an important addition to the toolkit available for researchers conducting pharmaco-economic evaluations in Arabic-speaking countries particularly those focusing on therapeutic areas where the EQ-5D might not be the best measure, for example conditions related to hearing and vision.[33]

However, our study is not without limitations. We conducted the validation study using online administration and hence, our sample might not be representative of all Arabic speaking people particularly those with no internet access or lower level of literacy. However, based on the most

recent surveys of internet access and coverage in the Arabic speaking countries included in our study, we believe that this is unlikely to affect the representativeness of our sample. Nevertheless, undertaking other validation studies using the traditional pen and paper administration method would be useful to see if the administration method could have affected our conclusions.

We have also tested the translated version in a limited number of Arabic-speaking countries and further validation studies would confirm the acceptability of this version in other countries. However, as we have used the standard written Arabic (Fus-ha) which is common across all Arabic-speaking countries, we do not foresee any issues relating to acceptability of this version in other countries. Valuation studies are also underway in a number of Arabic-speaking countries to estimate utility values for the health states defined by this translated version of SF-6D.

## Conclusion

This Arabic translation of SF-6D has maintained conceptual, semantic and cultural validity, as compared to the original English version. There was strong correlation between SF-6D domains and perceived health. The Arabic version also possesses good discriminative ability between groups based on the presence of chronic disease.

## Acknowledgment

The authors would like to thank the project advisory committee members: Dr Layla Al-Hayas, Dr Mohamed Amin, Dr Hemat Moustafa and Dr Moustafa Dawoud. The authors would also like to thank the professional translators: Professor Wasseem Abdel Haleem and Mr Ahmed Amin, Professor Sam Salek for his editorial comments and Professor John Brazier for his input throughout the project.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

1. Simons GF, Gary F, Charles D (eds.). *Ethnologue: Languages of the World*, 21 ed.. Dallas, Tx: SIL International; 2018. Online version: http://www.ethnologue.com. Retrieved 9 Nov, 2018.
2. Versteegh, K. *The Arabic Language*. 2nd Ed. Edinburgh: Edinburgh University Press; 2014.
3. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press; 2005.
4. Brazier J, Roberts J, Deverill M (2002). The Estimation of a Preference-Based Measure of Health from the SF-36. *J Health Econ*. 2002;21(2):271-292.
5. Brazier J, Roberts J. The Estimation of a Preference-Based Measure of Health From the SF-12. *Med Care.* 2004;42:851-859.
6. Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of life in rural and urban populations in Lebanon using SF-36 health survey. *Health and Qual Life Outcomes*. 2003;1:30. doi: 0.1186/1477-7525-1-30
7. Ferreira LN, Ferreira PL, Pereira LN, Rowen D, Brazier JE. Exploring the consistency of the SF-6D. *Value Health*. 2013;16(6):1023-31. doi: 10.1016/j.jval.2013.06.018.
8. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37:53-72.

9. Horsman,J, Furlong W, Feeny, D, Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1: 54. doi: 10.1186/1477-7525-1-54
10. Bekairy AB, Bustami RT, Almotairi M et al. Validity and reliability of the Arabic version of the the EuroQOL (EQ-5D). A study from Saudi Arabia. *Int J Health Sci (Qassim)*. 2018;12(2):16–20.
11. Aburuz S, Bulatova N, Twalbeh M, Gazawi M. The validity and reliability of the Arabic version of the EQ-5D: A study from Jordan. *Ann Saudi Med.* 2009;29:304-8.
12. Wu J, Han Y, Zhao FL, Zhou J, Chen Z, Sun H. Validation and comparison of EuroQoL-5 dimension (EQ-5D) and Short Form-6 dimension (SF-6D) among stable angina patients. *Health Qual Life Outcomes*. 2014;12:156. doi:10.1186/s12955-014-0156-6
13. Torrance N, Lawson KD, Afolabi E et al. Estimating the burden of disease in chronic pain with and without neuropathic characteristics: Does the choice between the EQ-5D and SF-6D matter? *Pain*. 2014;155(10):1996–2004. doi: 10.1016/j.pain.2014.07.001
14. Teckle P, McTaggart-Cowan H, Van der Hoek K et al. Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health Qual Life Outcomes*. 2013; 11: 203. doi: 10.1186/1477-7525-11-203
15. Shiroiwa T, Fukuda T, Ikeda S, et al. Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D. *Qual Life Res*. 2015;25(3):707-19. doi: 10.1007/s11136-015-1108-2
16. Yang F, Lau T, Lee E, Vathsala A, Chia KS, Luo N. Comparison of the preference-based EQ-5D-5L and SF-6D in patients with end-stage renal disease (ESRD). *Eur J Health Econ.* 2015;16(9):1019-26. doi: 10.1007/s10198-014-0664-7.

17. Chen J, Wong CKH, McGhee SM, Pang PKP, Yu W. A Comparison between the EQ-5D and the SF-6D in Patients with Chronic Obstructive Pulmonary Disease (COPD). *PLoS One.* 2014; 9(11): e112389. doi:10.1371/journal.pone.0112389

18. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health Utilities Index. *Pharmaco Economics.* 1995;7:503-520.

19. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract*. 2011;17:268–274. doi: 10.1111/j.1365-2753.2010.01434.x

20. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005;8:94–104. doi: 10.1111/j.1524-4733.2005.04054.x

21. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of quality of life measures: Literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12): 1417-32.

22. Krejcie RV, Morgan DW. Determining sample size for research activities. *Educ Psychol Meas.*1970;30:607–610

23. Byrne BM. *Structural Equation Modeling with AMOS: Basic concepts, applications and programming*. 3rd ed. Abingdon:UK: Routledge; 2016.

24. Hu L, Bentler P. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999;6:1-55. doi: 10.1080/10705519909540118

25. Lam CLK, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable and valid in a Chinese population. *Value Health.* 2008; 11:295-303.

26. Brazier JE, Fukuhara S, Roberts J, Kharoubi S et al. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidem*. 62(12): 1323-1331.
27. Ferreira LN, Ferreira PL, Brazier J, Rowen D. A Portugese value set for the SF-6D. *Value Health.* 2010; 13(5): 624-630.
28. Norman R, Church J, Van den Berg B, Goodall S. Australian health-related quality of life population norms derived from the SF-6D. *Aust N Z J Public Health.* 2013;37(1):17-23.
29. Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of life in rural and urban populations in Lebanon using SF-36 health survey. *Health Qual Life Outcomes*. 2003 Dec;1(1):30.
30. Brazier, JE, Roberts, JR,. The estimation of a preference-based index from the SF-12. *Med Care.* 2004;42(9):851-859.
31. Chrzan K. Using partial profile choice experiments to handle large numbers of attributes. *Int J Mark Res*. 2010;52(6):827–40.
32. Craig BM, Pickard AS, Stolk E, Brazier JE. US valuation of the SF-6D. *Med Decis Making* 2013;33(6):793-803.
33. Payakachat N, Ali MM, Tilford JM. Can The EQ-5D Detect Meaningful Change? A Systematic Review. *Pharmacoeconomics*. 2015;33(11):1137–1154. doi:10.1007/s40273-015-0295-6