ORIGINAL RESEARCH

Evaluation of the Paper and Smartphone Versions of the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR16) and the Patient Health Questionnaire-9 (PHQ-9) in Depressed Patients in China

Long Zhen [1,2]
Gang Wang [1]
Gailing Xu [2]
Le Xiao [1]
Lei Feng [1]
Xu Chen [1]
Man Liu [1]
Xuequan Zhu [1]

[1]The National Clinical Research Center for Mental Disorders and Beijing Key Laboratory of Mental Disorders and Beijing Anding Hospital, Capital Medical University, Beijing 100088, People's Republic of China; [2]Tianjin Mental Health Center, Tianjin Anding Hospital, Tianjin 300222, People's Republic of China

Purpose: Smartphone-based questionnaires have advantages compared with their paper versions, but there is a lack of consistent research on depressive disorder questionnaires. This study aimed to assess the equivalence between the paper and smartphone versions of the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR16) and Patient Health Questionnaire-9 (PHQ-9) for patients with depressive disorders in psychiatric hospitals in China.

Patients and Methods: This was a randomized crossover study of 110 depressed patients recruited from the outpatient department of Beijing Anding Hospital from March 2016 to September 2018. Group 1 completed both the QIDS-SR16 and PHQ-9 in paper format and then completed the smartphone version 1–2 h later. Group 2 completed the scales in the reverse order. Reliability was evaluated using intraclass correlation coefficients (ICCs) with 95% confidence intervals (CI). The expected ICC was 0.9 (α=0.05).

Results: The overall ICC score of the QIDS-SR16 paper and smartphone versions was 0.904 (95% CI: 0.861–0.934), and the ICCs of each item ranged from 0.769 to 0.923. The overall ICC score of the PHQ-9 paper and smartphone versions was 0.951 (95% CI: 0.929–0.967), and the ICCs of each item ranged from 0.779 to 0.914.

Conclusion: This study demonstrated the equivalence of the paper and smartphone versions of the PHQ-9 and QIDS-SR16 in depressed patients in China.

Keywords: depressive disorder, equivalence, intraclass correlation coefficient, questionnaires, smartphone

Introduction

Depression is a recurring mental disorder that represents an important burden to the individuals, the society, the health care systems, and the economy.[1] The lifetime prevalence of depression in adults is 20% worldwide.[1] The incidence of depression is increasing year by year, and the onset of the disorder is occurring at an increasingly younger age.[2] At present, depression is one of the largest medical burdens in Chinese society.[3]

Many tools are available for the evaluation of depression. In 2011, the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR16) was introduced

Correspondence: Gang Wang
The National Clinical Research Center for Mental Disorders & Beijing Key Laboratory of Mental Disorders & Beijing Anding Hospital, Capital Medical University, Beijing 100088, People's Republic of China
Tel +86-13672146258
Fax +86-02288188856
Email andingjs@163.com

by Liu et al.[4,5] The QIDS-SR16 has been shown to have good reliability and validity for the screening of depressive disorders and the measurement of depressive symptoms, and it is widely used in China.[6] The Patient Health Questionnaire-9 (PHQ-9) was introduced in China by Bian et al.[7] in 2009. The PHQ-9 has high sensitivity and specificity for depressive disorders[8] and has been found to have good reliability and validity in Chinese patients.[9]

Those questionnaires are usually filled by the patients in paper versions. In fact, the paper versions are the only validated versions of these questionnaires. Paper versions require giving the questionnaire to the patient and taking it back. It requires to have the patient in the office or to send the questionnaire by mail. Therefore, filling the questionnaire at any time and anywhere is not possible. Smartphone apps based on psychiatric measurement-based care (MBC) may alleviate these difficulties.[10] Moreover, currently, mobile-based treatment is predominantly based on web-based interventions, but with the development of technology, smartphones now play an increasingly important role in mobile-based therapies and MBC. Nevertheless, little is known about the use of smartphone versions of the scales of depression in clinical work. Care is needed when migrating questionnaires to electronic formats to ensure that measurement equivalence with the original is demonstrated and that the measurement characteristics of the scale remain unchanged.[11] The available research on measurement equivalence mainly aimed to examine the factors influencing the validity of the results when migrating from paper to electronic formats, the operability of electronic versions, and questionnaires regarding somatic diseases and health conditions,[12–15] but direct comparisons between electronic and paper versions of psychometric questionnaires are rare. These studies showed that the demographic characteristics of the respondents and the management model of the app would affect the measurement results of the electronic versions. Some patients may encounter difficulties when operating a smartphone, and the different ways in which the entries and options are presented to the patient may cause problems. Individuals with smartphone anxiety may report more negative emotions when completing an emotional assessment on a smartphone. On the other hand, some studies reported that patients are more comfortable with the use of electronic versions and find them quicker to complete than paper versions.[16] In addition, some patients are more likely to express their mood and be more relaxed when completing the electronic versions.

Studies showed that paper and electronic scale measurements are equivalent,[17,18] but only using handheld computers, and there is a lack of consistent research on depressive disorders. Thus, the aim of the present study was to assess the equivalence between the paper and the smartphone versions of QIDS-SR16 and PHQ-9 for patients with depressive disorders in psychiatric hospitals. This could allow for more efficient monitoring of changes in depressive symptoms in MBC.

Patients and Methods
Study Design and Patients
This prospective study was conducted from March 2016 to September 2018 by convenience sampling at the Outpatient Department of Beijing Anding Hospital. It was the part of a project testing the effect of MBC in the management of major depression. Beijing Anding Hospital hospital receives 1500 outpatient visits daily and serves approximately 21 million people. The study protocol was approved by the Clinical Research Ethics Committee of Beijing Anding Hospital. All patients provided written informed consent to participate in this study and agreed to the publication of the data. This study was conducted in accordance with the Declaration of Helsinki.

All patients who were diagnosed with major depressive disorders were assessed using the Chinese version of the MINI version 5.0 modules on major depression[19] by a study investigator. The inclusion criteria were 1) met the criteria for major depressive disorder according to DSM-IV; 2) 21–65 years of age; 3) able to understand and fill in the questionnaires; and 4) no major physical diseases. The exclusion criteria were 1) previously diagnosed mania (light mania), bipolar disorder, schizophrenia, schizoaffective disorder, or other mental disorders; 2) depressive episodes with psychotic symptoms; 3) alcohol addiction or history of acute poisoning; or 4) questionnaires were not completed.

Study Procedure
In order to inform the research, a preliminary investigation was conducted before the formal investigation. All the scale raters were experienced psychiatrists, who received consistency training on the diagnosis of diseases, the determination of symptoms, the assessment of scales, and the use of the smartphone app. In this study, the paper and smartphone scales were completed in a randomized

crossover design, with an interval of 1–2 h between tests, separated by other neuropsychological tests.

The app (URL for IOS: https://www.pgyer.com/lb2R; for Android: https://www.pgyer.com/iyFC; Department of Software Engineering, Beijing University of Technology, China; Beijing Anding Hospital has the copyright and ownership) was downloaded and installed onto the participants' phones with the assistance of the researchers. The participants first completed a demographic characteristics form and were randomly divided by a central system into two groups. Group 1 completed paper versions of the depression scale QIDS-SR16 and PHQ-9 and then completed the smartphone versions 1 to 2 h later. Group 2 completed the same scales in the reverse order. Brief training on operating the app was given by the researchers. The participants completed scale measurements independently but were allowed to ask for assistance.

App Design

The app was designed to be as simple and easy to use as possible in order to reach a wide range of people with depression. All the contents in the smartphone versions are the same as those in the paper versions. Basic instructions were provided on the app homepage, and all patients were given time to read these prior to completing the scales. Only one item was shown on each page. If the item was not completely visible on one page, the participants could scroll down to see it all. Answers could be modified while answering each question, but could not be modified when completed. The program could not continue without a response being submitted to each item. Once submitted, data were analyzed automatically, and the results were displayed on the medical side of the app.

Measurements

The demographic data questionnaire included questions on the general background (eg, sex, age, education level, onset age, duration of illness, and family history of psychiatric disorders) of the patients, which were completed by the patients themselves.

The QIDS-SR16 is a 16-item, self-administered scale that measures the severity of depressive symptoms relating to nine symptom domains over the last 7 days.[4,5] The overall score of the scale ranges between 0 and 27, and a higher score indicates more severe depressive symptoms. It takes 5–7 minutes to complete. The scores may be graded as mild (score of 6–10), moderate (score of 11–15), severe (score of 16–20), and very severe (score of 21–27).[5]

The PHQ-9 is a multipurpose instrument that assists in the screening and measurement of depressive symptoms within the past 2 weeks.[20] It is a brief self-report tool, comprising nine items. Respondents rate the frequency of the symptoms on a 4-point rating scale: 0 (not at all), 1 (several days), 2 (more than half the days), and 3 (nearly every day). The overall scores range between 0 and 27. The scores may also be graded as mild (score of 5–9), moderate (score of 10–14), moderate-severe (score of 15–19), and severe (score 20–27).[8] The PHQ-9 only takes 3–5 minutes to complete and is rapidly scored.

In 2009, Si et al.[19] compared the reliability and validity of the Mini International Neuropsychiatric Interview (MINI) with that of the Composite International Diagnostic Interview (CIDI) and the Structured Clinical Interview for Disorders (SCID). The results showed that the MINI, which is a diagnostic tool for all mental disorders and not only depression, had very acceptable reliability and validity scores. The MINI is a brief structured clinical interview based on the DSM-IV criteria.[19] It has similar reliability and validity properties compared with the CIDI and the SCID but can be administered in less time. It includes 130 questions.

Sample Size

The sample size was determined according to the ISPOR guidelines.[11] The study power was set at 95%, and the expected intraclass correlation coefficient (ICC) was 0.9 (α=0.05), resulting in a target sample size of 110 patients.

Statistics

SPSS 25.0 (IBM Corp., Armonk, NY, USA) was used for data analysis. Continuous variables are presented as means ± standard deviations (SD) or as medians and interquartile ranges (IQR) according to their distribution, as determined by the Kolmogorov–Smirnov test; comparisons between groups were performed with the Student's t-test or Mann–Whitney U-test, as appropriate. Categorical variables are reported as frequencies with percentages and were compared with the chi-square test. Concordance of scale scores between the paper and smartphone versions was analyzed using a two-way fixed-effects ICC model, including intra-groups and inter-groups. High positive ICCs indicate that different versions' measurements covary and that the mean and variability of the scores are similar. Two-sided P<0.05 was considered statistically significant.

Results

Sociodemographic Data

Out of the 112 patients who fulfilled the inclusion criteria, 110 gave informed consent and provided complete data. In Group 1, 54 patients completed the paper versions first. In Group 2, 56 patients completed the smartphone version first. The mean age was 30.6±7.3 years, and 40.9% of the patients were male (n=45). The majority of patients (74.6%) had a high level of education: university or higher (n=82) (Table 1). All patients had used smartphones previously. The majority of patients (70.9%, n=78) expressed a preference for the mobile app version. There were no significant differences in any of the patient characteristics between the two groups (P>0.05, Table 1).

Intra-Group Consistency Measurement

High concordances were noted between the paper and smartphone versions for the overall and individual item scores of the QIDS-SR16 scale (Table 2) and PHQ-9 scale (Table 3) in the two groups.

Inter-Version Consistency Measurement

High concordances were noted between the paper and smartphone versions of the QIDS-SR16 scale for the overall score (ICC=0.904, 95% CI 0.861–0.934) and individual item scores (ICC range: 0.769–0.923) (Table 4).

High concordances were noted between the paper and smartphone versions of the PHQ-9 scale for the overall scores (ICC=0.951, 95% CI 0.929–0.967) and individual item scores (range 0.779–0.914) (Table 5).

Discussion

Smartphone-based questionnaires have advantages compared with their paper versions, but there is a lack of consistent research on depressive disorder questionnaires. Therefore, the present study aimed to assess the equivalence between the paper and the smartphone versions of the QIDS-SR16 and PHQ-9 for patients with depressive disorders in psychiatric hospitals in China. This study demonstrated the equivalence in the measurement properties of the paper and smartphone versions of the PHQ-9 and QIDS-SR16 in depressed patients in China.

Table 1 Sociodemographic Characteristics of the Study Population

Characteristics		Total (n=110)	Group 1 (n=54)	Group 2 (n=56)	Z/X^2	P
Age (years), median (IQR)		24.00–35.25	24–36.25	24–34.75	−0.551	0.582
Age at onset (years), median (IQR)		21.00–25.00	20–27	21.25–25	−1.198	0.231
Duration of illness (years), median (IQR)		1.00–8.25	1–8	1–9	−0.015	0.988
Hamilton Depression Rating Scale score, median (IQR)		18.75–25	18–25.25	20–25	−0.775	0.439
Young Mania Rating Scale score, median (IQR)		0–2	0–2	0–2	−0.776	0.438
Sex, n (%)	Male	45 (40.9)	24 (53.3)	21 (46.7)	0.548	0.561
	Female	65 (59.1)	30 (46.2)	35 (53.8)		
Education, n (%)	High school or less	13 (11.8)	8 (61.5)	5 (38.5)	5.256	0.156
	Technical secondary to junior college	15 (13.6)	9 (60)	6 (40)		
	Undergraduate degree	55 (50.0)	21 (38.2)	34 (61.8)		
	Master's degree or higher	27 (24.5)	16 (59.3)	11 (40.7)		
Version preference, n (%)	Paper	32 (29.1)	12 (37.5)	20 (62.5)	2.426	0.144
	Electronic	78 (70.9)	42 (53.8)	36 (46.2)		
Family history of psychiatric disorders, n (%)	Positive	25 (22.7)	10 (40)	15 (60)	1.070	0.366
	Negative	85 (77.3)	44 (51.8)	41 (48.2)		
Employment	Full-time	62 (56.4)	32 (51.6)	30 (48.4)	0.432	0.812
	Student	23 (20.9)	11 (47.8)	12 (52.2)		
	No job or retirement	25 (22.7)	11 (44)	14 (56.0)		

Notes: Group 1: paper version, followed by the electronic version 1–2 h later. Group 2: electronic version, followed by the paper version 1–2 h later.
Abbreviation: IQR, interquartile range.

Table 2 Intra-Group ICC (95% CI) of the QIDS-SR16 Scale

QIDS-SR16 Item	Group 1 (n=54)	Group 2 (n=56)
1. Sleep	0.860 (0.760–0.919)	0.857 (0.756–0.916)
2. Sad mood	0.742 (0.555–0.850)	0.794 (0.648–0.879)
3. Appetite/Weight	0.922 (0.866–0.955)	0.927 (0.876–0.957)
4. Concentration/Decision-making	0.868 (0.773–0.924)	0.854 (0.750–0.914)
5. Self-outlook	0.830 (0.707–0.901)	0.839 (0.725–0.906)
6. Thoughts of death or suicide	0.826 (0.699–0.899)	0.873 (0.783–0.925)
7. Involvement	0.883 (0.799–0.932)	0.908 (0.844–0.946)
8. Energy level	0.848 (0.739–0.912)	0.885 (0.804–0.933)
9. Agitation/Retardation	0.851 (0.744–0.914)	0.800 (0.659–0.883)
Overall score	0.889 (0.809–0.936)	0.923 (0.869–0.955)

Notes: Group 1: paper version, followed by the electronic version 1–2 h later. Group 2: electronic version, followed by the paper version 1–2 h later.
Abbreviations: QIDS-SR16, Quick Inventory of Depressive Symptomatology-Self-Report; ICC, intraclass correlation coefficient; CI, confidence interval.

Table 3 Intra-Group ICC (95% CI) of the PHQ-9 Scale

PHQ-9 Item	Group 1 (n=54)	Group 2 (n=56)
1. Little interest or loss of pleasure	0.854 (0.748–0.915)	0.854 (0.751–0.914)
2. Feeling down, depressed, or hopeless	0.878 (0.790–0.929)	0.805 (0.667–0.886)
3. Trouble falling or staying asleep or sleeping too much	0.861 (0.760–0.919)	0.912 (0.850–0.948)
4. Feeling tired or having little energy	0.875 (0.784–0.927)	0.908 (0.843–0.946)
5. Poor appetite or over-eating	0.915 (0.854–0.951)	0.915 (0.856–0.950)
6. Poor self-esteem	0.898 (0.824–0.941)	0.833 (0.715–0.902)
7. Trouble concentrating	0.732 (0.538–0.845)	0.837 (0.721–0.904)
8. Moving or speaking so slowly that other people could notice	0.918 (0.858–0.953)	0.891 (0.814–0.936)
9. Thoughts that you would be better off dead or hurting yourself	0.896 (0.821–0.940)	0.929 (0.879–0.958)
Overall score	0.950 (0.913–0.988)	0.953 (0.920–0.972)

Notes: Group 1: paper version, followed by the electronic version 1–2 h later. Group 2: electronic version, followed by the paper version 1–2 h later.
Abbreviations: PHQ-9, Patient Health Questionnaire-9; ICC, intraclass correlation coefficient; CI, confidence interval.

Table 4 Consistency Tests Between the Paper and Smartphone Versions of the QIDS-SR16

Item	Paper Mean ± SD	Electronic Mean ± SD	ICC (95% CI)
1. Sleep	2.55 ± 0.55	2.55 ± 0.56	0.858 (0.793–0.903)
2. Sad mood	2.32 ± 0.63	1.90 ± 0.50	0.769 (0.663–0.842)
3. Appetite/Weight	1.47 ± 0.23	1.44 ± 0.35	0.923 (0.887–0.947)
4. Concentration/Decision-making	1.78 ± 0.68	1.63 ± 0.65	0.860 (0.796–0.904)
5. Self-outlook	1.87 ± 0.64	1.90 ± 0.61	0.832 (0.755–0.885)
6. Thoughts of death or suicide	0.86 ± 0.24	0.87 ± 0.23	0.854 (0.788–0.900)
7. Involvement	2.03 ± 0.56	1.95 ± 0.54	0.898 (0.851–0.930)
8. Energy level	1.54 ± 0.42	1.48 ± 0.43	0.868 (0.807–0.909)
9. Agitation/Retardation	1.71 ± 0.47	1.65 ± 0.44	0.835 (0.760–0.887)
Overall score	16.51 ± 2.63	16.12 ± 2.70	0.904 (0.861–0.934)

Abbreviations: QIDS-SR16, Quick Inventory of Depressive Symptomatology-Self-Report; SD, standard deviation; ICC, intraclass correlation coefficient; CI, confidence interval.

Selection of the Questionnaires

In the present study, the QIDS-SR16 and PHQ-9 questionnaires were selected because they are often used in the literature and because they are short, easy to administer, and self-rating, which allows their application as smartphone versions. In the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, both the QIDS-SR16 and PHQ-9 were used.[21] In an MBC

Table 5 Consistency Tests Between the Paper and Smartphone Versions of the PHQ-9

Item	Paper Mean ± SD	Electronic Mean ± SD	ICC (95% CI)
1. Little interest or loss of pleasure	2.36 ± 0.72	2.35 ± 0.70	0.854 (0.787–0.900)
2. Feeling down, depressed, or hopeless	2.33 ± 0.79	2.30 ± 0.70	0.840 (0.766–0.890)
3. Trouble falling or staying asleep or sleeping too much	2.11 ± 1.01	2.08 ± 0.94	0.886 (0.834–0.922)
4. Feeling tired or having little energy	2.57 ± 0.65	2.54 ± 0.69	0.890 (0.839–0.924)
5. Poor appetite or over-eating	1.66 ± 1.08	1.68 ± 0.99	0.914 (0.875–0.941)
6. Poor self-esteem	2.17 ± 0.91	2.17 ± 0.93	0.864 (0.802–0.907)
7. Trouble concentrating	1.89 ± 0.96	1.83 ± 0.96	0.779 (0.677–0.848)
8. Moving or speaking so slowly that other people could notice	1.33 ± 0.92	2.30 ± 0.70	0.906 (0.863–0.936)
9. Thoughts that you would be better off dead or hurting yourself	0.84 ± 0.74	1.21 ± 0.81	0.914 (0.875–0.941)
Overall score	16.96 ± 4.79	16.90 ± 4.47	0.951 (0.929–0.967)

Abbreviations: QIDS-SR16, Quick Inventory of Depressive Symptomatology-Self-Report; ICCs, intraclass correlation coefficients; CI, confidence intervals; PHQ-9, Patient Health Questionnaire-9; MINI, Mini International Neuropsychiatric Interview; CIDI, Composite International Diagnostic Interview; SCID, Structured Clinical Interview for Disorders; MBC, measurement-based care; SD, standard deviations; IQR, interquartile ranges; PHQ-9, Patient Health Questionnaire-9; SD, standard deviation; ICC, intraclass correlation coefficient; CI, confidence interval.

project on major depression in China, the psychometric properties of the QIDS-SR and PHQ-9 in depressed inpatients were examined, and it was found that they have similar and acceptable psychometric properties in most domains.[22] Moreover, the QIDS-SR16 and PHQ-9 have different characteristics. PHQ-9 consists of only one page, and the questions are concise and easier to answer. The QIDS-SR16 consists of three pages, and only people with high education levels find it as easy to answer as the PHQ-9.[23] Therefore, those two questionnaires were selected for the present validation study.

Study Population

In comparison to epidemiological data from a previous study on psychiatric disorders in China,[24] the average age of the patients in this study was relatively young, the average duration of illness was relatively short, and the average education level was relatively high. The reasons for this may be that older patients are inexperienced in the use of smartphones or do not have smartphones, and thus were unwilling to take part in the study. In addition, the participants in this study were outpatients from a tertiary hospital at the Center of Science Technology and Culture of China in Beijing. These patients tend to have a short illness duration and high levels of education. Ali et al.[16] reported that patients prefer iPad versions of scales of depression over paper versions. Similarly, in this study, the majority of patients (73.8%) also preferred the electronic version.

Consistency Between Electronic and Paper Versions

Previous studies generally showed good concordance between the electronic and paper versions of various questionnaires across a wide variety of diseases and conditions,[25–27] but those studies were not controlled clinical trials, they did not use smartphone electronic versions, and did not include scales used in depression. Nevertheless, the results of the present study for QIDS-SR16 and PHQ-9 are consistent between the paper and electronic versions, as supported by the previous studies mentioned above. For outpatients suffering from depression, this study showed that paper and electronic versions had a high concordance on the whole.

It is generally believed that ICC values lower than 0.70 indicate poor reliability, values higher than 0.75 indicate good reliability for group comparisons, while ICC should be between 0.85 to 0.95 for applications at the individual level.[11] In this study, all the items in the intra-group ICC analysis showed good reliability in both groups, further supporting the reliability of the ICC between two versions of questionnaires. Between the two versions, the ICCs for the overall scores were all higher than 0.90, ie, 0.904 for the QIDS-SR16 and 0.951 for the PHQ-9. For the individual items, the ICC for three items on the QIDS-SR16 and two items on the PHQ-9 were less than 0.85. The items showing a low level of concordance included "Sad mood", "Self-outlook", and "Agitation/Retardation" in the QIDS-SR16, and "Feeling down, depressed, or hopeless" and "Trouble concentrating" in the PHQ-9. These low ICCs

may have arisen because there is no clear criterion for judging mood or emotional state. The ICC of the Appetite/weight item was the highest, at 0.923. This might be because both appetite and weight are mainly manifested as behavioral problems and easy to judge. Items that are difficult to judge, ie, those without clear criteria, may be more affected by a switch to a different version.

The PHQ-9 showed good consistency between the electronic and paper versions. This is supported by a previous study that showed that computerization of the PHQ-9 did not affect its psychometric properties.[28] Compared with PHQ-9, the descriptions in the responses in the QIDS-SR16 are complicated. Sung et al.[23] determined that the QIDS-SR16 score was able to distinguish minor from major depression, while the PHQ-9 could not. This may be because the QIDS-SR16 has more items in which symptoms must be assessed more precisely. This feature, however, may not necessarily be better for the electronization of the scale, and this may lead to difficulties for patients in distinguishing between the different possible answers to an item. Based on these features, the PHQ-9 may be a good choice for app quantitative treatment, as it is simple but effective. Currently, the PHQ-9 is being further simplified, and we expect the results of relevant studies to be tested in the MBC field.[29]

In this study, the expected ICC was 0.9, but the items' ICCs were mostly lower than 0.90. This is possible because the participants in this study had major depressive disorders, with decreased energy and attention, which may be required to complete the scale. Nevertheless, the ICCs of the overall score of both measurements were higher than 0.90, and the ICCs of items were mostly higher than 0.85, ie, the consistencies were within the acceptable range.

Limitations

The main limitations of this study pertain to sample heterogeneity and the lack of qualitative data. Only outpatients from one hospital were recruited, and therefore the findings may not be applicable to all areas of China. In addition, in the process of moving from a paper version to an electronic version, differences in patient groups, electronic equipment, or operation procedures may affect the measurement results.[12,30] Therefore, the population characteristics of the patients are important. In this study, there was no simultaneous qualitative interview to collect information from patients to better understand the impact of different populations on measurement results. The factors influencing population characteristics should be considered in future research. Furthermore, in this study, the application environment was controlled and stable, so the results may have poor predictability for real clinical circumstances. In the present study, the interval between the two questionnaires was 1–2 h. The determination of the interval time was based on previous research experience and the specific operating environment of the subject, to reduce the risk of carryover effect and risk of change in mindset as much as possible. There is currently no scientifically proven standard time for this time issue. This is a limitation, and therefore, we tried to minimize this bias by answering the questionnaires in the reverse order in these two groups in our study. Finally, clinimetrics is a new field focusing on the science of clinical measurements and is considered an emerging topic in the field of measurement-based care,[31–33] and this was not evaluated in the present study. We will explore this further in future studies.

Conclusion

In conclusion, this study demonstrated equivalence between the paper and smartphone versions of two scales of depression in depressed patients in China. This demonstrates that both the QIDS-SR16 and PHQ-9 scales are appropriate for use in both paper and electronic versions. These findings support the use of the electronic versions of the PHQ-9 and QIDS-SR16 via smartphone apps. This is particularly the case for the PHQ-9 because it is simple and easier for patients to complete, with good psychometric equivalence. Both the PHQ-9 and QIDS-SR16 can be flexibly selected and applied in clinical practice and/or scientific research, as required.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Otte C, Gold SM, Penninx BW, et al. Major depressive disorder. *Nat Rev Dis Primers*. 2016;2(1):16065. doi:10.1038/nrdp.2016.65

2. Ferrari AJ, Charlson FJ, Norman RE, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med*. 2013;10(11):e1001547. doi:10.1371/journal.pmed.1001547

3. Qin X, Wang S, Hsieh CR. The prevalence of depression and depressive symptoms among adults in China: estimation based on a National Household Survey. *China Economic Rev*. 2018;51:271–282. doi:10.1016/j.chieco.2016.04.001

4. Liu J, Xiang YT, Wang G, et al. Psychometric properties of the Chinese versions of the quick inventory of depressive symptomatology - clinician rating (C-QIDS-C) and self-report (C-QIDS-SR). *J Affect Disord*. 2013;147(1–3):421–424. doi:10.1016/j.jad.2012.08.035

5. Liu J, Xiang YT, Lei H, et al. Guidance on the conversion of the Chinese versions of the quick inventory of depressive symptomatology-self-report (C-QIDS-SR) and the montgomery-asberg scale (C-MADRS) in Chinese patients with major depression. *J Affect Disord*. 2014;152–154:530–533. doi:10.1016/j.jad.2013.09.023

6. Zhao N, Wang XH, Shi JJ, et al. Using the quick inventory of depressive symptomatology to assess gender differences in residual symptoms of depressed patients after acute phase treatment. *Chin Ment Health J*. 2018;32:903–909.

7. Bian CD, He XY, Qian J, Wu WY, Li CB. The liability and validity of a modified patient health questionnaire for screening depressive syndrome in general hospital outpatients. *J Tongji Univ (Med Sci)*. 2009;30:136–140.

8. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x

9. Zeng QZ, Liu H, Miao JM, et al. Screening value of the patient health questionnaire depression scale in outpatients from traditional Chinese internal department. *J Clin Psychiatry*. 2013;23:229–232.

10. Maresova P, Klimova B, Kuca K. Mobile applications as good intervention tools for individuals with depression. *Ceska Slov Farm*. 2017;66(2):55–61.

11. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value Health*. 2009;12(4):419–429. doi:10.1111/j.1524-4733.2008.00470.x

12. Muehlhausen W, Byrom B, Skerritt B, McCarthy M, McDowell B, Sohn J. Standards for instrument migration when implementing paper patient-reported outcome instruments electronically: recommendations from a qualitative synthesis of cognitive interview and usability studies. *Value Health*. 2018;21(1):41–48. doi:10.1016/j.jval.2017.07.002

13. Griffiths-Jones W, Norton MR, Fern ED, Williams DH. The equivalence of remote electronic and Paper Patient Reported Outcome (PRO) collection. *J Arthroplasty*. 2014;29(11):2136–2139. doi:10.1016/j.arth.2014.07.003

14. Rasmussen SL, Rejnmark L, Ebbehoj E, et al. High Level of agreement between electronic and paper mode of administration of a thyroid-specific patient-reported outcome, ThyPRO. *Eur Thyroid J*. 2016;5(1):65–72. doi:10.1159/000443609

15. Delgado-Herrera L, Banderas B, Ojo O, Kothari R, Zeiher B. Diarrhea-predominant irritable bowel syndrome: creation of an electronic version of a patient-reported outcome instrument by conversion from a pen-and-paper version and evaluation of their equivalence. *Patient Relat Outcome Meas*. 2017;8:83–95. doi:10.2147/PROM.S126605

16. Ali FM, Johns N, Finlay AY, Salek MS, Piguet V. Comparison of the paper-based and electronic versions of the dermatology life quality index: evidence of equivalence. *Br J Dermatol*. 2017;177(5):1306–1315. doi:10.1111/bjd.15314

17. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010. *Lancet (London, England)*. 2012;380(9859):2163–2196. doi:10.1016/S0140-6736(12)61729-2

18. Goldstein LA, Connolly Gibbons MB, Thompson SM, et al. Outcome assessment via handheld computer in community mental health: consumer satisfaction and reliability. *J Behav Health Serv Res*. 2011;38(3):414–423. doi:10.1007/s11414-010-9229-4

19. Si TU, Shu L, Dang WM, Su YA, Chen VX, Dong WT. Evaluation of the reliability and validity of Chinese version of the mini-international neuropsychiatric interview in patients with mental disorders. *Chin Ment Health J*. 2009;23:493–497.

20. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD the PHQ primary care study. *JAMA*. 1999;282(18):1737–1744. doi:10.1001/jama.282.18.1737

21. Trivedi MH. Tools and strategies for ongoing assessment of depression: a measurement-based approach to remission. *J Clin Psychiatry*. 2009;70(Suppl 6):26–31. doi:10.4088/JCP.8133su1c.04

22. Feng Y, Huang W, Tian TF, et al. The psychometric properties of the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR) and the Patient Health Questionnaire-9 (PHQ-9) in depressed inpatients in China. *Psychiatry Res*. 2016;243:92–96. doi:10.1016/j.psychres.2016.06.021

23. Sung SC, Low CC, Fung DS, Chan YH. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: A comparison of the nine-item Patient Health Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). *Asia Pac Psychiatry*. 2013;5(4):249–258. doi:10.1111/appy.12101

24. Yin H, Xu G, Tian H, Yang G, Wardenaar KJ, Schoevers RA. The prevalence, age-of-onset and the correlates of DSM-IV psychiatric disorders in the Tianjin Mental Health Survey (TJMHS). *Psychol Med*. 2018;48(3):473–487. doi:10.1017/S0033291717001878

25. White MK, Maher SM, Rizio AA, Bjorner JB. A meta-analytic review of measurement equivalence study findings of the SF-36(R) and SF-12(R) health surveys across electronic modes compared to paper administration. *Qual Life Res*. 2018;27(7):1757–1767. doi:10.1007/s11136-018-1851-2

26. Muehlhausen W, Doll H, Quadri N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health Qual Life Outcomes*. 2015;13(1):167. doi:10.1186/s12955-015-0362-x

27. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. *Qual Life Res*. 2015;24(8):1949–1961. doi:10.1007/s11136-015-0937-3

28. Erbe D, Eichert HC, Rietz C, Ebert D. Interformat reliability of the patient health questionnaire: validation of the computerized version of the PHQ-9. *Internet Interv*. 2016;5:1–4. doi:10.1016/j.invent.2016.06.006

29. Shin C, Lee SH, Han KM, Yoon HK, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. *Psychiatry Investig*. 2019;16(4):300–305. doi:10.30773/pi.2019.02.01

30. MacKenzie H, Thavaneswaran A, Chandran V, Gladman DD. Patient-reported outcome in psoriatic arthritis: a comparison of web-based versus paper-completed questionnaires. *J Rheumatol*. 2011;38(12):2619–2624. doi:10.3899/jrheum.110165

31. Carrozzino D. Clinimetric approach to rating scales for the assessment of apathy in parkinson's disease: A systematic review. *Prog Neuropsychopharmacol Biol Psychiatry*. 2019;94:109641. doi:10.1016/j.pnpbp.2019.109641

32. Fava GA, Carrozzino D, Lindberg L, Tomba E. The clinimetric approach to psychological assessment: a tribute to Per Bech, MD (1942–2018). *Psychother Psychosom*. 2018;87(6):321–326. doi:10.1159/000493746

33. Fleck MP, Carrozzino D, Fava GA. The challenge of measurement in psychiatry: the lifetime accomplishments of Per Bech (1942–2018). *Rev Bras Psiquiatria*. 2019;41(5):369–372. doi:10.1590/1516-4446-2019-0509