

The Medical Outcome Study-HIV Health Survey: A Systematic Review and Reliability Generalization Meta-analysis

Abstract

The Medical Outcome Study-HIV (MOS-HIV) is one of the most used questionnaires for the evaluation of the health-related quality of life (HRQoL) in people living with HIV (PLWHIV) in both medical settings and research studies. This study aimed to estimate the average reliability of the Medical Outcome Study-HIV (MOS-HIV) scores and to evaluate the characteristics of the studies that could explain the variability between reliability estimates. Furthermore, the study aimed to estimate the induction rate of the reliability of the MOS-HIV. A systematic review of the previous literature, including studies that reported alpha and/or test-retest coefficients with the data at hand for the total score of the MOS-HIV and the subscales, was conducted. Fifty studies (52 samples; N=14,132) were included in the reliability generalization meta-analysis. The average alpha coefficient for the total score of MOS-HIV was .91 and above .80 for all of the subscales, except for role functioning, which had an average coefficient of .76. Different study dimensions were related to the heterogeneity of reliability between studies. Reliability induction was found to be 76.1%. The results obtained in the present study indicate that the MOS-HIV is a reliable instrument for HRQoL evaluation in PLWHIV, for clinical and research purposes. In the clinical practice of health services, nurses could employ this gold standard for reliable evaluations of HRQoL in PLWHIV.

Keywords: Medical Outcome Study-HIV (MOS-HIV), Health-Related Quality of Life, HIV, Meta-analysis, systematic review, reliability generalization, people living with HIV (PLWHIV)

Introduction

According to the World Health Organization (WHO) as of 2018, 36.9 million people were living with HIV globally (Murillo-Bonilla, Godínez, Cruz, & González, 2018). The prevalence of HIV was highest in Africa, accounting for almost 70% of this population (25.7 million), followed by Asia (9.4%), America (9.2%), and Europe (6.2%). The majority of new infections occurred in men (83.9%) and men who have sex with men (MSM) accounted for 53.1% of the diagnoses («World Health Organization, HIV/AIDS», n. d.).

The semiology is dependent on viral load and the number of CD4 lymphocytes. As the infection progresses, fever, myalgias, diarrhea, anemia, renal failure, skin lesions, pneumonia, lymphomas, progressive reduction of cognitive and motor functions, sleep disorders, and chronic pain can appear, affecting the general state of health (Hudson, Kirksey, & Holzemer, 2004; Portillo, Holzemer, & Chou, 2007; Zhu, Zhao, & Hu, 2019). These symptoms, which are associated with rejection and social stigma, cause an increase in stress, anxiety, and depression in this group, significantly reducing the health-related quality of life (HRQoL) in people living with HIV (PLWHIV) (Burgoyne & Saunders, 2001; Cardona-Arias & Higuaita-Gutiérrez, 2014; Cooper, Clatworthy, Harding, Whetham, & Emerge Consortium, 2017; Dunne et al., 2019; Duong, Torre, Springer, Cox, & Plankey, 2016; Gielen, McDonnell, Wu, O'Campo, & Faden, 2001; Hudson, Kirksey, & Holzemer, 2004; Melhuish & Lewthwaite, 2018; Sánchez & Guiteras, A. F.).

Low HRQoL can be a repercussion of a pathological condition on the health status. The subjective perception of the negative influence of such conditions can affect physical, psychological, and social aspects (Castellvi, Ferrer, & Alonso, 2013). Specifically, in PLWHIV, this term also refers to the subjective perception that this population has about the disease and their quality of life when it is influenced by the disabilities or physical, psychological, and social limitations caused by HIV, considering the consequences of both the disease and the treatment (Castillo & Luz, 2006; Schwartzmann, 2003).

As indicated above, previous studies have widely demonstrated that PLWHIV show a significant decrease in their well-being and quality of life (Burgoyne & Saunders, 2001; Dunne et al., 2019; Gielen et al., 2001). This has a significant negative impact on the global health status of PLWHIV, as HRQoL can be a significant predictor of disability, treatment adherence, clinical progression, and impairment of psychosocial functioning (O'Brien, Bayoumi, Strike, Young, & Davis, 2008; Ryu, West, & Sousa, 2009). Considering this, the assessment of HRQoL in PLWHIV using reliable and valid instruments is crucial.

Although studies have generally shown a general impairment of HRQoL in this population, most of them have used generic HRQoL assessment tools for evaluation, despite the limitations (Cooper et al., 2017). Considering that the idiosyncratic characteristics of each pathology can lead to a specific affectation of HRQoL, it is necessary to evaluate the HRQoL in PLWHIV through specific and validated instruments. Although there are various evaluation instruments to analyze HRQoL in PLWHIV (Cooper et al., 2017), the most widely applied instrument in clinical and research contexts is the Medical Outcome Study-HIV Health Survey (MOS-HIV), as indicated in several reviews (Clayson et al., 2006; Cooper et al., 2017; Drewes, Gusy & Rüden, 2013; Simpson et al., 2013, Wu et al., 2013). The MOS-HIV, used to assess HRQoL, is a tool designed by Albert Wu and colleagues (Wu et al., 1997) to evaluate the physical and emotional state of PLWHIV. During the initial development of the MOS-HIV, previous tools designed for the evaluation of medical results were employed, serving as a base for the 20-item short-form (SF-20) questionnaire (Stewart, Hays, & Ware, 1988). After the implementation of the MOS-HIV, and the analyses of the items, opinions, and contributions of patients and volunteers of clinical trials about the characteristics of HIV/AIDS were considered. Thus, the subscales energy/fatigue, cognitive functioning, health distress, and quality of life were added to the original questionnaire, leading to a questionnaire comprising 30 items (MOS-30) (Wu et al., 1991). Furthermore, four new items for assessing general health and one item for assessing pain were added. Considering the modifications to the original questionnaire, the MOS-HIV questionnaire consists of 35 items.

The MOS-HIV measures the following 11 subscales of HRQoL: general health perceptions (5 items), pain (2 items), physical functioning (6 items), role functioning (2 items), social functioning (1 item), mental health (5 items), energy/fatigue (4 items), cognitive functioning (4 items), health distress (4 items), quality of life (1 item), and health transition (1 item). The score obtained in these subscales can be quantified through the calculation of two general indexes: physical health summary (PHS) and mental health summary (MHS). Concretely, the PHS summary is composed by general health perceptions, pain, physical functioning, role functioning, social functioning, and energy/fatigue subscales and the MHS summary is composed by mental health, cognitive functioning, health distress, quality of life, and energy/fatigue subscales. The items refer to the last two weeks and are rated on 2-, 3-, 5-, and 6-point scales with a final score reflected on a scale from 0 to 100. Higher scores indicate greater health (Wu et al., 1997).

The MOS-HIV has been widely applied in both clinical practice and research studies (Cooper et al., 2017). In effect, it has been published in different cultures and languages such as Chinese (Huang, Tian, Dai, & Ye, 2013), Dutch,

Italian, German, French, UK English (Scott-Lennox, Wu, Boyer, & Ware, 1999), Korean (Shim et al., 2018), Greek (Stasinopoulou et al., 2010), Persian (Alamolhoda, Jafari, Ayatollahi, & Kazerooni, 2017) and Spanish (Badia, Podzamczer, Garcia, López-Lavid, & Consiglio, 1999), showing good psychometric properties. Most studies that have applied the MOS-HIV have reported good internal consistency for the total scale and also for the subscales (coefficients alpha higher than .70) (Burgess, Dayer, Catalan, Hawkins, & Gazzard, 1993; Revicki et al., 1998; Revicki, Swartz, Wu, Haubrich, & Collier, 1999; Revicki, Wu, & Murray, 1995; Safrin et al., 1996; Wu et al., 1997). The reliability of the summary scores has also been satisfactory, with coefficients alpha ranging from .90 to .92 for PHS and from .91 to .94 for MHS (Revicki et al., 1998).

Reliability is a psychometric property that must be calculated in each application of the instrument (Crocker & Algina, 1986; Lord & Novick, 1968; McDonald, 1999; Streiner, Norman, & Cairney, 2015). Nonetheless, frequently, several studies have induced reliability from previous studies in which the test had been applied to other samples (Henson & Thompson, 2002). According to Shields and Caruso (Shields & Caruso, 2004), two forms of reliability induction could be identified: (a) reliability induction by omission, when the authors did not make any reference to the reliability of the applied test, or (b) reliability induction by the report, when the authors included the reliability obtained from previous studies. As reliability varies with each test administration, the reliability generalization (RG), coined by Vacha-Haase (Vacha-Haase, 1998), is a meta-analytic procedure used to statistically compute the reliability estimates of various studies in which the test had been applied (e.g., in different contexts and different types of people). When variability in the reliability coefficients is found, an RG meta-analysis allows us to explore characteristics of the studies that might be related to the reliability coefficients to examine, for example, the samples and contexts for which the instrument is the most reliable. The precise and suitable assessment of HRQoL in PLWHIV is significantly relevant for assessing the physical and emotional state of these patients. Thus, it is necessary to understand the psychometric properties of the scores of a measurement instrument widely used in practice by both nursing methodologists and clinicians. Similarly, an RG meta-analysis plays a key role in improving our knowledge about the reliability of test scores, and in helping methodological and clinical nurses to select the most accurate instruments available for evaluating HRQoL in research studies and clinical practice.

To our knowledge, there are no previous studies that exhaustively analyzed the reliability of the different applications of the MOS-HIV in different contexts and groups of PLWHIV. Thus, the aim of this research was to conduct a systematic review and RG meta-analysis of the applied studies that have used the MOS-HIV. The specific

objectives of the present research were: (a) to estimate the mean reliability (for the total scales and the subscales of the MOS-HIV), (b) to analyze the variability within the reliability estimates, and (c) to detect the characteristics of the studies that could influence the heterogeneity of the reliability estimates. Moreover, the present study aimed to estimate the reliability induction rate of the MOS-HIV.

Methods

The current systematic review and meta-analysis were conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009). Supplementary material presents the PRISMA checklist for this meta-analysis.

Selection criteria

The criteria for the selection of studies to be included in the systematic review and meta-analysis were: (1) an empirical research where the MOS-HIV, or an adaptation of the MOS-HIV maintaining the 35-item structure, was applied; (2) reporting the reliability of the MOS-HIV or any of its subscales (internal consistency, test–retest) based on the study-specific sample; (3) written in English or Spanish; and (4) published and evaluated by experts. The following exclusion criteria were applied: (1) meta-analyses and systematic reviews, and (b) studies that applied any other version that did not retain the 35-item structure of the MOS-HIV.

Search strategies

Electronic databases (PubMed, ProQuest, and Web of Science) were searched in January 2019. The keywords “Medical Outcomes Study HIV Health Survey” OR “MOS-HIV” were linked with every electronic database for all fields. No limits on the date of publication of the study were established. Moreover, references of the recovered studies were assessed to identify studies that may fulfill the selection criteria. The procedure for the determination of eligibility of the study was conducted independently by two researchers (C.A.B and M.R.A). The discrepancies were resolved to reach agreement.

Data extraction

A protocol for extracting the characteristics of the studies was established by distinguishing among three broad categories of potential moderator variables: substantive, methodological, and extrinsic (Lipsey, 2009). The following substantive characteristics were coded: (a) mean and standard deviation of the total scores and each of the subscales of the MOS-HIV, (b) mean and standard deviation of the participants' age (in years), (c) gender (% male), (d) ethnicity (% Caucasian), (e) mean and standard deviation of time since diagnosis, (f) percentage of participants with detectable viral load (%), (g) percentage of participants with previous diagnosis of AIDS, (h) percentage of participants with HIV symptomatology, (i) mean CD4 count, and (j) sample setting (hospital, outpatient, community). To extract methodological characteristics, the following variables were coded: (k) MOS-HIV version (English original vs. other), (l) study focus (psychometric vs. applied), (m) type of design, (n) sample size, and (o) administration format (self-report vs. clinical interview). Finally, two additional extrinsic variables were coded: (p) continent in which the study was performed, and (q) year of the study.

Along with these characteristics, reliability coefficients (alpha and test-retest) were retrieved for the total scores and the subscales of the MOS-HIV.

The codification of the characteristics of the studies was carried out by two researchers independently (C.A.B and M.R.A). In general, the results showed a high agreement, with kappa coefficients ranging from .81 to 1 (M=.92) for qualitative variables, and intraclass correlations ranging from .89 to 1 (M=.95) for continuous variables. Disagreements between the two researchers that coded these characteristics were resolved.

Reliability estimates

In the meta-analysis, two kinds of reliability coefficients were considered: the alpha coefficient to evaluate the internal consistency reliability of the measures and the Pearson correlation coefficient to identify test-retest temporal reliability. Therefore, alpha coefficients and Pearson correlations were extracted from the MOS-HIV score and for each subscale. To normalize their distributions and stabilize their variances, both types of reliability coefficients were transformed before the meta-analytic integration. Bonett's formula (2002) was applied to transform the alpha coefficients, α , and Fisher's Z transformation (Sánchez-Meca, López-López, & López-Pina, 2013) was used for Pearson correlation coefficients, r .

Statistical analyses

Independent meta-analyses were performed for each reliability estimate (alpha coefficients and test-retest reliability coefficients). Furthermore, independent meta-analyses were carried out for the reliability coefficients retrieved from the total score of the MOS-HIV and for each subscale. The random-effects model was used to compute the average reliability coefficients in separate meta-analyses (Borenstein, Hedges, Higgins, & Rothstein, 2011; Sánchez-Meca et al., 2013). This model implies that each reliability coefficient is weighted by the inverse variance, defined as the sum of the within-study and between-studies variance. The latter was estimated by restricted maximum likelihood (López-López, Botella, Sánchez-Meca, & Marín-Martínez, 2013). All the statistical analyses were performed using the Bonett (for alpha coefficients) and Fisher's Z (for test-retest coefficients) transformations. The sampling variances of the transformed Bonett coefficients were calculated using Bonett (2002), and Fisher's Z transformations using: $V(Z) = 1/(n-3)$, with n being the sample size. A 95% confidence interval around the average reliability coefficient was calculated using the improved method proposed by Hartung (Hartung, 1999; Sánchez-Meca & Marín-Martínez, 2008). The mean reliability coefficients and their confidence limits, computed with Bonett's or Fisher's transformations, were back-transformed into the alpha coefficient and Pearson correlation metrics, respectively, to make the interpretation of the results easier.

To assess the variability among the reliability coefficients, forest plots were developed and the Q statistic and I^2 index were calculated in each meta-analysis. The Q statistic was used to evaluate the homogeneity of hypothesis among the reliability coefficients (Borenstein et al., 2011). The I^2 index quantifies the amount of variability shown by the reliability coefficients. Values of 25%, 50%, and 75% correspond to low, moderate, and large heterogeneity, respectively (Higgins, Thompson, Deeks, & Altman, 2003). The forest plot is a graph that represents the individual reliability coefficients by a square and their 95% CIs by horizontal lines, together with the average reliability coefficient plotted as a diamond and lateral points indicating its 95% CI.

Using mixed-effects models, the effect of moderator variables on the reliability coefficients was conducted for meta-analyses with at least 20 reliability coefficients when the presence of variability was identified (López-López et al., 2013; Sánchez-Meca et al., 2013). Weighted ANOVAs and meta-regressions were carried out for categorical and continuous moderators (Rubio-Aparicio, Sánchez-Meca, López-López, Marín-Martínez, & Botella, 2017; Viechtbauer, López-López, Sánchez-Meca, & Marín-Martínez, 2015). The method developed by Knapp and Hartung (Knapp & Hartung, 2003) was used to study the statistical significance of the moderators and the R^2 was computed to estimate the proportion of variance explained by each moderator (López-López, Marín-Martínez, Sánchez-Meca,

Noortgate, & Viechtbauer, 2014). As a further step, multiple meta-regression models were used to identify the set of study characteristics that best explained variability among the reliability coefficients. The moderator variables included in the model were selected considering the following three criteria: statistical significance achieved in the previous moderator analyses, substantive relevance, and a smaller presence of missing data. Moreover, to avoid inflation of Type I error rates, the ratio used was the inclusion of a moderator variable in 10 studies (Borenstein et al., 2011).

The Egger test and the construction of funnel plots with the trim-and-fill method were performed to analyze the risk of publication bias (Duval & Tweedie, 2000). The funnel plot displays the reliability coefficients on the x -axis and the standard error in the y -axis to check for the existence of publication bias. A symmetrical funnel plot indicates the absence of publication bias.

Finally, a sensitivity analysis was performed to assess whether our meta-analytic results were substantially influenced by the presence of outlier reliability coefficients (e.g., reliability coefficients below .60) by systematically removing each study and recalculating the significance of the overall results.

The package *metafor* in *R* was employed to carry out all statistical analyses (Viechtbauer, 2010).

Results

Selection process and reliability induction

Figure 1 presents the flow diagram of the screening and the procedure for the selection of the studies. A total of 797 results were obtained. After reading titles and abstracts, and eliminating duplicates, 357 references were preselected based on the inclusion and exclusion criteria. After reviewing the full text, 264 empirical studies were selected. Of the 264 studies that applied the MOS-HIV, only 63 reported some type of reliability coefficient, whereas the other 201 induced reliabilities from previous research, which entails a 76.1% reliability induction for this test. Of the 201 studies that induced reliability, 125 (62.2%) did not include any reference to the MOS-HIV scores reliability, and 76 (37.8%) studies induced reliability from previous investigations.

(Insert figure 1)

Characteristics of the studies

The present RG study analyzed 63 studies that reported reliability coefficients from the data at hand. Out of the 63 studies, 13 of them could not be included in the meta-analysis for several reasons (see Figure 1): seven studies reported a range of reliability coefficients (Bekele et al., 2013; Bucciardini et al., 2007; Burrage & Demi, 2003; George, Bergin, Clarke, Courtney, & Codd, 2016; Kamen et al., 2012; Maluccio, Palermo, Kadiyala, & Rawat, 2015; Tozzi et al., 2003), five studies stated that the reliability coefficients were higher than the cut-off value (Baigis et al., 2002; Bunjounmanee, Chunloy, Tangsathapornpong, Khawcharoenporn, & Apisarntharak, 2014; Golub, Tomassilli, & Parsons, 2009; Mutabazi-Mwesigire, Katamba, Martin, Seeley, & Wu, 2015; Yu, 2017) and one study (Badia et al., 1999) administered the MOS-HIV to the same participants of another study included already (Badia, Podzamczek, López-David, & García, 1999). Thus, 50 studies were included in the RG meta-analysis. Two studies (Curtin et al., 2019; Henderson et al., 2010) reported reliability coefficients from two different samples, the database of our RG study included a total of 52 independent samples.

Table S1 (Supplementary Material) contains the main characteristics of the 50 studies (52 samples) included in the meta-analysis and the references of these studies.

The total number of participants was $N=14,132$ (min.=46; max.=899), with an average of 272 participants per sample (median =204; SD=209). The mean of participants' ages ranged from 26.68 to 54.50 ($M=40.46$; $SD=5.38$), standard deviations of participants' ages ranged from 4.20 to 14.22 ($M=8.23$; $SD=1.88$), and the percentage of males from 0 to 100 ($M=59.81$; $SD=27.64$). With regard to the locations in which the studies were conducted, four continents were identified: North America, 22 samples (42.3%), Asia, 15 samples (28.8%), Africa, 8 samples (15.4%), and Europe, 7 samples (13.5%). Finally, it was found that 27 samples (51.9%) used the MOS-HIV original version, and 25 samples (48.1%) used adaptations of MOS-HIV to other languages and cultures. All studies were written in English.

Mean reliability and heterogeneity

Meta-analytic computations were conducted for eight of the eleven subscales of the MOS-HIV, as three subscales (Social Functioning, SF; Quality of Life, QL; and Health Transition, HT) were composed of only one item.

Table 1 shows the mean alpha coefficients obtained for the total scores for the two summary scores (physical health summary, PHS, and mental health summary, MHS) and the eight subscales of the MOS-HIV. A mean coefficient of 0.91 (95%CI: 0.85. 0.94) was reached by the 10 samples that reported alpha coefficients for the total score.

Regarding the two summary scores, the mean alpha coefficient obtained for the MHS was 0.86 (95%CI: 0.43. 0.97),

slightly higher than the mean estimate of 0.83 (95%CI: 0.67 and 0.91) for the PHS. The mean alpha coefficients for the subscales ranged from 0.76 (95%CI: 0.69. 0.81) for the Role Functioning (RF) subscale to 0.89 (95%CI: 0.87 and 0.91) for the Health Distress (HD) subscale. The supplementary material contains the forest plots constructed for the MOS-HIV total scores, the PHS and MHS summary scores, and the eight subscale scores in each study (see Figures from S1 to S11).

Alpha coefficients for the total scale, summary scores, and subscales exhibited a statistically significant heterogeneity ($p < .05$) in all cases I^2 above 85% (see Table 1). The variability found in the alpha coefficients is also reflected in the forest plots (see Figures from S1 to S11 in Supplementary Material).

(Insert Table 1)

A sensitivity analysis was carried out by deleting the alpha coefficients below .60 to check if extreme values affected the mean alpha coefficients. Alpha coefficients $<.60$ were found for the PHS and MHS summaries, and GHP and RF subscales (see Table S1 in Supplementary Material). For the PHS summary, when a mean alpha coefficient (and its 95%CI) was recalculated with the $k = 5$ coefficient alpha, the mean alpha coefficient increased to $\alpha_+ = .86$ (95%CI: .74 and .92). For the MHS summary, when a mean alpha coefficient (and its 95%CI) was recalculated with the $k = 3$ coefficients, the mean alpha coefficient increased to $\alpha_+ = .91$ (95% CI: .74 and .97). The recalculated mean alpha coefficients with $k = 23$ and $k = 21$ alpha coefficients for the GHP and RF subscales were $\alpha_+ = .82$ (95%CI: .79 and .84) and $\alpha_+ = .79$ (95%CI: .75 and .84), respectively. When the new mean reliability coefficients were compared with those obtained with the original alpha coefficients for the PHS and MHS summaries, and the GHP and RF subscales (see Table 1), small increases were noted.

Although the statistical analyses were conducted using Bonett's transformation for alpha coefficients, Table 1 and Supplementary Figures (from S1 to S11) display the averages and their confidence intervals after being back-transformed to the alpha coefficient metric to facilitate their interpretation.

Regarding the test-retest reliability, no study reported test-retest reliabilities for the total scale or for the PHS and MHS summaries; only two estimates were reported for the eight subscales (see Table S1 in Supplementary Material). Although we intended to include in our meta-analytic computations test-retest coefficients, the scarce references (e.g., Badia et al., 1999; Liu, Zhu, & Qu, 2018) that reported this type of reliability did not allow us to conduct the meta-analytic computations.

Analyses of the moderator variables

As the alpha coefficients for the General Health Perceptions (GHP), Pain (P), Physical Functioning (PF), Role Functioning (RF), Energy/Fatigue (EF), Mental Health (MH), Health Distress (HD), and Cognitive Functioning (CF) subscales presented more than 20 reliability estimates, the analyses of moderator variables were performed for these indicated subscales. Weighted ANOVAs and simple meta-regression were performed for continuous and categorical variables, respectively. The results regarding the analyses of both continuous and categorical moderator variables can be found in the supplementary material (see Tables from S2 to S17).

Regarding continuous moderator variables, the standard deviation of age exhibited a positive and statistically significant relationship with the alpha coefficients for the GHP and HD subscales, with a percentage of variance explained of 37% and 38%, respectively (see Tables S2 and S14, respectively). For the HD subscale, the mean age of the samples showed a positive association with the alpha coefficients, with an R^2 of 39% (see Table S14). The positive sign of the regression coefficients of these two moderators indicated a higher standard deviation of age and mean age were associated with higher reliability.

The mean of the PF and EF subscale scores reached a negative association with the alpha coefficients ($p = .036$; $R^2 = .17$ for PF, and $p = .027$; $R^2 = .19$ for EF), where higher PF and EF subscale scores were associated with lower alpha coefficients (as can be seen in Tables S6 and S10, respectively).

Another continuous moderator that yielded a statistically significant association with the alpha coefficients was the percentage of males. The positive relationship found means that higher alpha coefficients were obtained as the number of male participants increased. In particular, the percentages of variance explained were 21% for the EF subscale (see Table S10), 32% for the MH subscale (see Table S12), and 33% for the CF subscale (see Table S16).

Moreover, the mean CD4 count also proved to be a continuous moderator related statistically significant to the alpha coefficients for the MH scale ($p = .021$) with a percentage of variance explained by 80% (see Table S12). The negative relationship indicates that the larger the mean CD4 count, the larger is the reliability.

With regard to qualitative moderator variables, the MOS-HIV version, classified as original English versus other languages, had a statistically significant influence on the alpha coefficients through most subscales of the MOS-HIV (GHP, P, PF, EF, and CF). The pattern found was similar in all subscales, with larger overall reliabilities for the original English version than for other languages (as can be seen in Tables S3, S5, S7, S11, and S17). For the PF and CF subscales, the place where the study was carried out showed a statistically significant association with the alpha coefficients ($p = .012$; $R^2 = .29$ for PF, and $p = .016$; $R^2 = .30$ for CF), with larger mean reliabilities for studies performed in North America and Europe, and lower means yielded by those carried out in Africa and Asia (see Tables S7 and S17). Finally, the form of administering the questionnaire also revealed a statistically significant relationship with the alpha coefficients for the PF subscale ($p = .026$), explaining 24% of the variance, with the largest mean alpha coefficients for self-report and the lowest mean for clinical interview (see Table S7).

Multiple meta-regression models

To explain part of the heterogeneity among the alpha coefficients when moderator variables were analyzed individually (see Tables from S2 to S17), weighted multiple meta-regression analyses were also performed. The results regarding the multiple meta-regression models linked with general health perceptions (GHP), physical functioning (PF), energy/fatigue (EF), health distress (HD), and cognitive functioning (CF) subscales can be found in the supplementary material (see Tables from S18 to S22). Regarding the GHP subscale, two moderators were included in the model: standard deviation of age and the test version. A statistically significant result was found for the full model ($p = .010$), which accounted for 43% of the variance; only the standard deviation of age reached statistical significance ($p = .033$) once the influence of test version was controlled for (see Table S18). For both PF and EF subscales, the predictors included in the meta-regression models were the mean of the scores and the test version. Both models yielded a statistically significant difference ($p = .017$ for PF and $p = .001$ for EF) explaining a large percentage of variance (34% for PF and 51% for EF), with the test version as the predictor that showed a statistically significant result once the mean of the scores for PF and EF subscales were controlled (see Tables S19 and S20). A statistically significant result was also found for the full model fitted for the HD subscale ($p = .009$),

with a 40% variance accounted for by the two moderators included in the model: the mean age and the percentage of males. Moreover, only the mean age exhibited a positive, statistically significant result ($p = .033$) with the alpha coefficient for HD (see Table S21). Finally, regarding the CF subscale, the percentage of males and the test version were included in the meta-regression model. The full model reached a statistical significance ($p = .000$) and explained 50% of the variance, and both predictors reached statistical significance after controlling for the influence of the other (see Table S22).

Publication bias

Publication bias was evaluated for the general health perceptions (GHP), pain (P), physical functioning (PF), role functioning (RF), energy/fatigue (EF), mental health (MH), health distress (HD), and cognitive functioning (CF) subscales. The supplementary material presents the funnel plots-derived trim and fill method obtained for the eight subscales (see Figures from S12 to S19). When the trim-and-fill method was applied to achieve symmetry in the graphs, imputed reliability coefficients were added to three subscales: CF, MH, and HD subscales where one, two, and three reliability coefficients were imputed, respectively. Lastly, non-significant results for the intercepts of the eight subscales were obtained using Egger's regression ($p = .217$ for GHP; $p = .786$ for P; $p = .943$ for PF; $p = .472$ for RF; $p = .798$ for E/F; $p = .381$ for MH; $p = .548$ for HD; and $p = .997$ for CF). Consequently, publication bias could be discarded as a risk factor for the obtained results.

Discussion

The present RG meta-analysis aimed to estimate the average reliability coefficients of the MOS-HIV and identify the characteristics of the studies that could be associated with the variability in the reliability coefficients.

For the total scale, the average reliability coefficients were .91, .83, and .86 for the summary scores, PHS, and MHS, respectively. For the average reliability of the different subscales of the MOS-HIV, most of them were above .80, except in the case of RF, with an average coefficient of .76. Although the number of items is not large for the subscales (most of them with less than 6 items), on average, the mean reliability coefficients were highly satisfactory. Based on the criteria proposed by Nunnally and Bernstein (1994), reliability coefficients above .70 are adequate for exploratory research. In the case of clinical practice, coefficients higher than .90 are recommended (Charter, 2003; Nunnally & Bernstein, 1994). Moreover, to evaluate the clinical relevance of the alpha coefficients

of test scores, Cicchetti (1994) recommended the following ranges: unacceptable for coefficients lower than .70, fair for the range from .70 to .80, good for .80 to .90, and excellent for values over .90. In this context, the obtained total score and scores from the subscales of the instrument can be considered as excellent and good, respectively, being fair for RF. For research purposes, both the total score and scores from the subscales exhibited excellent reliability coefficients, being adequate for RF.

Regarding the heterogeneity of the evaluated alpha coefficients, a large variability has been found among the reliability coefficients reported by the evaluated studies. This shows that the scores of this questionnaire could not be generalized, as it seems to depend on the application context, the evaluated population, the characteristics of the study, or the methodology of application, among others. Based on this observed heterogeneity, we explored the possible moderators that could explain this variability in reliability coefficients across studies.

The higher percentage of men in the evaluated studies was related to higher reliability coefficients in some subscales, such as EF, MH, and CF. Taking note of the initial development of the questionnaire, it is necessary to show that it was initially developed and configured employing the previous MOS but including additional scales relevant for PLWHIV, based on a literature review and discussions with clinicians and patients, mainly men (Wu et al., 1991). Hence, the initial validations of these instruments were developed using samples mainly represented by men, such as the preliminary validation, with samples of 97% and 98% of men (Wu et al., 1991) and indicated in the second study, which analyzed its psychometric properties, that this instrument was poorly employed in women (Wu et al., 1997), something that remains to the current time (Cooper et al., 2017). This fact probably had a significant influence on item content, representing more consequences of HIV on HRQoL from the men's perspective than in women, explaining the obtained results. Hence, more information is needed on the performance of the scale in women (Cooper et al., 2017).

Only in the case of MH, the mean CD4 count was revealed as a moderator. Although previous studies have found controversial results regarding the association between CD4 and HRQoL measured with the MOS-HIV (Call et al., 2000; Ruiz Perez et al., 2005; Weinfurt, Willke, Glick, Freimuth, & Schulman, 2000), this variable could explain the reliability variability in this subscale.

The version of the questionnaire was also revealed as a moderator in GHP, P, PF, EF, and CF. As in several previous RG meta-analyses, the original versions in English of the questionnaires exhibited higher reliability coefficients in comparison to the adaptations (Rubio-Aparicio, Núñez-Núñez, Sánchez-Meca, López-Pina, Marín-Martínez, &

López-López, 2020; Sánchez-Meca, Rubio-Aparicio, Núñez-Núñez, López-Pina, Marín-Martínez, & López-López, 2017; Vicent, Rubio-Aparicio, Sánchez-Meca, & González, 2019). In addition, in two of the subscales, PF and CF, the use of the questionnaire in North America exhibited higher reliability. As in the case of gender, given that this instrument was preliminarily validated in English and in the United States, these moderation results are not surprising. These results reveal the necessity of improving the cultural and language validations and adaptations of the questionnaires evaluating HRQoL in this population, as employing adaptations instead of the original version seems to decrease the reliability. The variables related to the age of the patients were also found to be relevant predictors as in previous similar studies like other RG meta-analyses (Vicent et al., 2019; Rubio-Aparicio, Badenes-Ribera, Sánchez-Meca, Fabris, & Longobardi, 2020).

The researchers also aimed to analyze reliability induction in studies that employ the MOS-HIV. It has been found that 76.1% of the analyzed articles induced reliability. This percentage is very high, as only a quarter of the studies employing this evaluation instrument reported the original reliability of their studies. As indicated previously, this issue could be a significant limitation of the studies, as diverse moderators could influence the reliability of this questionnaire. Although several initiatives have been developed to promote the report of original reliability coefficients, such as the American Psychological Association Task Force on Statistical Inference (Wilkinson, 1999), and the American Psychological Association for quantitative studies (Appelbaum et al., 2018), greater efforts are needed in this type of evaluation instrument of HRQoL, which has significant implications in clinical practice.

Relevance to Clinical Practice

Clinicians can use the information provided by this study to deepen their knowledge about the psychometric properties, mainly reliability, of one of the most employed evaluation instruments for the assessment of HRQoL in PLWHIV, in clinical settings. In the clinical practice of health services, nurses could employ this gold standard for reliable evaluations of HRQoL in PLWHIV. Hence, this scale can be used by health professionals to improve the standards of care provided to this patient group.

Although the present study entails an advance in the study of reliability indices of one of the most applied evaluation instruments to analyze the HRQoL in PLWHIV, some limitations should be addressed. First, the fact that only 10 studies reported reliability for the total score of MOS-HIV, and six for PHS and MHS limits the generalizability of the obtained results in this regard and restricts the identification of possible moderators that could

influence the reliability estimations. Second, multivariate meta-analysis could not be carried out due to an insufficient number of studies. Another limitation of the study could be based on the high induced reliability found in previously published studies. Lastly, no meta-analytical computations could be carried out for test-retest estimations because only two studies reported reliability in this regard.

Conclusion

The findings of this systematic review and RG meta-analysis demonstrate adequate reliability coefficients of the “gold standard” for the HRQoL evaluation in PLWHIV. Although the total score demonstrated the most adequate reliability coefficient, reaching .90, the subscales of the instrument could be also employed, but mainly in a research context, as all of them except RF exhibited an alpha coefficient above .80. Furthermore, the results indicated that reliability could depend on some moderators that could influence it, characteristics that should be considered by researchers and clinicians when applying the questionnaire. The MOS-HIV is a multidimensional instrument and its use in clinical and research contexts usually consists of the application of certain subscales, rather than in applying the global scale and MHS and PHS summaries. This fact can be shown in the large number of studies that applied the subscales in comparison to the few studies that applied the global scale and both summaries. The information extracted from the application of each subscale can be more enriching and detailed for the practitioners than those obtained through the global scale and PHS and MHS dimensions. Finally, the high rate of reliability induction for the MOS-HIV confirms the relevance of reporting reliability coefficients in future research that employs this questionnaire, avoiding the incorrect practice of inducing reliability through the lack of reporting alpha coefficients from the data being analyzed.

In summary, from a psychometric perspective, MOS-HIV could be an adequate instrument to reliably evaluate HRQoL in PLWHIV in research and medical settings, notably when the total score is employed.

Conflict of interest

The authors declare that they have no conflicts of interest.

Funding

This research was funded by the Office of the Vice President of Research and Knowledge Transfer of the University of Alicante (GRE-18-17B).

References

- Alamolhoda, M., Jafari, P., Ayatollahi, S. M. T., & Kazerooni, P. A. (2017). Reliability and validity of Persian version of Medical Outcome Study-HIV health survey in Iranian people living with HIV. *Journal of Traditional Chinese Medicine, 37*(5), 695-701. [https://doi.org/10.1016/S0254-6272\(17\)30325-4](https://doi.org/10.1016/S0254-6272(17)30325-4)
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3. <https://doi.org/10.1037/amp0000191>
- Badia, X., Podzamczar, D., Garcia, M., & Consiglio, E. (1999). A randomized study comparing instruments for measuring health-related quality of life in HIV-infected patients. *AIDS, 13*(13), 1727-1735. <https://doi.org/10.1097/00002030-199909100-00017>
- Badia, X.; Podzamczar, D., LópezLavid, C., & García, M. (1999). Medicina basada en la evidencia y la validación de cuestionarios de calidad de vida: La versión española del MOS-HIV para la evaluación de la calidad de vida en pacientes infectados por el VIH [Evidence-based medicine and validation of quality of life questionnaires: The Spanish version of MOS-HIV for the evaluation of quality of life in patients infected with HIV]. *Enfermedades Infecciosas y Microbiología Clínica, 17*, 103-113.
- Baigis, J., Korniewicz, D. M., Chase, G., Butz, A., Jacobson, D., & Wu, A. W. (2002). Effectiveness of a home-based exercise intervention for HIV-infected adults: A randomized trial. *The Journal of the Association of Nurses in AIDS Care: JANAC, 13*(2), 33-45. [https://doi.org/10.1016/S1055-3290\(06\)60199-4](https://doi.org/10.1016/S1055-3290(06)60199-4)
- Bekele, T., Rourke, S. B., Tucker, R., Greene, S., Sobota, M., Koornstra, J., ... Hwang, S. W. (2013). Direct and indirect effects of perceived social support on health-related quality of life in persons living with HIV/AIDS. *AIDS Care, 25*(3), 337-346. <https://doi.org/10.1080/09540121.2012.701716>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.

- Bucciardini, R., Fragola, V., Massella, M., Polizzi, C., Mirra, M., Goodall, R., ... Initio Trial International Coordinating Committee. (2007). Health-related quality of life outcomes in HIV-infected patients starting different combination regimens in a randomized multinational trial: The INTIO-QoL substudy. *AIDS Research and Human Retroviruses*, 23(10), 1215-1222. <https://doi.org/10.1089/aid.2007.0067>
- Bunjoungmanee, P., Chunloy, K., Tangsathapornpong, A., Khawcharoenporn, T., & Apisarnthanarak, A. (2014). Quality of life assessment among patients living with HIV/AIDS at a tertiary care hospital in Thailand. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 45(4), 834-842.
- Burgess, A., Dayer, M., Catalan, J., Hawkins, D., & Gazzard, B. (1993). The reliability and validity of two HIV-specific health-related Quality-of-Life measures: A preliminary analysis. *AIDS*, 7(7), 1001-1008. <https://doi.org/10.1097/00002030-199307000-00015>
- Burgoyne, R. W., & Saunders, D. S. (2001). Quality of life among urban Canadian HIV/AIDS clinic outpatients. *International Journal of STD & AIDS*, 12(8), 505-512. <https://doi.org/10.1258/0956462011923598>
- Burrage, J., & Demi, A. (2003). Buddy programs for people infected with HIV. *The Journal of the Association of Nurses in AIDS Care: JANAC*, 14(1), 52-62. <https://doi.org/10.1177/1055329002239190>
- Call, S. A., Klapow, J. C., Stewart, K. E., Westfall, A. O., Mallinger, A. P., DeMasi, R. A., ... Saag, M. S. (2000). Health-related quality of life and virologic outcomes in an HIV clinic. *Quality of Life Research*, 9(9), 977-985. <https://doi.org/10.1023/a:1016668802328>
- Cardona-Arias, J. A., & Higuera-Gutiérrez, L. F. (2014). Impact of HIV/AIDS on quality of life: Meta-analysis 2002-2012. *Revista Española De Salud Pública*, 88(1), 87-101. <https://doi.org/10.4321/S1135-57272014000100006>
- Castellvi, P., Ferrer, M., & Alonso, J. (2013). The patient-reported outcomes in research: Definition, impact, classification, measurement and assessment. *Medicina Clínica*, 141(8), 358-365. <https://doi.org/10.1016/j.medcli.2013.07.013>
- Castillo, M., & Luz, M. de la. (2006). Calidad de vida [Quality of life]. *Revista Mexicana de Medicina Física y Rehabilitación*, 18(1), 5-6.

- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability methods, and the clinical implications of low reliability. *The Journal of General Psychology*, *130*(3), 290–304. <https://doi.org/10.1080/00221300309601160>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Clayson, D. J., Wild, D. J., Quarterman, P., Duprat-Lomon, I., Kubin, M., & Coons, S. J. (2006). A comparative review of health-related quality-of-life measures for use in HIV/AIDS clinical trials. *Pharmacoeconomics*, *24*(8), 751-765. <https://doi.org/10.2165/00019053-200624080-00003>
- Cooper, V., Clatworthy, J., Harding, R., Whetham, J., & Emerge Consortium. (2017). Measuring quality of life among people living with HIV: A systematic review of reviews. *Health and Quality of Life Outcomes*, *15*(1), 220. <https://doi.org/10.1186/s12955-017-0778-6>
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston.
- Curtin, K. B., Cheng, Y. I., Wang, J., Scott, R. K., Squires, L., Benator, D. A., & Lyon, M. E. (2019). Quality of life of persons living with HIV and congruence with surrogate decision-makers. *Quality of Life Research*, *28*(1), 47-56 <https://doi.org/10.1007/s11136-018-2002-5>
- Drewes, J., Gusy, B., & Rüden, U. V. (2013). More than 20 years of research into the quality of life of people with HIV and AIDS — a descriptive review of study characteristics and methodological approaches of published empirical studies. *Journal of the International Association of Providers of AIDS Care (JIAPAC)*, *12*(1), 18-22. <https://doi.org/10.1177/1545109712456429>
- Dunne, E. M., Balletto, B. L., Donahue, M. L., Feulner, M. M., DeCosta, J., Cruess, D. G., ... Scott-Sheldon, L. A. (2019). The benefits of yoga for people living with HIV/AIDS: A systematic review and meta-analysis. *Complementary Therapies in Clinical Practice*, *34*, 157-164. <https://doi.org/10.1016/j.ctcp.2018.11.009>
- Duong, N., Torre, P., Springer, G., Cox, C., & Plankey, M. W. (2016). Hearing Loss and Quality of Life (QOL) among Human Immunodeficiency Virus (HIV)-Infected and Uninfected Adults. *Journal of AIDS & Clinical Research*, *7*(12), 645. <https://doi.org/10.4172/2155-6113.1000645>

- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>
- George, S., Bergin, C., Clarke, S., Courtney, G., & Codd, M. B. (2016). Health-related quality of life and associated factors in people with HIV: An Irish cohort study. *Health and Quality of Life Outcomes*, *14*, 115. <https://doi.org/10.1186/s12955-016-0517-4>
- Gielen, A. C., McDonnell, K. A., Wu, A. W., O'Campo, P., & Faden, R. (2001). Quality of life among women living with HIV: The importance violence, social support, and self-care behaviors. *Social Science & Medicine*, *52*(2), 315-322. [https://doi.org/10.1016/s0277-9536\(00\)00135-0](https://doi.org/10.1016/s0277-9536(00)00135-0)
- Golub, S. A., Tomassilli, J. C., & Parsons, J. T. (2009). Partner serostatus and disclosure stigma: Implications for physical and mental health outcomes among HIV-positive adults. *AIDS and Behavior*, *13*(6), 1233-1240. <https://doi.org/10.1007/s10461-008-9466-1>
- Hartung, J. (1999). An Alternative Method for Meta-Analysis. *Biometrical Journal*, *41*(8), 901-916. [https://doi.org/10.1002/\(SICI\)1521-4036\(199912\)41:8<901::AID-BIMJ901>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1521-4036(199912)41:8<901::AID-BIMJ901>3.0.CO;2-W)
- Henderson, W. A., Schlenk, E. A., Kim, K. H., Hadigan, C. M., Martino, A. C., Sereika, S. M., & Erlen, J. A. (2010). Validation of the MOS-HIV as a measure of health-related quality of life in persons living with HIV and liver disease. *AIDS Care*, *22*(4), 483-490. <https://doi.org/10.1080/09540120903207292>
- Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting «reliability generalization» studies. *Measurement and Evaluation in Counseling and Development*, *35*(2), 113-127. <https://doi.org/10.1080/07481756.2002.12069054>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557-560. <https://doi.org/10.1136/bmj.327.7414.557>
- Huang, Z.-J., Tian, M., Dai, S.-Y., & Ye, D.-Q. (2013). Feasibility, reliability and validity of the Chinese simplified version of the MOS-HIV health survey among AIDS patients in China. *Quality of Life Research*, *22*(2), 403-407. <https://doi.org/10.1007/s11136-012-0148-0>
- Hudson, A., Kirksey, K., & Holzemer, W. (2004). The Influence of Symptoms on Quality of Life among HIV-Infected Women. *Western Journal of Nursing Research*, *26*(1), 9-23. <https://doi.org/10.1177/0193945903259221>

- Kamen, C., Taniguchi, S., Student, A., Kienitz, E., Giles, K., Khan, C., ... Koopman, C. (2012). The impact of denial on health-related quality of life in patients with HIV. *Quality of Life Research*, 21(8), 1327-1336. <https://doi.org/10.1007/s11136-011-0045-y>
- Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17), 2693–2710. <https://doi.org/10.1002/sim.1482>
- Lipsey, M.W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (2nd ed., pp. 147-158). Russell Sage Foundation.
- Liu, J., Zhu, Y., & Qu, B. (2018). Reliability and validity of the Chinese version of the Medical Outcomes Study HIV Health Survey (MOS-HIV) in people living with HIV/AIDS (PLWHA) in China. *PLoS One*, 13(7), e0201177. <https://doi.org/10.1371/journal.pone.0201177>
- López-López, J. A., Botella, J., Sánchez-Meca, J., & Marín-Martínez, F. (2013). Alternatives for mixed-effects meta-regression models in the reliability generalization approach: A simulation study. *Journal of Educational and Behavioral Statistics*, 38(5), 443-469. <https://doi.org/10.3102/1076998612466142>
- López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Noortgate, W. V. den., & Viechtbauer, W. (2014). Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology*, 67(1), 30-48. <https://doi.org/10.1111/bmsp.12002>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Pub. Co.
- Maluccio, J. A., Palermo, T., Kadiyala, S., & Rawat, R. (2015). Improving Health-Related Quality of Life among People Living with HIV: Results from an Impact Evaluation of a Food Assistance Program in Uganda. *PLOS ONE*, 10(8), e0135879. <https://doi.org/10.1371/journal.pone.0135879>
- McDonald, R. P. (1999). *Test theory: a unified treatment*. Lawrence Earlbaum Associates. Inc.
- Melhuish, A., & Lewthwaite, P. (2018). Natural history of HIV and AIDS. *Medicine*, 46(6), 356-361. <https://doi.org/10.1016/j.mpmed.2018.03.010>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

- Murillo-Bonilla, L. M., Godínez, J. M. G., Cruz, M. A. S., & González, R. P. U. (2018). Frecuencia del Virus de la Inmunodeficiencia Humana (VIH) a Nivel Mundial [Worldwide Frequency of the Human Immunodeficiency Virus (HIV)] *Revista de Medicina Clínica*, 2(1), 35-37. <https://doi.org/10.5281/zenodo.1186932>
- Mutabazi-Mwesigire, D., Katamba, A., Martin, F., Seeley, J., & Wu, A. W. (2015). Factors That Affect Quality of Life among People Living with HIV Attending an Urban Clinic in Uganda: A Cohort Study. *PLoS One*, 10(6), e0126810. <https://doi.org/10.1371/journal.pone.0126810>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.
- O'Brien, K. K., Bayoumi, A. M., Strike, C., Young, N. L., & Davis, A. M. (2008). Exploring disability from the perspective of adults living with HIV/AIDS: Development of a conceptual framework. *Health and Quality of Life Outcomes*, 6(1), 76. <https://doi.org/10.1186/1477-7525-6-76>
- Portillo, C. J., Holzemer, W. L., & Chou, F.-Y. (2007). HIV Symptoms. *Annual Review of Nursing Research*, 25(1), 259-291. <https://doi.org/10.1891/0739-6686.25.1.259>
- Revicki, D. A., Sorensen, S., & Wu, A. W. (1998). Reliability and validity of physical and mental health summary scores from the Medical Outcomes Study HIV Health Survey. *Medical Care*, 36(2), 126-137. <https://doi.org/10.1097/00005650-199802000-00003>
- Revicki, D. A., Swartz, C., Wu, A. W., Haubrich, R., & Collier, A. C. (1999). Quality of life outcomes of saquinavir, zalcitabine and combination saquinavir plus zalcitabine therapy for adults with advanced HIV infection with CD4 counts between 50 and 300 cells/mm³. *Antiviral Therapy*, 4(1), 35-44.
- Revicki, Dennis A., Wu, A. W., & Murray, M. I. (1995). Change in Clinical Status, Health Status, and Health Utility Outcomes in HIV-Infected Patients. *Medical Care*, 33(4), 173-182.
- Rubio-Aparicio, M., Badenes-Ribera, L., Sánchez-Meca, J., Fabris, M. A., & Longobardi, C. (2020). A reliability generalization meta-analysis of self-report measures of muscle dysmorphia. *Clinical Psychology: Science and Practice*, 27(1), 1-24. <https://doi.org/10.1111/cpsp.12303>
- Rubio-Aparicio, M., Núñez-Núñez, R. M., Sánchez-Meca, J., López-Pina, J. A., Marín-Martínez, F., & López-López, J. A. (2020). The Padua Inventory–Washington State University Revision of obsessions and compulsions: A reliability generalization meta-analysis. *Journal of Personality Assessment*, 102(1), 113-123. <https://doi.org/10.1080/00223891.2018.1483378>

- Rubio-Aparicio, M., Sánchez-Meca, J., López-López, J. A., Botella, J., & Marín-Martínez, F. (2017). Analysis of categorical moderators in mixed-effects meta-analysis: Consequences of using pooled versus separate estimates of the residual between-studies variances. *British Journal of Mathematical and Statistical Psychology*, *70*(3), 439-456. doi:10.1111/bmsp.12092
- Ruiz Perez, I., Rodriguez Baño, J., Lopez Ruz, M. A., del Arco Jimenez, A., Causse Prados, M., Pasquau Liaño, J., ... Marcos, M. (2005). Health-related quality of life of patients with HIV: Impact of sociodemographic, clinical and psychosocial factors. *Quality of Life Research*, *14*(5), 1301-1310. <https://doi.org/10.1007/s11136-004-4715-x>
- Ryu, E., West, S. G., & Sousa, K. H. (2009). Mediation and moderation: Testing relationships between symptom status, functional health, and quality of life in HIV patients. *Multivariate Behavioral Research*, *44*(2), 213-232. <https://doi.org/10.1080/00273170902794222>
- Safrin, S., Finkelstein, D. M., Feinberg, J., Frame, P., Simpson, G., Wu, A., ... Black, J. R. (1996). Comparison of three regimens for treatment of mild to moderate *Pneumocystis carinii* pneumonia in patients with AIDS. A double-blind, randomized, trial of oral trimethoprim-sulfamethoxazole, dapsone-trimethoprim, and clindamycin-primaquine. ACTG 108 Study Group. *Annals of Internal Medicine*, *124*(9), 792-802. <https://doi.org/10.7326/0003-4819-124-9-199605010-00003>
- Sánchez, M. D. G., & Guiteras, A. F. (2004). Evaluación de la calidad de vida en personas VIH positivas con lipodistrofia [Evaluation of quality of life in people HIV positive with lipodystrophy]. *Psicología y Salud*, *14*(1), 5-11.
- Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *The British Journal of Mathematical and Statistical Psychology*, *66*(3), 402-425. <https://doi.org/10.1111/j.2044-8317.2012.02057.x>
- Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*(1), 31-48. <https://doi.org/10.1037/1082-989X.13.1.31>
- Sánchez-Meca, J., Rubio-Aparicio, M., Núñez-Núñez, R. M., López-Pina, J., Marín-Martínez, F., & López-López, J. A. (2017). A reliability generalization meta-analysis of the Padua Inventory of obsessions and compulsions. *The Spanish Journal of Psychology*, *20*, 1-15. doi:10.1017/sjp.2017.65

- Schwartzmann, L. (2003). Calidad de vida relacionada con la salud: aspectos conceptuales [Health-related quality of life: conceptual aspects]. *Ciencia y Enfermería*, 9(2), 09-21. <https://doi.org/10.4067/S0717-95532003000200002>
- Scott-Lennox, J. A., Wu, A. W., Boyer, J. G., & Ware, J. E. (1999). Reliability and validity of French, German, Italian, Dutch, and UK English translations of the Medical Outcomes Study HIV Health Survey. *Medical Care*, 37(9), 908-925. <https://doi.org/10.1097/00005650-199909000-00007>
- Shields, A. L., & Caruso, J. C. (2004). A Reliability Induction and Reliability Generalization Study of the Cage Questionnaire. *Educational and Psychological Measurement*, 64(2), 254-270. <https://doi.org/10.1177/0013164403261814>
- Shim, E.-J., Ha, H., Lee, S. H., Kim, N. J., Kim, E. S., Bang, J. H., ... Hahm, B.-J. (2018). Psychometric properties of the Korean version of the medical outcomes study HIV health survey: Results from a multicenter survey in Korea. *Health and Quality of Life Outcomes*, 16(1), 92. <https://doi.org/10.1186/s12955-018-0919-6>
- Simpson, K. N., Hanson, K. A., Harding, G., Haider, S., Tawadrous, M., Khachatryan, A., ... Wu, A. W. (2013). Patient reported outcome instruments used in clinical trials of HIV-infected adults on NNRTI-based therapy: a 10-year review. *Health and Quality of Life Outcomes*, 11(1), 164. <https://doi.org/10.1186/1477-7525-11-164>.
- Stasinopoulou, P. G., Tzavara, C., Dimitrakaki, C., Georgiou, O., Baraboutis, I. G., Skoutelis, A., ... Tountas, Y. (2010). Reliability and validity of the Greek translation of the MOS-HIV health survey in HIV-infected individuals. *Quality of Life Research*, 19(2), 199-205. <https://doi.org/10.1007/s11136-009-9573-0>
- Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS short-form general health survey. Reliability and validity in a patient population. *Medical Care*, 26(7), 724-735. <https://doi.org/10.1097/00005650-198807000-00007>
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press.
- Tozzi, V., Balestra, P., Galgani, S., Murri, R., Bellagamba, R., Narciso, P., ... Wu, A. W. (2003). Neurocognitive performance and quality of life in patients with HIV infection. *AIDS Research and Human Retroviruses*, 19(8), 643-652. <https://doi.org/10.1089/088922203322280856>

- Vacha-Haase, T. (1998). Reliability Generalization: Exploring Variance in Measurement Error Affecting Score Reliability Across Studies. *Educational and Psychological Measurement*, 58(1), 6-20. <https://doi.org/10.1177/0013164498058001002>
- Vicent, M., Rubio-Aparicio, M., Sánchez-Meca, J., & González, C. (2019). A reliability generalization meta-analysis of the child and adolescent perfectionism scale. *Journal of Affective Disorders*, 245, 533-544. <https://doi.org/10.1016/j.jad.2018.11.049>
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36(1), 1-48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods*, 20(3), 360-374. <https://doi.org/10.1037/met0000023>
- Weinfurt, K. P., Willke, R. J., Glick, H. A., Freimuth, W. W., & Schulman, K. A. (2000). Relationship between CD4 count, viral burden, and quality of life over time in HIV-1-infected patients. *Medical Care*, 38(4), 404-410. <https://doi.org/10.1097/00005650-200004000-00007>
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604. <https://doi.org/10.1037/0003-066X.54.8.594>
- World Health Organization, HIV/AIDS. (n. d.). Retrieved August 27, 2019, from <https://apps.who.int/gho/data/view.main.22100WHO?lang=en>
- Wu, A. W., Hanson, K. A., Harding, G., Haider, S., Tawadrous, M., Khachatryan, A., ... Simpson, K. N. (2013). Responsiveness of the MOS-HIV and EQ-5D in HIV-infected adults receiving antiretroviral therapies. *Health and Quality of Life Outcomes*, 11(1), 42. <https://doi.org/10.1186/1477-7525-11-42>
- Wu, A. W., Revicki, D. A., Jacobson, D., & Malitz, F. E. (1997). Evidence for reliability, validity and usefulness of the Medical Outcomes Study HIV Health Survey (MOS-HIV). *Quality of Life Research*, 6(6), 481-493.
- Wu, A. W., Rubin, H. R., Mathews, W. C., Ware, J. E., Brysk, L. T., Hardy, W. D., ... Richman, D. D. (1991). A health status questionnaire using 30 items from the Medical Outcomes Study. Preliminary validation in persons with early HIV infection. *Medical Care*, 29(8), 786-798. <https://doi.org/10.1097/00005650-199108000-00011>

Yu, N. X. (2017). The effects of beliefs about AIDS-related death on quality of life in Chinese married couples with both husband and wife infected with HIV: Examining congruence using the actor-partner interdependence model. *Health and Quality of Life Outcomes*, 15(1), 127. <https://doi.org/10.1186/s12955-017-0703-z>

Zhu, Z., Zhao, R., & Hu, Y. (2019). Symptom Clusters in People Living With HIV: A Systematic Review. *Journal of Pain and Symptom Management*, 58(1), 115-133. <https://doi.org/10.1016/j.jpainsymman.2019.03.018>

Figure 1. Flow diagram of study selecting process