Journal of
Clinical Medicine

MDPI

Article

Equivalence of Paper and Smartphone Versions of the Beck Depression Inventory-II

Hiroyuki Uchida [1,2,†], Takumi Igusa [1,3,†], Yurika Higashi [4], Minami Takeda [4], Kenji Tsuchiya [5], Senichiro Kikuchi [1,4] and Kazuki Hirao [1,4,*]

1 Graduate School of Health Sciences, Gunma University, Maebashi 371-8514, Japan
2 Department of Rehabilitation, Kurashiki Heisei Hospital, Kurashiki 710-0826, Japan
3 Department of Rehabilitation, Medical Corporation Taiseikai, Uchida Hospital, Numata 378-0005, Japan
4 Department of Occupational Therapy, Faculty of Medicine, Gunma University, Maebashi 371-8514, Japan
5 Department of Rehabilitation, Faculty of Health Sciences, Nagano University of Health and Medicine, Nagano 381-2227, Japan
* Correspondence: kazuki.hirao@gunma-u.ac.jp; Tel./Fax: +81-27-220-8952
† These two authors contributed equally to this work.

Abstract: Background: The Beck Depression Inventory-II (BDI-II) is a widely used patient-reported outcome (PRO) tool designed to screen for depressive symptoms and assess their severity. In recent years, with advancements in digital technology, the BDI-II has been adapted for use as an electronic PRO (ePRO) tool. However, to the best of our knowledge, the reliability of the smartphone version of the BDI-II has not been thoroughly investigated. This study aimed to assess the equivalence of the traditional paper and smartphone versions of the BDI-II. **Methods:** This study employed a randomized crossover design with adults (n = 100) from the Gunma Prefecture, Japan. Participants completed both the paper and smartphone versions of the BDI-II at 1-week intervals. The equivalence between the two versions was evaluated using the intraclass correlation coefficient ($ICC_{agreement}$). Additionally, Cronbach's alpha and McDonald's omega were calculated for both versions. **Results:** The mean age of the participants was 19.78 years (SD = 0.94, 10% male). The $ICC_{agreement}$ between the paper and smartphone versions of the BDI-II was 0.81 (95% CI 0.74–0.87). Cronbach's alpha was 0.88 (95% CI 0.84–0.91) for the paper version and 0.88 (95% CI 0.84–0.91) for the smartphone version. McDonald's omega was 0.88 (95% CI 0.81–0.95) for the paper version and 0.89 (95% CI 0.85–0.93) for the smartphone version. **Conclusions:** The BDI-II is suitable for use in its smartphone version, and the smartphone version of the BDI-II is a valuable addition to the mental healthcare professional's toolkit.

Keywords: patient-reported outcomes; electronic; depression; smartphone; equivalence

check for updates

Academic Editor: Gerd Wagner

Received: 22 November 2024
Revised: 5 January 2025
Accepted: 13 January 2025
Published: 14 January 2025

Citation: Uchida, H.; Igusa, T.; Higashi, Y.; Takeda, M.; Tsuchiya, K.; Kikuchi, S.; Hirao, K. Equivalence of Paper and Smartphone Versions of the Beck Depression Inventory-II. *J. Clin. Med.* **2025**, *14*, 500. https://doi.org/10.3390/jcm14020500

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Using patient-reported outcomes (PROs) is a key recommendation recognized globally. PROs refer to outcomes directly reported by patients about their feelings and functioning, including symptoms, physical capabilities, and quality of life, without modification or interpretation by healthcare professionals [1–3]. Incorporating PROs in clinical practice offers several advantages. Previous studies have demonstrated that systematically monitoring patient symptoms using PROs enhances patient–physician communication and reduces the likelihood of missed symptoms [3–5]. Notably, PRO-based care has been associated with improved outcomes compared to standard care [6]. Therefore, integrating PROs into routine clinical practice and research is essential. Furthermore, the recent development of digital

technology and the proliferation of Internet-based interventions have increased the use of electronic PROs (ePROs), streamlining their use and expanding their application [7–10].

ePROs offer several advantages over traditional paper-version PROs. ePROs are often preferred by patients, because they may disclose more sensitive information than paper-based versions [8,10–15]. In addition, ePROs reduce score calculation errors, data entry mistakes, and missing data, thereby improving data quality and facilitating reliable analysis and reporting [8,16–18]. As a result, these improvements can enhance the quality of patient care and increase clinical efficiency [9,19]. In some cases, can also reduce the cost associated with data collection [9].

However, transitioning from paper-based PROs to ePROs requires careful consideration of potential challenges. In fact, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines highlight that differences in administrative methods—such as changing from circling an answer on paper to selecting it on the screen—and presentation formats, such as scrolling or font size adjustments, necessitate thorough cognitive debriefing, usability tests, and reliability evaluations (e.g., intraclass correlation coefficients [ICCs]) to ensure equivalence or superiority [8]. Many studies have been conducted on the use of ePROs across various fields [9,10,20,21]. For instance, in mental health, ePROs are increasingly utilized for screening depressive symptoms and assessing their severity [21–26]. One notable example is the Beck Depression Inventory-II (BDI-II, which has been widely investigated for its application as an ePRO [27–30].

BDI-II is one of the most widely used PROs for detecting possible depression in healthy populations and assessing the severity of depressive symptoms in diagnosed patients [31–36]. Originally developed in 1961, the BDI was slightly modified in 1979 and substantially revised in 1996 to align with contemporary diagnostic criteria for major depression [37]. The BDI-II consists of 21 items and offers notable advantages, including that it can be administered to a individuals aged 13–80 years old and completed in just 5–10 min [31,38,39]. A distinctive feature of the BDI-II is that, unlike conventional PROs, it uses descriptive response options tailored for each item, rather than using a uniform scale. This approach provides clearer context for respondents, making questions easier to understand and interpret. For instance, in assessing suicidal ideation, the response options are as follows:

0 = I do not have any thoughts of killing myself;
1 = I have thoughts of killing myself, but I would not carry them out;
2 = I would like to kill myself;
3 = I would kill myself if I had the chance.

This format enhances clarity compared to ambiguous answer choices such as “rarely” or “often” [40], allowing BDI-II to capture more nuanced details of depressive symptoms [40,41]. Because of these advantages, the BDI-II has been extensively tested for reliability and validity across diverse populations and countries. Furthermore, equivalence studies have explored its use as an ePRO [27–33,35,36]. However, to our knowledge, no study has specifically evaluated the equivalence of the BDI-II between its paper-based and smartphone versions.

Smartphones offer several over personal computers and tablets. With a growing number of smartphone users worldwide, most people always carry a smartphone to access the Internet [16,42]. Thus, smartphones allow more real-time access to ePRO than personal computers or tablets. Previous studies have highlighted the benefits of using smartphones as ePRO devices [16,22,24,26]. Given the advantages of smartphones, as well as the unique features of the BDI-II, which captures detailed and nuanced depressive symptoms, a smartphone version of the BDI-II could assess depressive symptoms in real time [40,41]. In turn, this may reduce the number of depressive symptoms overlooked by physicians and

improve communication between patients and healthcare providers [3–5]. For this reason, it is important to develop a smartphone version of the BDI-II.

However, the ISPOR guidelines emphasize that differences in question phrasing and presentation formats between ePROs and original paper-based versions could adversely affect their reliability and validity [8]. Moreover, the type of device used may also influence these factors [8]. Therefore, a comparison of the ePRO and the original PRO should be conducted on a device-by-device basis.

This study aimed to evaluate the equivalence of the paper-based and smartphone-based versions of the BDI-II in accordance with the ISPOR guidelines. The hypothesis of this study was that the scores obtained from the smartphone version of the BDI-II would be equivalent to those from the paper-based version.

2. Material and Methods

2.1. Study Design and Ethics

This randomized crossover study aimed to evaluate the format-to-format equivalence of the paper-based and smartphone versions of the BDI-II (Figure 1). The study participants were recruited from the Gunma University in Gunma, Japan, from October 2023 to December 2023. The study was conducted according to the ISPOR guidelines and was approved by the Ethical Review Board for Medical Research Involving Human Subjects of Gunma University (approval no. HS2022-109) [8]. Written informed consent was obtained from each participant before study participation.

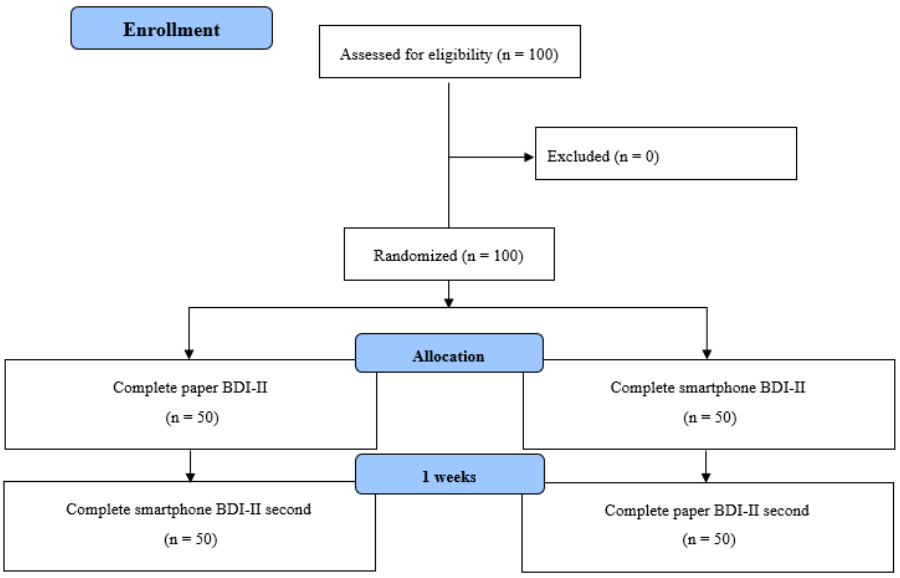

Figure 1. Flowchart of the study.

2.2. Participants and Procedure

Participants in this study were recruited from Gunma University in the Gunma Prefecture, Japan, using advertisements, emails, and social networking services. The eligibility criteria for this study were as follows: (1) 18 years of age or older, (2) male or female, (3) smartphone owner, (4) native Japanese speaker, and (5) obtained informed consent. Eligible participants were randomly assigned to either the smartphone-first (groups responding to the smartphone-based version BDI-II first) or paper-first (groups responding to the paper-based version BDI-II first) group after providing their demographic information (age and gender) and lifestyle habits (drinking, exercise, and smoking habits). One week after the first survey response, the smartphone-first group responded to the paper-based version of the BDI-II and the paper-first group responded to the smartphone version of the

BDI-II. This 1-week interval was chosen based on a prior study investigating the equivalence of paper-based and electronic PROs, aimed at reducing potential recall and carryover effects [26]. These questionnaires were administered under the researcher's supervision and conducted in a controlled environment, minimizing the risk of omitted responses and standardizing test conditions by reducing noise, distraction, and fatigue.

2.3. Randomization

Participants were randomly assigned to the paper- or smartphone-first groups in a 1:1 ratio before completing the survey. The randomized list was generated using the permuted block method (block size 4) and Excel software by a third-party unrelated to the study and sent to the central registry center of Kurashiki Heisei Hospital in Okayama, Japan. Once the participants were enrolled in the study, the researchers informed the central registry center at Kurashiki Heisei Hospital, which performed the random assignment.

2.4. Sample Size

The sample size was determined on the basis of the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative and the ISPOR guidelines. The ISPOR guidelines used Walter's methodology to calculate the sample size required for ICCs. With an ICC of 0.85 in the underlying population and two assessments conducted at 80% power, 43 complete data points would be needed to conform that the ICC of the true population exceeds 0.70 with 95% confidence [8,43]. However, the COSMIN initiative recommends a sample size of at least 100 to achieve adequate statistical power for assessing testing–retesting reliability (i.e., equivalence) [44]. Thus, the target sample size for this study was set at 100.

2.5. Measure: Beck Depression Inventory-II

BDI-II is a 21-item self-administered questionnaire designed to measure the severity of depressive symptoms experienced over the previous two weeks [31,38]. Each item is rated using a 4-point Likert scale ranging from 0 to 3 points. The total scores ranged from 0 to 63 points, with higher scores indicating greater severity of depressive symptoms. Previous studies have well documented the reliability and validity of BDI-II [31–33,35,37,39,40,45–48].

2.6. Software

An electronic version of the BDI-II was provided to participants via Google Forms on their smartphones. The question items, answer choices, and question order in the smartphone version of the BDI-II were identical to those in the paper version. All questions for each survey were displayed on the screen, allowing participants to scroll down to answer subsequent questions. Responses were selected by tapping radio buttons on the screen. It was impossible to skip questions or select multiple answers for each item, though participants could revise their previous answers.

2.7. Statistical Analysis

The ISPOR guidelines categorize modifications from PRO to ePRO into three levels, namely Minor, Moderate, and Substantial [8]. Minor modifications involve transferring a scale from paper to screen without significantly altering font size, item content, recall period, or response options. Moderate modifications include splitting items into multiple screens, significantly reducing font size, and requiring patients to scroll to view all text or responses. Substantial modifications include deleting items or drastically changing item text.

In this study, the font size of the smartphone version of the BDI-II administered was different from that of the paper version and scrolling was required to navigate between ques-

tions. These changes correspond to moderate modifications under the ISPOR guidelines. When PROs undergo such modifications, it is recommended to formally establish the equivalence of electronic measures [8]. Therefore, we adopted a randomized crossover design to evaluate the equivalence of the smartphone version of the BDI-II and the paper version of the BDI-II and calculated the $ICC_{agreement}$ and its 95% confidence interval based on a two-way random-effects model. In this study, we adopted a randomized crossover design for several reasons and used $ICC_{agreement}$ as the measure of equivalence. ICC is one of the most commonly used measures for assessing equivalence between ePROs and PROs, and the ISPOR guidelines recommend using a crossover design in such cases [8,13,24,49–51]. Another advantage of using crossover designs is that they improve the power of the test, making it possible to evaluate equivalence with fewer samples [8]. In addition, $ICC_{consistency}$, the Pearson correlation coefficient, and the Spearman correlation coefficient only take random error into account, as opposed to systematic errors; however, $ICC_{agreement}$ considers both random and systematic errors, and it has been suggested that it is an extremely rigorous parameter for evaluating equivalence [8,44,52]. The ICC is expressed as a value between 0 and 1, with an ICC of 0.70 or greater indicating good reliability [44,52]. We also calculated Cronbach's alpha and McDonald's omega and their 95% CIs to assess the internal consistency of paper PRO and ePRO. Cronbach's alpha and McDonald's omega are expressed as 0–1, with values above 0.7 indicating good internal consistency [52,53].

In addition, linear mixed models (LMMs) were used to evaluate carryover effects [54]. A LMM considers the administration format (paper or smartphone), the order of administration, and their interaction as fixed effects, while treating participants as a random effect. Statistical significance was set at $p < 0.05$ with two-tailed tests. All analyses were performed using R version 4.3.1. We used the lmerTest package and the lme4 package to conduct the LMM analysis. In addition, we used the irr package to calculate $ICC_{agreement}$, and the psych and MBESS packages for Cronbach's alpha and McDonald's omega.

3. Results

3.1. Characteristics of the Study Participants

Of the 100 participants who met the eligibility criteria, all completed both the paper- and smartphone-based versions of the BDI-II and provided complete data. Table 1 presents the demographic and lifestyle characteristics of the participants. The mean age was 19.78 years (SD = 0.94), with ten participants (10%) being male. Seven participants (7%) reported a drinking habit, none (0%) had a smoking habit, and 28 participants (28%) engaged in regular exercise.

Table 1. Baseline characteristics of the two groups.

	Total (n = 100)	**Paper-First (n = 50)**	**Smartphone-First (n = 50)**
Characteristics			
Age (years)	19.78 (0.94)	19.78 (0.86)	19.78 (1.02)
Sex			
Male	10 (10%)	5 (10%)	5 (10%)
Female	90 (90%)	45 (90%)	45 (90%)
Drinker			
Yes	7 (7%)	3 (6%)	4 (8%)
No	93 (93%)	47 (94%)	46 (92%)
Smoker			
Yes	0 (0%)	0 (0%)	0 (0%)
No	100 (100%)	50 (100%)	50 (100%)
Exercise habits			
Presence	28 (28%)	17 (34%)	11 (22%)
Absence	72 (72%)	33 (66%)	39 (78%)

Data are presented as means (standard deviation) or numbers (%).

3.2. Mean and LMM Results for the Paper- and Smartphone-First Groups

Table 2 presents the means and LMM results for the paper- and smartphone-first groups. In the paper-first group, the mean BDI-II score for the paper version was 6.82 (SD = 5.01), while in the smartphone-first group, the mean score was 7.78 (SD = 7.24). For the smartphone version, the mean BDI-II score was 5.82 (SD = 4.88) in the paper-first group, and 9.56 (SD = 7.36) in the smartphone-first group. The interaction between administration format and order was not significant in the LMM (p = 0.29; 95% CI −2.20–0.64), indicating that no carryover effect was observed.

Table 2. Mean and LMM results for the paper-first and smartphone-first groups.

	Total (n = 100)	**Paper-First (n = 50)**	**Smartphone-First (n = 50)**	**LMM**			
Outcomes	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Effect**	**Estimate**	**p**	**95% CI**
BDI-II							
Paper	7.30 (6.21)	6.82 (5.01)	7.78 (7.24)	Format	−1	0.06	−2.01–0.01
Smartphone	7.69 (6.49)	5.82 (4.88)	9.56 (7.36)	Order	2.74	0.03	0.30–5.18
				Interaction	−0.78	0.29	−2.20–0.64

BDI-II: Beck Depression Inventory-II
SD: Standard deviation
LMM: Linear mixed models
CI: Confidence interval

3.3. Agreement Between the Smartphone and Paper Versions of BDI-II

Table 3 shows the results of the $ICC_{agreement}$ for the BDI-II scores. The $ICC_{agreement}$ between the paper-based and smartphone-based BDI-II scores was 0.81 (95% CI 0.74–0.87).

Table 3. Agreement between the smartphone and paper versions of the BDI-II.

Outcomes	**$ICC_{agreement}$**	**95% CI**
BDI-II	0.81	0.74–0.87

BDI-II: Beck Depression Inventory-II
ICC: Intraclass correlation coefficient
CI: Confidence interval

3.4. Internal Consistency (Cronbach's Alpha and McDonald's Omega) for the Two Groups

Table 4 shows the Cronbach's alpha and McDonald's omega values for the BDI-II in each group. The Cronbach's alpha for the paper version was 0.88 (95% CI 0.84–0.91) and 0.88 (95% CI 0.84–0.91) for the smartphone version. McDonald's omega for the paper version was 0.88 (95% CI 0.81–0.95) and 0.89 (95% CI 0.85–0.93) for the smartphone version.

Table 4. Internal consistency (Cronbach's alpha and McDonald's omega) for the two groups.

Outcomes	**Cronbach's Alpha**	**95% CI**	**McDonald's Omega**	**95% CI**
BDI-II				
Paper	0.88	0.84–0.91	0.88	0.81–0.95
Smartphone	0.88	0.84–0.91	0.89	0.85–0.93

BDI-II; Beck Depression Inventory-II; CI: confidence interval.

4. Discussion

In addition to the established advantages of PROs, such as reducing missed symptoms and improving communication between patients and healthcare providers, the use of

ePROs is increasing worldwide, driven by advancements in digital technology and the internet [3–10]. ePROs are widely used in the field of mental health, particularly for screening depressive symptoms and assessing their severity, with the BDI-II being one such tool [21–33,35,36]. However, to the best of our knowledge, no studies have evaluated the equivalence between the paper-based and smartphone-based versions of the BDI-II. Given the global rise in smartphone users, the ubiquity of smartphones, and the increasing number of people using them to access the internet, the smartphone version of the BDI-II presents a promising alternative to PCs and other devices [16,42]. Therefore, this study aimed to evaluate the equivalence between the smartphone-based and paper-based versions of the BDI-II, following the ISPOR guidelines.

The $ICC_{agreement}$ for the BDI-II in this study was 0.81 (95% CI 0.74–0.87). Cronbach's alpha was 0.88 (95% CI 0.84–0.91) for both the smartphone and paper versions, indicating comparable reliability. The internal consistency measured by McDonald's omega was 0.89 (95% CI 0.85–0.93) for the smartphone version and 0.88 (95% CI 0.81 to 0.95) for the paper version, suggesting almost identical internal consistency. Previous studies indicate that an ICC, Cronbach's alpha, and McDonald's omega value of 0.70 or higher is considered good [44,52,53]. In addition, the 95% CI results for $ICC_{agreement}$, Cronbach's alpha, and McDonald's omega for the BDI-II in this study suggested a 95% probability that even in the worst-case scenario, these values would remain at 0.7 or higher [55]. Thus, the smartphone version of the BDI-II is likely to yield results similar to those of the paper version. Given the study results and the convenience of smartphones, using the BDI-II via smartphones may be a more promising strategy than using other devices, such as PCs or tablets. However, ICC and Cronbach's alpha should be at least 0.7 for population-level use and between 0.85 and 0.95 for individual-level use [8]. Therefore, while the smartphone version of the BDI-II is suitable for population-level use, it may not be appropriate for individual-level use.

This study provided evidence that the smartphone version of the BDI-II is equivalent to the paper version. However, despite these promising results, the conventional paper-based version may still be more appropriate for use with older populations. Previous studies suggested that older adults' e-health literacy—defined as the ability to effectively use digital technologies with internet access, such as smartphones, wearables, tablets, and computers in order to make health decisions—should be considered when introducing digital technologies [56–59]. Additionally, the utilization rate of e-health services tends to be low in older populations [56,57,59]. A previous study investigating the equivalence of smartphone-based ePROs also indicated that older adults may have been hesitant to participate, either due to limited experience with smartphones or not owing one [24]. Similarly to these findings, it cannot be denied that older individuals may experience resistance to using the smartphone version of the BDI-II. As such, the traditional paper version may still be necessary when administering the BDI-II to older populations. For this reason, it is expected that in both clinical practice and research, the smartphone version and paper version of the BDI-II should be used appropriately, depending on the patient's age.

Our findings contribute to expanding the evidence for using the BDI-II as an ePRO and provide valuable insights for future equivalence studies. To our knowledge, no studies have specifically evaluated whether the BDI-II could be used as an ePRO on smartphones alone, although several studies have explored its use on other devices or across multiple platforms. For instance, the equivalence of the internet and paper versions of the BDI-II was assessed in 87 patients from primary and psychiatric care in Sweden, yielding a Pearson correlation coefficient of 0.89 between the two versions [27]. The Cronbach's alpha for the Internet version ranged from 0.87 to 0.89, while the paper version's Cronbach's alpha was between 0.89 and 0.90 [27]. In addition, the equivalence between the computer version of the BDI-II and the paper version was tested in 180 college students, with a

correlation coefficient of 0.98 [28]. The Cronbach's alphas for the computer and paper versions of the BDI-II were 0.91 and 0.88, respectively [28]. Another study investigating the equivalence of the internet and paper versions of the BDI-II in 494 Swedish patients with panic disorder found a Pearson correlation coefficient of 0.94, with Cronbach's alphas for the internet and paper versions ranging from 0.88 to 0.92 and 0.89–0.90, respectively [29]. A study validating the web-based version of the BDI-II in 185 Spanish university staff also found a Cronbach's alpha of 0.90 [30]. However, several methodological problems in these studies limit the generalizability of the BDI-II as an ePRO. While many studies use the Pearson or Spearman correlation coefficients to investigate equivalence, their measurements are not considered highly rigorous for equivalence evaluation because they do not account for systematic errors [44,52]. Therefore, assuming equivalence between the paper and computer or web versions of the BDI-II based solely on these correlations may be misleading. In addition, for internal consistency, many studies rely on Cronbach's alpha, which is a widely used measure, but assumes that the tau equivalence model is met. This assumption may not hold for many PRO measures, potentially leading to the over- or underestimation of internal consistency [53]. To address these statistical concerns, we used the $ICC_{agreement}$ which accounts for both systematic and random errors. In addition, we calculated McDonald's omega, a more accurate measure of internal consistency than Cronbach's alpha. Consequently, our study provides a more precise validation of BDI-II equivalence and internal consistency, addressing issues that previous studies encountered when adapting the BDI-II to ePRO formats.

Nevertheless, several limitations exist in this study. First, the participants represented a relatively young Japanese population (ages 18–22). Therefore, the results may not be applicable to other countries or age groups. Second, this study did not assess differential item functioning (DIF). The ISPOR guidelines suggest performing item response theory (IRT) analysis to evaluate DIF in the scale [8]. However, conducting a DIF analysis for the BDI-II in this study was challenging because IRT requires a sample of at least 200 individuals [8]. Future studies should evaluate DIF in the smartphone version of BDI-II with a larger sample size. Third, while the crossover design was used, potential carryover cannot be entirely ruled out. Thus, our results should be interpreted with caution. However, based on previous research, a one-week interval between the first and second survey was chosen to minimize carryover effects, and no statistically significant difference in carryover effects was observed. Fourth, minor modifications recommended by the ISPOR guidelines, such as cognitive debriefing and usability testing, were not performed to minimize participant burden [8]. As a result, the usability of the smartphone version of the BDI-II remains unclear. Additionally, previous studies have suggested that healthcare providers may have negative attitudes toward implementing digital technologies in clinical practice [60]. Future research should include cognitive debriefing and usability testing for both the recipients and the providers of the smartphone version of the BDI-II. Fifth, this study was conducted with a sample of 100 people extracted from a single facility. Although the sample size was consistent with the COSMIN recommendations, it remains relatively small [44]. Moreover, participants were recruited from only one facility, which limits the generalizability of our results. Therefore, future research should involve larger sample sizes drawn from multiple facilities to replicate these findings.

5. Conclusions

This study demonstrated the equivalence between the smartphone and paper versions of the BDI-II, indicating that the BDI-II could be effectively used in clinical and research settings through smartphones, which are convenient and user-friendly tools. With the continued growth of digital technology in mental health, the smartphone version of the BDI-

II will be an important asset for mental healthcare professionals. However, the traditional paper versions of the BDI-II may still be recommended for older individuals. Future studies should aim to recruit large, diverse samples from multiple institutions, and replicate the findings of the present study.

Author Contributions: Conceptualization, H.U., T.I., and K.H.; methodology, H.U., T.I., and K.H.; formal analysis, H.U.; investigation, Y.H. and M.T.; resources, K.H.; data curation, H.U., T.I., and K.H.; writing—original draft preparation, H.U., T.I., and K.H.; writing—review and editing, H.U., T.I., Y.H., M.T., K.T., S.K., and K.H.; project administration, K.H.; funding acquisition, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI [grant number 22K11111].

Institutional Review Board Statement: The Ethical Review Board approved this study for Medical Research Involving Human Subjects of Gunma University (approval no. HS2022-109, dated 21 October 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: Due to the restrictions imposed by the Ethics Committee, the data used in this study cannot be publicly shared. However, data can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Narra, L.R.; Verdini, N.; Lapen, K.; Nipp, R.; Gillespie, E.F. Patient-Reported Outcomes in Clinical Trials: From an Endpoint to an Intervention in Cancer Care. *Semin. Radiat. Oncol.* **2023**, *33*, 358–366. [CrossRef] [PubMed]
2. LeBlanc, T.W.; Abernethy, A.P. Patient-reported outcomes in cancer care—Hearing the patient voice at greater volume. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 763–772. [CrossRef]
3. Basch, E.; Barbera, L.; Kerrigan, C.L.; Velikova, G. Implementation of Patient-Reported Outcomes in Routine Medical Care. *Am. Soc. Clin. Oncol. Educ. Book* **2018**, *38*, 122–134. [CrossRef]
4. Snyder, C.F.; Aaronson, N.K.; Choucair, A.K.; Elliott, T.E.; Greenhalgh, J.; Halyard, M.Y.; Hess, R.; Miller, D.M.; Reeve, B.B.; Santana, M. Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations. *Qual. Life Res.* **2012**, *21*, 1305–1314. [CrossRef] [PubMed]
5. Valderas, J.M.; Kotzeva, A.; Espallargues, M.; Guyatt, G.; Ferrans, C.E.; Halyard, M.Y.; Revicki, D.A.; Symonds, T.; Parada, A.; Alonso, J. The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. *Qual. Life Res.* **2008**, *17*, 179–193. [CrossRef]
6. Lewis, C.C.; Boyd, M.; Puspitasari, A.; Navarro, E.; Howard, J.; Kassab, H.; Hoffman, M.; Scott, K.; Lyon, A.; Douglas, S.; et al. Implementing Measurement-Based Care in Behavioral Health: A Review. *JAMA Psychiatry* **2019**, *76*, 324–335. [CrossRef] [PubMed]
7. Austin, D.W.; Carlbring, P.; Richards, J.C.; Andersson, G. Internet administration of three commonly used questionnaires in panic research: Equivalence to paper administration in Australian and Swedish samples of people with panic disorder. *Int. J. Test.* **2006**, *6*, 25–39. [CrossRef]
8. Coons, S.J.; Gwaltney, C.J.; Hays, R.D.; Lundy, J.J.; Sloan, J.A.; Revicki, D.A.; Lenderking, W.R.; Cella, D.; Basch, E. 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*, 419–429. [CrossRef] [PubMed]
9. Meirte, J.; Hellemans, N.; Anthonissen, M.; Denteneer, L.; Maertens, K.; Moortgat, P.; Van Daele, U. Benefits and Disadvantages of Electronic Patient-reported Outcome Measures: Systematic Review. *JMIR Perioper. Med.* **2020**, *3*, e15588. [CrossRef] [PubMed]
10. Campbell, N.; Ali, F.; Finlay, A.Y.; Salek, S.S. Equivalence of electronic and paper-based patient-reported outcome measures. *Qual. Life Res.* **2015**, *24*, 1949–1961. [CrossRef] [PubMed]
11. Davis, R.N. Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behav. Res. Methods Instrum. Comput.* **1999**, *31*, 572–577. [CrossRef] [PubMed]
12. Buchanan, T. Online assessment: Desirable or dangerous? *Prof. Psychol. Res. Pract.* **2002**, *33*, 148–154. [CrossRef]
13. Ali, F.M.; Johns, N.; Finlay, A.Y.; Salek, M.S.; Piguet, V. Comparison of the paper-based and electronic versions of the Dermatology Life Quality Index: Evidence of equivalence. *Br. J. Dermatol.* **2017**, *177*, 1306–1315. [CrossRef]

14. Fritz, F.; Balhorn, S.; Riek, M.; Breil, B.; Dugas, M. Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. *Int. J. Med. Inf.* **2012**, *81*, 303–313. [CrossRef]
15. Bushnell, D.M.; Martin, M.L.; Parasuraman, B. Electronic versus paper questionnaires: A further comparison in persons with asthma. *J. Asthma* **2003**, *40*, 751–762. [CrossRef]
16. Aiyegbusi, O.L. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. *Qual. Life Res.* **2020**, *29*, 325–333. [CrossRef]
17. Marsh, J.D.; Bryant, D.M.; Macdonald, S.J.; Naudie, D.D. Patients respond similarly to paper and electronic versions of the WOMAC and SF-12 following total joint arthroplasty. *J. Arthroplast.* **2014**, *29*, 670–673. [CrossRef] [PubMed]
18. Lee, S.J.; Kavanaugh, A.; Lenert, L. Electronic and computer-generated patient questionnaires in standard care. *Best. Pract. Res. Clin. Rheumatol.* **2007**, *21*, 637–647. [CrossRef] [PubMed]
19. Basch, E.; Stover, A.M.; Schrag, D.; Chung, A.; Jansen, J.; Henson, S.; Carr, P.; Ginos, B.; Deal, A.; Spears, P.A.; et al. Clinical Utility and User Perceptions of a Digital System for Electronic Patient-Reported Symptom Monitoring During Routine Cancer Care: Findings From the PRO-TECT Trial. *JCO Clin. Cancer Inf.* **2020**, *4*, 947–957. [CrossRef] [PubMed]
20. Karamanidou, C.; Natsiavas, P.; Koumakis, L.; Marias, K.; Schera, F.; Schäfer, M.; Payne, S.; Maramis, C. Electronic Patient-Reported Outcome-Based Interventions for Palliative Cancer Care: A Systematic and Mapping Review. *JCO Clin. Cancer Inf.* **2020**, *4*, 647–656. [CrossRef] [PubMed]
21. van Ballegooijen, W.; Riper, H.; Cuijpers, P.; van Oppen, P.; Smit, J.H. Validation of online psychometric instruments for common mental health disorders: A systematic review. *BMC Psychiatry* **2016**, *16*, 45. [CrossRef]
22. Bush, N.E.; Skopp, N.; Smolenski, D.; Crumpton, R.; Fairall, J. Behavioral screening measures delivered with a smartphone app: Psychometric properties and user preference. *J. Nerv. Ment. Dis.* **2013**, *201*, 991–995. [CrossRef]
23. Spangenberg, L.; Glaesmer, H.; Boecker, M.; Forkmann, T. Differences in Patient Health Questionnaire and Aachen Depression Item Bank scores between tablet versus paper-and-pencil administration. *Qual. Life Res.* **2015**, *24*, 3023–3032. [CrossRef]
24. Zhen, L.; Wang, G.; Xu, G.; Xiao, L.; Feng, L.; Chen, X.; Liu, M.; Zhu, X. 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. *Neuropsychiatr. Dis. Treat.* **2020**, *16*, 993–1001. [CrossRef] [PubMed]
25. Erbe, D.; Eichert, H.C.; 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. [CrossRef]
26. Hirao, K.; Takahashi, H.; Kuroda, N.; Uchida, H.; Tsuchiya, K.; Kikuchi, S. Differences in Center for Epidemiologic Studies Depression Scale, Generalized Anxiety Disorder-7 and Kessler Screening Scale for Psychological Distress Scores between Smartphone Version versus Paper Version Administration: Evidence of Equivalence. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4773. [CrossRef] [PubMed]
27. Holländare, F.; Andersson, G.; Engström, I. A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *J. Med. Internet Res.* **2010**, *12*, e49. [CrossRef] [PubMed]
28. Schulenberg, S.E.; Yutrzenka, B.A. Equivalence of computerized and conventional versions of the Beck Depression Inventory-II (BDI-II). *Curr. Psychol.* **2001**, *20*, 216–230. [CrossRef]
29. Carlbring, P.; Brunt, S.; Bohman, S.; Austin, D.; Richards, J.; Öst, L.-G.; Andersson, G. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Comput. Hum. Behav.* **2007**, *23*, 1421–1434. [CrossRef]
30. Asensio-Cuesta, S.; Bresó, A.; Saez, C.; García-Gómez, J.M. Robustness and Findings of a Web-Based System for Depression Assessment in a University Work Context. *Int. J. Environ. Res. Public Health* **2019**, *16*, 644. [CrossRef]
31. Kojima, M.; Furukawa, T.A.; Takahashi, H.; Kawai, M.; Nagaya, T.; Tokudome, S. Cross-cultural validation of the Beck Depression Inventory-II in Japan. *Psychiatry Res.* **2002**, *110*, 291–299. [CrossRef] [PubMed]
32. Kjaergaard, M.; Arfwedson Wang, C.E.; Waterloo, K.; Jorde, R. A study of the psychometric properties of the Beck Depression Inventory-II, the Montgomery and Åsberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a sample from a healthy population. *Scand. J. Psychol.* **2014**, *55*, 83–89. [CrossRef] [PubMed]
33. Hiroe, T.; Kojima, M.; Yamamoto, I.; Nojima, S.; Kinoshita, Y.; Hashimoto, N.; Watanabe, N.; Maeda, T.; Furukawa, T.A. Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. *Psychiatry Res.* **2005**, *135*, 229–235. [CrossRef]
34. Takagaki, K.; Okamoto, Y.; Jinnin, R.; Mori, A.; Nishiyama, Y.; Yamamura, T.; Yokoyama, S.; Shiota, S.; Miyake, Y.; Ogata, A.; et al. Behavioral activation for late adolescents with subthreshold depression: A randomized controlled trial. *Eur. Child. Adolesc. Psychiatry* **2016**, *25*, 1171–1182. [CrossRef] [PubMed]
35. Wang, Y.P.; Gorenstein, C. Assessment of depression in medical patients: A systematic review of the utility of the Beck Depression Inventory-II. *Clinics* **2013**, *68*, 1274–1287. [CrossRef]
36. Leung, K.Y.Q.; Cartoon, J.; Hammond, N.E. Depression screening in patients with aneurysmal subarachnoid haemorrhage and their caregivers: A systematic review. *Aust. Crit. Care* **2023**, *36*, 1138–1149. [CrossRef]
37. Furukawa, T.A. Assessment of mood: Guides for clinicians. *J. Psychosom. Res.* **2010**, *68*, 581–589. [CrossRef] [PubMed]

38. Beck, A.T.; Steer, R.A.; Brown, G. Beck depression inventory–II. *Psychol. Assess.* **1996**, *1*, 210.
39. Mostafa Alim, S.; Ahmed, M.N.; Mullick, M.S.I.; Chowdhury, N.F.; Akhter, F.; Alam, M.S. Validation of the Bangla version of Beck Depression Inventory-II. *Brain Behav.* **2020**, *10*, e01563. [CrossRef]
40. Williams, Z.J.; Everaert, J.; Gotham, K.O. Measuring Depression in Autistic Adults: Psychometric Validation of the Beck Depression Inventory-II. *Assessment* **2021**, *28*, 858–876. [CrossRef]
41. Schutt, P.E.; Kung, S.; Clark, M.M.; Koball, A.M.; Grothe, K.B. Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) Depression Measures in an Outpatient Bariatric Clinic. *Obes. Surg.* **2016**, *26*, 1274–1278. [CrossRef] [PubMed]
42. Boettcher, J.; Magnusson, K.; Marklund, A.; Berglund, E.; Blomdahl, R.; Braun, U.; Delin, L.; Lundén, C.; Sjöblom, K.; Sommer, D.; et al. Adding a smartphone app to internet-based self-help for social anxiety: A randomized controlled trial. *Comput. Hum. Behav.* **2018**, *87*, 98–108. [CrossRef]
43. Walter, S.D.; Eliasziw, M.; Donner, A. Sample size and optimal designs for reliability studies. *Stat. Med.* **1998**, *17*, 101–110. [CrossRef]
44. Prinsen, C.A.C.; Mokkink, L.B.; Bouter, L.M.; Alonso, J.; Patrick, D.L.; de Vet, H.C.W.; Terwee, C.B. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual. Life Res.* **2018**, *27*, 1147–1157. [CrossRef] [PubMed]
45. Smarr, K.L.; Keefer, A.L. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care Res.* **2011**, *63* (Suppl. 11), S454–S466. [CrossRef]
46. Lee, E.H.; Lee, S.J.; Hwang, S.T.; Hong, S.H.; Kim, J.H. Reliability and Validity of the Beck Depression Inventory-II among Korean Adolescents. *Psychiatry Investig.* **2017**, *14*, 30–36. [CrossRef]
47. Osman, A.; Kopper, B.A.; Barrios, F.; Gutierrez, P.M.; Bagge, C.L. Reliability and validity of the Beck depression inventory--II with adolescent psychiatric inpatients. *Psychol. Assess.* **2004**, *16*, 120–132. [CrossRef] [PubMed]
48. Toledano-Toledano, F.; Contreras-Valdez, J.A. Validity and reliability of the Beck Depression Inventory II (BDI-II) in family caregivers of children with chronic diseases. *PLoS ONE* **2018**, *13*, e0206917. [CrossRef] [PubMed]
49. Nishimura, K.; Kusunose, M.; Sanda, R.; Tsuji, Y.; Hasegawa, Y.; Oga, T. Comparison between electronic and paper versions of patient-reported outcome measures in subjects with chronic obstructive pulmonary disease: An observational study with a cross-over administration. *BMJ Open* **2019**, *9*, e032767. [CrossRef]
50. van Kollenburg, R.A.A.; de Bruin, D.M.; Wijkstra, H. Validation of the Electronic Version of the International Index of Erectile Function (IIEF-5 and IIEF-15): A Crossover Study. *J. Med. Internet Res.* **2019**, *21*, e13490. [CrossRef] [PubMed]
51. Robles, N.; Rajmil, L.; Rodriguez-Arjona, D.; Azuara, M.; Codina, F.; Raat, H.; Ravens-Sieberer, U.; Herdman, M. Development of the web-based Spanish and Catalan versions of the Euroqol 5D-Y (EQ-5D-Y) and comparison of results with the paper version. *Health Qual. Life Outcomes* **2015**, *13*, 72. [CrossRef] [PubMed]
52. Terwee, C.B.; Bot, S.D.; de Boer, M.R.; van der Windt, D.A.; Knol, D.L.; Dekker, J.; Bouter, L.M.; de Vet, H.C. Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* **2007**, *60*, 34–42. [CrossRef]
53. Dunn, T.J.; Baguley, T.; Brunsden, V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* **2014**, *105*, 399–412. [CrossRef]
54. Gueorguieva, R.; Krystal, J.H. Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Arch. Gen. Psychiatry* **2004**, *61*, 310–317. [CrossRef] [PubMed]
55. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]
56. Elgamal, R. Meta-analysis: eHealth literacy and attitudes towards internet/computer technology. *Patient Educ. Couns.* **2024**, *123*, 108196. [CrossRef] [PubMed]
57. Verma, R.; Saldanha, C.; Ellis, U.; Sattar, S.; Haase, K.R. eHealth literacy among older adults living with cancer and their caregivers: A scoping review. *J. Geriatr. Oncol.* **2022**, *13*, 555–562. [CrossRef]
58. van der Vaart, R.; Drossaert, C.H.; de Heus, M.; Taal, E.; van de Laar, M.A. Measuring actual eHealth literacy among patients with rheumatic diseases: A qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. *J. Med. Internet Res.* **2013**, *15*, e27. [CrossRef]
59. Tennant, B.; Stellefson, M.; Dodd, V.; Chaney, B.; Chaney, D.; Paige, S.; Alber, J. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. *J. Med. Internet Res.* **2015**, *17*, e70. [CrossRef]
60. Giannouli, V.; Stoyanova, S.; Drugas, M.; Ivanova, D. Attitudes towards e-Health during the COVID-19 Pandemic: Untangling the Gordian Knot in Greece, Bulgaria and Romania in Healthcare Professionals and Students? *Psychiatr. Danub.* **2021**, *33*, 415–419. [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.