Commentary on using the SF-36 or MOS-HIV in studies of persons with HIV disease
Jim Shahriar¹, Thomas Delate², Ron D Hays³ and Stephen Joel Coons*⁴

Address: ¹Quality Programs, Health Care Services, Blue Shield of California, San Francisco, CA 94105, USA, ²Express Scripts, Inc., Office of Research and Planning, Maryland Heights, MO 63043, USA, ³Division of General Internal Medicine and Health Services Research, University of California at Los Angeles, Los Angeles, CA 90095-1736, USA and ⁴Division of HIV Policy and Outcomes Research, College of Pharmacy, University of Arizona, Tucson, AZ 85721-0207, USA

Email: Jim Shahriar - jimsheed@aol.com; Thomas Delate - tdelate@express-scripts.com; Ron D Hays - hays@rand.org; Stephen Joel Coons* - coons@pharmacy.arizona.edu

* Corresponding author

Abstract
The purpose was to compare and comment on use of the SF-36 and MOS-HIV instruments in studies of persons with HIV disease. Three medical information databases were searched to identify examples of HIV studies that included the MOS-HIV or SF-36. Thirty-nine and 14 published articles were identified for illustration in comparing the use of the MOS-HIV and SF-36 in HIV disease, respectively. Support for the reliability and construct validity of the MOS-HIV and SF-36 was found. Ceiling and floor effects were reported for both the MOS-HIV and SF-36; however, ceiling effects were more common for the MOS-HIV, in part due to fewer items in the physical, social, and role functioning domains. The MOS-HIV measures three domains hypothesized to be associated with the health deterioration of HIV disease not measured by the SF-36; however, these domains may not assess aspects of HIV disease that typify the majority of the persons with HIV disease today. National norms for the U.S. adult population (and other nations) are available for the SF-36. In addition, the SF-36 has been used in a wide variety of patient populations, enabling comparisons of HIV-infected persons with persons with other health conditions. No national norms for the MOS-HIV are available. We conclude that there is currently insufficient evidence in the literature to recommend the use of the MOS-HIV over the SF-36 in HIV-infected persons. Although the SF-36 is not targeted at HIV, it may be preferable to use the SF-36 over the MOS-HIV due to fewer ceiling effects, availability of national norms, and the vast amount of data for other populations in the U.S. and around the world. Head-to-head comparisons demonstrating the unique value of the MOS-HIV over the SF-36 are clearly needed. More importantly, additional work needs to be directed at comparing the MOS-HIV and other putatively HIV-targeted instruments to one another to help demarcate aspects of HRQOL that are truly generic versus specific to HIV disease. Using both a generic and targeted HRQOL measure is a good general strategy, but this has not been a typical practice in studies of HIV because the MOS-HIV is so similar in content to the SF-36.

Introduction
Human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS) entered the public consciousness over two decades ago. In the ensuing years,
interest in the measurement of self-reported functioning and well-being, or health-related quality of life (HRQOL), in HIV-infected individuals has been extensive [1–4]. While the treatment for HIV-infection remains non-curative, the improvements in mortality and AIDS-free survival for HIV-infected individuals have been substantial [5,6]. The modification of the natural history of the disease with multi-pharmaceutical regimens that have diverse beneficial as well as toxic effects [7] makes the measurement of HRQOL in this patient population more important than ever. However, there is no consensus regarding the best measurement approach.

When a particular disease is being considered, there is a tendency to assume that disease-targeted measures are superior to general or generic measures. HIV/AIDS is no exception. Numerous HIV/AIDS-targeted HRQOL measures have been developed. A recent review [8] evaluated the psychometric properties of HIV disease-targeted HRQOL instruments. Based on their review, the authors could not recommend the use of any of the instruments reviewed; however, the MOS-HIV was found to have the most available evidence for the evaluation criteria applied. Regardless of the lack of compelling arguments for its use, the MOS-HIV appears to be the most popular HRQOL instrument currently reported in the HIV literature. The purpose of this investigation is to examine the use of the MOS-HIV, a measure targeted at HIV disease, with the leading generic HRQOL instrument, the SF-36, in studies of persons with HIV.

The Medical Outcomes Study
The Medical Outcomes Study (MOS) was a four-year observational study that investigated the changes in physician practice styles and patient outcomes under different healthcare settings such as health maintenance organizations, large physician groups, or individual physician fee-for-service practices [9]. One of the goals of this longitudinal study was to construct reliable and valid tools for measuring and monitoring patient-reported functioning and well-being [10]. To complement the conventional clinical outcomes in the study, a spectrum of patient-reported outcome measures was created [11].

The MOS Short Form 20-Item Health Survey (SF-20), a brief, generic health status instrument that provides six scale scores (general health perceptions, physical function, role function, social function, pain, and mental health), was the first short form developed from the MOS and was used for the screening of patients for chronic disease status during the cross sectional phase of the study [11]. The SF-36, a second generation of the short form, includes an additional health concept (energy/fatigue), increases the precision of previous single-item measures (pain, social functioning) and multi-item measures (physical functioning) by adding additional items, measures the extent of physical limitations rather than the duration of the limitation, and focuses on a wider array of role limitations. The SF-36 was developed for the longitudinal phase of the MOS. At least nine short form instruments developed from the MOS scales have been used in studies of HIV-infected persons [11]. Among those instruments, the MOS-HIV has been reported to be the most widely used by researchers in patients with HIV infection [2,12,13].

Background to MOS-HIV and SF-36 Instruments
Owing to the perceived need for a succinct instrument to evaluate HRQOL in HIV-infected patients in multi-center AIDS clinical trials, the development of the MOS-HIV was begun in 1987 [14]. Sixteen items selected from the six scales of the SF-20 were the foundation for the construction of MOS-HIV. Four additional scales that were hypothesized to be related to the health status of HIV-infected persons (i.e., cognitive functioning [4 items], energy/fatigue [4 items], health distress [4 items] and quality of life [1 item]) as well as a single item assessing health transition were added to the original scales in the SF-20, resulting in a 30-item questionnaire [14]. The original MOS-HIV included only one general health perception item. Subsequently, the 4 other SF-20 current health items were added, leading to the 34-item MOS-HIV. With the addition of a second pain item, the current version of the MOS-HIV (distributed by the Medical Outcomes Trust) contains 35 items. It covers 11 dimensions of health including physical functioning, role functioning, pain, social functioning, emotional well-being, energy/fatigue, cognitive functioning, general health, health distress, overall QOL, and health transition. Mental (MHS) and physical health summary (PHS) scores can be calculated from the MOS-HIV scales [15]. The MOS-HIV scales range from 0 to 100, with higher scores representing better functioning and well-being. The MHS and PHS are scored using a method that transforms the scores to a standardized scale (T score) with a mean of 50 and a standard deviation of 10 in the sample in which the summary scores were developed [15]. Mean PHS and MHS scores above or below 50 can be interpreted as having better or worse HRQOL than the HIV-infected patient sample from which the summary measures were developed.

The SF-36 is one of the most widely used HRQOL instruments [16] and has demonstrated high levels of reliability and validity in diverse patient populations [17,18]. It has 36 items that measure eight multi-item health concepts (i.e., physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health). This instrument was developed to address the health-related concepts that are most influenced by disease states and their related treatments [11].
The SF-36 can be scored to yield two orthogonal factor-based component summary scores for mental and physical health. The mental (MCS-36) and physical component summary (PCS-36) scores were derived from the eight scales of the SF-36 using principal components analysis of the total patient sample from the MOS and a sample of the general US population [19]. These physical and mental components account for 82% of the reliable variance in the SF-36’s scales in the general US population. The SF-36 scales are scored on a 0 to 100 possible range, with higher scores representing better functioning and well-being. The MCS-36 and PCS-36 are scored using a method that transforms the scores to a standardized scale (T-scores) with a norm of 50 and a standard deviation of 10 in the general US population. Sample mean MCS-36 and PCS-36 scores above or below 50 can be interpreted as having better or worse HRQOL than the general US population. With norms established in subgroups based on gender and age and thirty medical conditions, including “healthy” with no chronic conditions, this standardized scoring provides a means of comparing results across patients with diverse medical conditions [20].

Table 1: Number of items in the scales of the MOS-HIV and SF-36.

<table>
<thead>
<tr>
<th>Scale</th>
<th>MOS-HIV</th>
<th>SF-36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Role functioning</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Pain</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Social functioning</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Energy / fatigue</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>General health</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Health distress</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Overall QoL</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Health transition</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The SF-36 can be scored to yield two orthogonal factor-based component summary scores for mental and physical health. The mental (MCS-36) and physical component summary (PCS-36) scores were derived from the eight scales of the SF-36 using principal components analysis of the total patient sample from the MOS and a sample of the general US population [19]. These physical and mental components account for 82% of the reliable variance in the SF-36’s scales in the general US population. The SF-36 scales are scored on a 0 to 100 possible range, with higher scores representing better functioning and well-being. The MCS-36 and PCS-36 are scored using a method that transforms the scores to a standardized scale (T-scores) with a norm of 50 and a standard deviation of 10 in the general US population. Sample mean MCS-36 and PCS-36 scores above or below 50 can be interpreted as having better or worse HRQOL than the general US population. With norms established in subgroups based on gender and age and thirty medical conditions, including “healthy” with no chronic conditions, this standardized scoring provides a means of comparing results across patients with diverse medical conditions [20].

Examples of Studies of Persons with HIV using MOS-HIV or SF-36

A series of literature searches was performed to identify studies that measured health status in persons with HIV with the SF-36 or MOS-HIV. The literature in three databases, Medline, HealthStar, and PsychInfo, was searched from 1975 through 2002. The terms “quality of life,” “HIV,” and the name of the MOS instrument (i.e., SF-36 or MOS-HIV) were cross-referenced in each search of the databases. Only articles written in English were included.

Discussion

Although other MOS-derived instruments have been used, studies incorporating either the MOS-HIV or SF-36 were found to be the most prevalent in studies of HIV disease. Thirty-nine citations for the MOS-HIV were found which presented empirical data [12,15,21–57]. Fourteen empirical articles were found for the SF-36 [20,58–70]. At the time of this review, more than 40 cross-cultural translations were available for the SF-36 and 14 translations for the MOS-HIV [14,71]. Only the SF-36 and the SF-12 (a subset of items from the SF-36 that reproduces > 90% of the variance of the SF-36’s summary scores) [19] have norms available for their summary/composite measures that have been calculated from nationally representative samples.

Table 1 compares the numbers of items for each of the scales of the MOS-HIV and SF-36. The SF-36 has four additional items for the measurement of the physical functioning domain (10 vs. 6), five more items for the role functioning domains (7 vs. 2), and one more item for the social functioning domain (2 vs. 1) than the MOS-HIV. However, the MOS-HIV measures three domains (two with multi-item scales [cognitive functioning and health distress] and one with a single item [overall quality of life]) that are not measured by the SF-36.

The two instruments take about five to ten minutes to be self-administered and can be interviewer administered in person or by telephone [48]. Reliability has been supported for the MOS-HIV [14,15]and the SF-36 [20,59,72], in this patient population. Support for item discrimination of the MOS-HIV has been shown in comparison with other HIV-targeted and generic HRQOL instruments [14]. One study found no differences in health distress and quality of life scale scores of the MOS-HIV in patients with early vs. late stages of HIV disease [12]. Differences have been found between HIV-infected and non-infected persons on all scales of the SF-36 [66] while only the summary scores of MOS-HIV distinguished between these two groups [22,23]. The physical health summary scores from the SF-12 and the MOS-HIV both have been found to dis-
The summary scores of the SF-36 have been shown to be responsive to HIV disease progression [20,74]. The MOS-HIV has been useful in illustrating changes in clinical status between treatment and control groups in intervention and observational studies; however, the role functioning, pain, mental health, health distress, and quality of life scales were shown to be non-responsive to treatment in one anti-retroviral intervention study [29]. Ceiling effects have been found for the physical functioning, role functioning and role emotional scales of the SF-36 [60] and MOS-HIV [37,55]. Ceiling effects have been found in the cognitive functioning, pain, and health transition scales of the MOS-HIV in HIV-infected patients with more advanced HIV disease [35]. In addition, the role functioning scale of the MOS-HIV has been found to have floor effects in HIV-infected patients with more advanced HIV disease [21,22,35,37].

Conclusions
One of the primary distinctions between the MOS-HIV and the SF-36 is the availability of nationally representative norms for the SF-36. Mental and physical health summary scores can be calculated for the MOS-HIV, but norms for these scores are available from only the subjects in the studies from which they were developed. The availability of nationally representative normative data permits the comparison of summary scores from one individual or a group of study subjects with scores from a sample representative of the general population. In addition, nationally representative norms for the MCS-36 and PCS-36 summary scores are available for males and females in seven age groups and for fourteen chronic conditions [75]. These norms permit healthcare decision makers to utilize the summary scores from the SF-36 to compare the health status of HIV-infected persons with other persons of similar gender and age, or with another chronic condition.

Perceived health of asymptomatic HIV-infected individuals does not appear to vary much from non-infected subjects. Wu and colleagues [76] demonstrated that HRQOL in HIV-infected patients with no symptoms or significant abnormalities was not different from that of healthy non-infected individuals. In a more recent study, Hays and colleagues found that the physical functioning of asymptomatic HIV-infected subjects was similar to that of the general US population [74]. In addition to increased measurement precision, using the SF-36 in patients in the early stages of HIV disease would allow a researcher to compare and contrast the health status of these HIV-infected individuals with the health status of patients with diverse chronic conditions in a range of cultures.

While the MOS-HIV has demonstrated evidence of reliability, construct validity and responsiveness among HIV-infected patient subjects, [12,15,42,43,77] there are limitations to the instrument. To allow for the addition of "disease-targeted" scales while not increasing respondent burden, the developers of the MOS-HIV sacrificed some measurement precision for the physical functioning, social functioning, and role functioning scales by reducing the number of items [21,24,35,37]. These scales, in particular, appear to be important in assessing HRQOL in this patient population since the majority of the persons infected with HIV are young adults who are still functioning normally in society and may not perceive themselves as having functional impairment [13,37]. For example, the full-length 10-item physical functioning scale of the SF-36 allows for the sampling of a wider range of severe and minor physical limitations and may provide for a better representation of the levels and types of physical limitations in this population [18]. The decreased measurement precision of the functioning scales in the MOS-HIV may explain why ceiling and floor effects have been found in a number of studies. Ceiling and floor effects become problematic when patients in longitudinal studies score the lowest or highest possible score at the baseline since subjects then can not report any lower or higher score that may occur if their health status deteriorates or improves at follow-up [78].

The MOS-HIV was developed over a decade ago when HIV disease and its treatment were very different from today [79]. At that time, rapid health deterioration from HIV disease and its associated sequelae was prevalent. With rapid progression into the late stages of HIV disease, monitoring the HRQOL of patients who were developing HIV-infection associated sequelae such as cancers (e.g., Kaposi’s sarcoma), opportunistic infections (e.g., pneumocystis carinii pneumonia), and AIDS dementia was imperative. The disease-targeted scales (i.e., cognitive functioning, health distress, and quality of life) added to the 16 items from the SF-20 to develop the MOS-HIV were included in the instrument to measure domains hypothesized to be associated with the HIV disease-related health deterioration [12]. Currently, however, with the availability of HAART, HIV disease is, for the majority of patients, a chronic rather than acute condition [79] and, as such, may require an instrument to assess HRQOL that was designed for chronic rather than acute diseases. Indeed, empirical evidence supporting the construct validity of the disease-targeted scales of the MOS-HIV is limited [12].

The current chronic nature of the disease may preclude the necessity to monitor the HIV disease-targeted domains of the MOS-HIV in the general population of HIV-infected patients. Nevertheless, if monitoring of HIV disease-targeted health concepts is important, the MCS-36 summary
score of the SF-36 is strongly correlated with the cognitive functioning ($r = 0.70$) and quality of life ($r = 0.68$) scales and negatively correlated with the health distress ($r = -0.57$) scale of the MOS-HIV [75]. This indicates that the MCS-36 represents these health concepts to some degree. Furthermore, in populations of patients in the late stages of HIV disease, augmenting the SF-36 with additional scales or instruments targeting the specific consequences of the disease and its treatment may be an effective approach. Two abstracts presented at the 1996 International Conference on AIDS found that the addition of HIV disease-targeted scales to the SF-36 did not detract from the measurement precision of the SF-36-specific scales or increase response burden appreciably and did provide for the effective monitoring of HIV disease [80,81].

Our results were based on studies identified through several databases. There are other databases that may have contained studies that were not included in our evaluation. In addition, there were only 14 studies identified that utilized the SF-36 in HIV-infected persons. It is possible that additional studies are needed to more fully document the strengths and shortcomings of this instrument. In sum, this study revealed that although the MOS-HIV has been used widely in the monitoring of HIV-infected persons, it has noteworthy limitations that may constrain its applications in this population. Hence, there is insufficient evidence in the literature to support the use of the MOS-HIV rather than the SF-36 in HIV-infected persons. Although the SF-36 is not targeted at HIV, it may be preferable to use the SF-36 over the MOS-HIV due to fewer ceiling effects, availability of national norms, and the vast amount of data for other populations in the U.S. and around the world. Head-to-head comparisons demonstrating the unique value of the MOS-HIV over the SF-36 are clearly needed. In addition, more work needs to be directed at comparing the MOS-HIV and other putatively HIV-targeted instruments to one another to help demarcate aspects of HRQOL that are truly generic versus specific to HIV disease.

**Authors’ contributions**

JS reviewed the literature and drafted the manuscript. TD helped with the literature review and drafting of the manuscript. RDH provided feedback about the initial idea and editorial input to the drafts of the manuscript. SJC oversaw the effort and was involved in drafting and revising the manuscript from start to finish.

**Acknowledgements**

Ron D. Hays, Ph.D., was supported in part by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities, (P20-MD00148-01) and the UCLA Center for Health Improvement in Minority Elders / Resource Centers for Minority Aging Research, National Institutes of Health, National Institute of Aging, (AG-02-004).

**References**


34. Delate T and Coons SJ: The discriminative ability of the 12-item short form health survey (SF-12) in a sample of persons infected with HIV Clin Ther 2000, 22(9):1112-120.


71. IQOLA Project SF-36 and SF-12 Translations [http://www.qualitymetric.com/cgi-bin/license/lang/matrix.cgi?i=0]


