Supplemental Information Requested by CMS for Patient Satisfaction, Quality of Life & Functional Status, and Behavioral Health & Substance Abuse Care Assessment Tools

Tool	Tool Information
Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) **IT-6.1.a**	**Summary** Purpose: The Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) provides a standardized survey instrument and data collection methodology for measuring patients' perspectives on hospital care. First, the survey is designed to produce comparable data on the patient's perspective on care that allows objective and meaningful comparisons between hospitals on domains that are important to consumers. Second, public reporting of the survey results is designed to create incentives for hospitals to improve their quality of care. Third, public reporting will serve to enhance public accountability in health care by increasing the transparency of the quality of hospital care provided in return for the public investment. Overview: The HCAHPS Survey is the first national, standardized, publicly reported survey of patients' perspectives of hospital care. HCAHPS, also known as the CAHPS® Hospital Survey*, is a 27-item survey instrument and data collection methodology for measuring patients' perceptions of their hospital experience. Format: The HCAHPS survey contains 18 patient perspectives on care and patient rating items that encompass eight key topics: communication with doctors, communication with nurses, responsiveness of hospital staff, pain management, communication about medicines, discharge information, cleanliness of the hospital environment, and quietness of the hospital environment. The survey also includes four screener questions and five demographic items, which are used for adjusting the mix of patients across hospitals and for analytical purposes. The survey is 27 questions in length. Scores: The Patient Experience of Care Domain score is comprised of two parts: the HCAHPS Base Score (maximum of 80 points) and the HCAHPS Consistency Points score (maximum of 20 points). Each of the eight HCAHPS dimensions contributes to the HCAHPS Base Score through either an Improvement or Achievement score. "Improvement" is the amount of change in an HCAHPS dimension from the earlier Baseline Period to the later Performance Period. "Achievement" is the comparison of each dimension in the Performance Period to the national median for that dimension during the Baseline Period. The larger of the Improvement or Achievement score for each dimension is used to calculate a hospital's HCAHPS Base Score. The second part of the Patient Experience of Care Domain is the Consistency Points score, which ranges from 0 to 20 points. Consistency Points are designed to target and further incentivize improvement in a hospital's *lowest performing* HCAHPS dimension. The Patient Experience of Care Domain Score is the sum of the HCAHPS Base Score (0 – 80 points) and HCAHPS Consistency Points score (0 – 20 points), thus ranges from 0 to 100 points, and comprises 30% of the Hospital VBP Total Performance Score. Administration Time: On average, it takes respondents about seven minutes to complete the HCAHPS survey items. [2i]

Tool	Tool Information
	Clinical/Quality information Reliability: The seven composites had a median internal consistency reliability of 0.69 and a median hospital-level reliability of 0.74 in the pilot study.[1]
	Availability of Benchmark Data The following summarizes the July 2013[2] 95th percentiles for the HCAHPS Top (near best) and Bottom Box (near worst) Scores (reported as [Bottom Box Score, Top Box Score]). Communication with Nurses: [10%, 87%]; Communication with Doctors: [8%, 90%]; Responsiveness of Hospital Staff: [18%, 83%]; Pain Management: [13%, 80%]; Communication about Medicine: [27%, 75%]; Cleanliness of Hospital Env.: [16%, 86%]; Quietness of Hospital Env.: [20%, 78%]; Discharge Information: [24%, 91%]; Overall Hospital Rating: [16%, 84%]; Recommend the Hospital: [11%, 86%]
	Additional Notes/Links Link to survey: http://www.hcahpsonline.org/home.aspx Distributor: http://www.hcahpsonline.org/app_vendor.aspx Versions/Languages: The HCAHPS Survey is available in English, Spanish, Chinese, Russian and Vietnamese in the mail format and in English and Spanish in the telephone and Interactive Voice Response formats. [2i] Original Publication Date: Voluntary collection of HCAHPS data for public reporting began in 2006, and public reporting of HCAHPS scores began in 2008. Since July 2007, hospitals subject to IPPS payment provisions ("subsection (d) hospitals") must collect, submit and publicly report HCAHPS data in order to receive their full IPPS annual payment update (APU). [2i] Copyright: Centers for Medicare & Medicaid Services, Baltimore, MD[2i] Web Access: http://www.hcahpsonline.org/home.aspx Contact/Availability: hcahps@azqio.sdps.org or 1-888-884-4007
CG-CAHPS 12-month Survey (Domains and Supplemental Items) – **IT-6.1.b.i – IT-6.1.b.iii.4**	**Summary**[3] Purpose: The Clinician and Group Consumer Assessment of Healthcare Providers and Systems (CG-CAHPS) survey is a standardized tool to measure patient perceptions of care by physicians in an office setting. The 12-month survey asks respondents about experiences during visits with their provider in the last 12 months. Overview: In addition to the 12-month CG-CAHPS survey, providers can specifically measure domains of patient satisfaction, or supplemental items. Domains: The following domains can be measured specifically using the CG-CAHPS survey: 1. Are getting timely care, appointments, and information

[1] http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/downloads/Hospital3State_Pilot_Analysis_Final200512.pdf
[2] http://www.hcahpsonline.org/files/Report_HEI_July_2013_Pctls.pdf
[2i] http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/downloads/HospitalHCAHPSFactSheet201007.pdf
[3] http://cahps.ahrq.gov/clinician_group/cgsurvey/fieldingcahps-cgsurveys.pdf

Tool	Tool Information
	2. How well their doctors communicate 3. Patient's rating of doctor access to specialist 4. Patient's involvement in shared decision making 5. Patient's overall health status/functional status Supplemental Items: Additional questions can be added to the CG-CAHPS survey in order to measure: 1. Cultural Competence 2. Health Information Technology 3. Health Literacy 4. Patient-Centered Medical Home Scores: CG-CAHPS uses multiple Likert-scales, as well as, ordinal 0 to 10 responses. Scores are calculated for top- (most positive) and bottom-box scores (most negative). Administration Time: 12-15 minutes Format: This survey employs a 4-point response scale – "Never/Sometimes/Usually/Always". Patient self-report or telephone survey.
	Clinical / Quality Information Reliability: CG-CAHPS Reliability data can be found under Appendix A: http://cahps.ahrq.gov/clinician_group/cgsurvey/fieldingcahps-cgsurveys.pdf
	Availability of Benchmarks **NOTE:** The contract that supports the Consumer Assessment of Healthcare Providers and Systems (CAHPS) program expired on June 27, 2012. This will result in a delay in updating the data that compose the CAHPS Database for the SECOND HALF OF CALENDAR YEAR 2012. The Agency for Healthcare Research and Quality is now working to award a new CAHPS contract, and when it is in place, the benchmarking database will be updated with the latest data available, including data collected in 2012.
	Additional Notes/Links Link to survey: http://www.cahps.ahrq.gov/clinician_group/ Distributor: Agency for Healthcare Research and Quality Versions/Languages: English and Spanish; versions include the adult and child 12-month survey (including the domains and supplemental items), and the Visit Specific survey. Original Publication Date: 2007 Copyright: N/A Web Access: http://www.cahps.ahrq.gov/ Contact/Availability: Agency for Healthcare Research and Quality, 540 Gaither Road Rockville, MD 20850 (301) 427-1364 Cost: Free

Tool	Tool Information
CG-CAHPS Visit Survey 2.0 **IT-6.1.b.iv**	**Summary:** Purpose: The Visit Survey asks respondents about experiences during their most recent visit with a provider, as opposed to all of their visits with that provider in the last 12 months. However, most questions about access to care refer to experiences over the last 12 months. The various CAHPS surveys ask consumers and patients to report on and evaluate their experiences with health care. These surveys cover topics that are important to consumers and focus on aspects of quality that consumers are best qualified to assess, such as the communication skills of providers and ease of access to health care services. Overview: Consumer experiences in obtaining health care, including the following five major areas: getting needed care; getting care without long waits; how well doctors communicate; courteous and helpful office staff customer service. Scores: If recommended protocols are followed data can be compared to national and state data. Format: All of the measures are on a Likert scale system. The original 1.0 versions were on a 6-point Likert while the newest version, 2.0, is on a 4-point Likert and has some revisions based on consumer interviews; however, the 2.0 version has not been approved by AHRQ as of yet so the 1.0 Adult version remains available.
	Clinical/Quality information: Reliability: As for internal reliability, studies using the plan level CAHPS version composite scores ranged from Cronbach alphas of .58 to .75 (Hargraves, Hays & Cleary, 2003).
	Availability of Benchmark Data: The contract that supports the Consumer Assessment of Healthcare Providers and Systems (CAHPS) program expired on June 27, 2012. This will result in a delay in updating the data that compose the CAHPS Database for the SECOND HALF OF CALENDAR YEAR 2012. The Agency for Healthcare Research and Quality is now working to award a new CAHPS contract, and when it is in place, the benchmarking database will be updated with the latest data available, including data collected in 2012.
	Additional Notes/Links Link to survey: http://www.cahps.ahrq.gov/clinician_group/cgsurvey/adultvisiteng2.pdf Distributor: Agency for Healthcare Research and Quality Versions/Languages: English; version 2.0 Original Publication Date: 2007 Copyright: N/A Web Access: http://www.cahps.ahrq.gov/ Contact/Availability: Agency for Healthcare Research and Quality, 540 Gaither Road Rockville, MD 20850 (301) 427-1364 Cost: Free

Tool	Tool Information
Client Satisfaction Questionnaire (CSQ Scales) **IT-6.2.a**	**Summary**[4]: Purpose: To measure and assess consumer satisfaction with health and human services. Overview: The CSQ Scales® were created in response to the need for a standard instrument to replace idiosyncratic, ad hoc, and/or untested tools. The goal was to develop a standardized measure with strong psychometric properties that could be used to assess general satisfaction across varied health and human services. The CSQ Scales® (CSQ) include a series of brief instruments. Scores: The overall score is produced by summing all item responses. For the CSQ-8 version, scores range from 8 to 32, with higher values indicating higher satisfaction. Scoring for other versions is similar after extrapolating for number of items. Administration time: 3 to 8 minutes. Reading Level: : 4.7 (Flesch-Kincaid) grade level; 69.4 (Flesch Reading Ease Index) Format: Self-administered, with data collected usually at the end of services. Items are questions inquiring about respondents' opinions and conclusions about services they have received or are currently receiving. Response options differ from item to item, but all are based on a four-point scale.
	Clinical/Quality information[5]**:** The CSQ is documented to have excellent reliability and internal consistency. The CSQ is reported to have high levels of client and staff acceptability when tested in numerous studies involving diverse client samples and a wide range of health and human service programs. In summary, the major strengths of the CSQ include its utility as a standard measure, excellent reliability and internal consistency, acceptability to clients and service providers, and sensitivity to different levels of program quality, and value to service providers committed to enhancement of quality and impact of services (Attkisson & Greenfield, 1996, 2004; Attkisson & Pascoe, 1983; Attkisson & Zwick, 1982; Greenfield, 1983; Larsen, Attkisson, Hargreaves, & Nguyen, 1979; Nguyen, Attkisson, & Stegner, 1983) Norms: Means and standard deviations are available from a series of studies involving approximately 8,000 clients. Most of the studies also report information on the demographics of sample members, who have considerable diversity across samples in terms of both demographic characteristics and services received. Reliability: In a variety of studies, the internal consistency of the CSQ-8, as measured by coefficient alpha, ranged from .83 to .93, with values of .86 and .87 in two of the largest samples. Test-retest results are not reported in the key references for the CSQ-8 but do exist for the CSQ-18A and CSQ-18B. Validity: Items were drawn from an initial pool of 81 statements developed to address nine hypothesized aspects or component dimensions of satisfaction. The pool was reduced to 31 items through reviews by panels of experts, after which these remaining items were tested on 248 mental health clients. The final eight items were selected based on their factor loadings. Predictive validity has been hypothesized and demonstrated by the presence of higher satisfaction scores for service completers as compared to non-completers. Also, treatment outcome comparisons between CSQ-8 scores and those on the Brief Psychiatric Rating Scale were, as predicted, moderately correlated. In another study, CSQ-8 scores, at

[4] http://www.csqscales.com/pdfs/Brief%20Summary%20of%20the%20Client%20Satisfaction%20Questionnaire%20(CSQ%20Scales).pdf

[5] http://www.csqscales.com/pdfs/Brief%20Summary%20of%20the%20Client%20Satisfaction%20Questionnaire%20(CSQ%20Scales).pdf

Tool	Tool Information
	outcome, were found to be correlated positively with symptom reduction, as measured by results on the Client Checklist. From the literature[6]: Client Satisfaction Questionnaire (CSQ-8). Originally developed for use in mental health programs (Larsen et al., 1979), it has since been applied in a variety of other arenas. It is a brief, 8-item scale with very good to excellent internal consistency, based on tested values for coefficient alpha that range from .83 (Roberts & Attkisson, 1983) to .94 (Cox, Brown, Peterson & Rowe, 1982). It is also one-dimensional, consisting of items shown to constitute a single factor in an earlier, 31-item scale (Nguyen, Attkisson, & Stegner, 1983). Findings indicating that it correlates with variables such as program completion provide evidence of its concurrent validity (Attkisson & Zwick, 1982). ***Limitations:*** The CSQ-8 does have drawbacks, however, such as the fact that it offers only four response options (numbered 1 to 4) for each item, which eliminates the possibility of neutral responses and provides less sensitivity than 5- or 7-point scales. Its response options vary throughout the measure, requiring respondents to adjust to different anchors as they move from item to item. In addition, reverse coding is done by changing the directionality of response anchors rather than by varying item wording from positive to negative, and this, too, may require more careful attention by respondents. Perhaps most important, tests conducted on the scale's readability show it has a grade-level equivalent of 6.8 (Kincaid, Aagard, O'Hara, & Cottrell, 1981). This is acceptable for many purposes but still limits the scale's use with important groups such as younger adolescents or adults with educational deficits.
	Availability of Benchmark Data: (1994)- Fisher & Corcoran: Measures for Clinical Practice: A Sourcebook: Adults "The CSQ-8 has been used with a number of populations. The largest single study involved 3268 clients from 76 clinical facilities including inpatients and outpatients. The study involved 42 Mexican Americans, 96 non-Mexican Hispanics, 361 blacks and 2605 whites. Both sexes and a wide range of other demographic variables were included. In essence, the CSQ-8 seems to operate about the same across all ethnic groups. This is also true for a version of the CSQ-8 that was translated into Spanish. The mean scores for the four groups ranged from 26.35 to 27.23 and were not significantly different."7 (2011)- Schulte, et al. Systematic review of dual diagnosis clients' treatment satisfaction- See link to table below for study specific means and SDs as it applies to CSQ-8. http://www.biomedcentral.com/1471-244X/11/64/table/T1
	Additional Notes/Links: Link to survey: http://www.csqscales.com/pdfs/Brief%20Summary%20of%20the%20Client%20Satisfaction%20Questionnaire%20(CSQ%20Scales).pdf Distributor: The CSQ Scales are exclusively distributed by Tamalpais Matrix Systems, LLC (TMS)* Versions and Languages: The CSQ Scales® are published in a variety of scale versions and languages. Version lengths include the CSQ-3, CSQ-4, CSQ-8, CSQ-18A and CSQ-18B and are used in a wide

[6] http://www.sagepub.com/fswrstudy/study/articles/McMurtry.pdf

[7] http://books.google.com/books?id=hwu2wQiIPUwC&pg=PA119&lpg=PA119&dq=CSQ+to+assess+patient+satisfaction&source=bl&ots=tT3aRzHOSn&sig=RpX0OIu1XjGMr6lbHOu5arpuhbw&hl=en&sa=X&ei=xsiCUaOyNcbZ0QH9koGQBg&ved=0CHUQ6AEwCQ#v=onepage&q=CSQ%20to%20assess%20patient%20satisfaction&f=false

Tool	Tool Information
	spectrum of clinical, human services, educational and governmental programs, legal and police services, administrative, and research settings. A longer 31-item version is also available for research purposes and selected evaluation studies in scientific investigations. The CSQ Scales® are used worldwide in the measurement of client/patient assessment of satisfaction with services and clinical care. Language translations now include Arabic, Castilian, Cambodian, Chinese, Czech, Dutch, UK English, French, German, Italian, Japanese, Laotian, Portuguese, Russian, Spanish, Slovak, Swedish, plus many others. Tamalpais Matrix Systems now offers the CSQ-8 in a single sheet, double-sided format containing English and Spanish. CSQ-8 "Big Print" versions are now also available in English and Spanish for readers preferring or requiring larger type size. In addition, the CSQ Scales® Reprint Portfolio containing CSQ Scales® publications is also available from TMS. A CSQ-8 version formatted for use with "Teleform" is also available. Original Publication Date: 1979 Copyright © 1979, 1989, 1990, 2006, 2012 Copyright: C. Clifford Attkisson, Tamalpais Matrix Systems, 35 Miller Avenue, Mill Valley, CA 94941-1903. Voice: 415-310-5396. Fax: 339-440-9537 or 866-770-4975 (US Toll Free). E-mail: Info@CSQscales.com. (*Sample items presented above are reprinted with permission.) Web Access: www.CSQscales.com Fax: 339-440-9537 866-770-4975 (U.S. Toll Free) Contact/Availability: Use is by written permission only from the Copyright holder (Attkisson) and remission of use fees. Cost: For the U.S. English version, cost is $.55 each for first 500 uses, $.45 (U.S. $) for each use thereafter. Inquire for prices on translated versions. Prices are subject to change without notice.
Visit-Specific Satisfaction Instrument (VSQ-9) **IT-6.2.b**	**Summary**: Purpose[8]: The VSQ-9 survey measures patient satisfaction with access to care (questions 1 to 4), with the direct interaction with the physician (questions 5 to 8), and with the visit overall (question 9) on a scale ranging from 1 (poor) to 5 (excellent) (Table 1). Overview: The VSQ-9 was developed for use in the Medical Outcomes Study and focuses specifically on satisfaction with a visit to a physician or other health care provider. The VSQ-9 was adapted for use by the Group Health Association of America, and more recently has been adopted by the American Medical Group Association as the recommended method to measure patient satisfaction with an office visit. The full 9-item survey is often used to compare patient satisfaction rates among physicians, even in large medical groups in which factors such as telephone access and geographic convenience may be beyond the control of the physician. Scores: To score the VSQ-9, the responses from each individual should be transformed linearly to a 0 to 100 scale, with 100 corresponding to "excellent" and 0 corresponding to "poor." Responses to the VSQ-9 items should then be averaged together to create a VSQ-9 score for each person. Administration time: The reliability and relative ease of administration of the VSQ-9 has led to it being adopted by a number of medical groups and health services researchers to measure patient satisfaction with care.[9]

[8] http://w.turner-white.com/pdf/jcom_sep00_problems.pdf
[9] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1492522/

Tool	Tool Information
	Reading Level: 8th grade Format: The VSQ-9 is typically administered in written form and has been administered retrospectively by phone.
	Clinical/Quality information[10]: Norms: Rubin et al. (1993), in their study involving 17,671 patients, reported that the use of a concise patient satisfaction survey was an acceptable method for assessing the level of satisfaction with an outpatient medical encounter. Reliability: Oermann (2003) reported the alpha reliability of the modified instrument to be .86. The modification of the instrument did not appear to affect the reliability of the instrument significantly, as Ware and Hays (1988) reported the internal consistency reliability of the VSQ-9 ranged from .82 to .94. Validity: Ware and Hays also established validity of the VSQ-9 in the 1988 report on two separate studies. From the literature: Testing of a preliminary version of the survey demonstrated that a 5% decrease in measured satisfaction is associated with a 3.4% increase in likelihood that a patient would change medical care providers in the next year. More than one third of the variation in patients' responses to the VSQ-9 satisfaction survey can be explained by a combination of patient characteristics and characteristics of the organizational system. Patients' personal characteristics explain 9% of the variance in patient satisfaction, while organizational characteristics alone explain 27%.
	Availability of Benchmark Data: In a recent study, a nine-item questionnaire commonly used to measure patient satisfaction with care, the Visit-Specific Satisfaction Questionnaire (VSQ-9) did not identify differences in satisfaction with care between minority and white patients. However, analysis of responses to four VSQ-9 questions that specifically asked about direct interaction with physicians revealed that minority patients were significantly less satisfied with physician interaction than white patients. Measurements of patient satisfaction that use multi-item, composite indicators should also include focused comparisons of satisfaction with the care provided by the physician, concludes Donald A. Barr, M.D., Ph.D., of Stanford University. In the study, which was supported by the Agency for Healthcare Research and Quality (HS09350), Dr. Barr11 compared responses to the VSQ-9 for overall patient satisfaction with a four-item subset of the VSQ-9 dealing with the quality of the direct physician-patient interaction. Participants were 537 highly educated primary care patients from an affluent area of California who completed the questionnaire during an office visit to one of the study physicians. For all nine questions of the VSQ-9 instrument, patients had to rate their responses from 1 for poor to 5 for excellent. To provide a benchmark for visit satisfaction, we compared our results on the MOS-VSQ to those obtained by Rubin et al.19 by calculating the percentage of patients who provided an overall rating of their visit as "excellent." Mean score for 537 subjects = 3.29 with SD of .38

[10] http://udini.proquest.com/view/effects-of-a-visit-specific-goid:304320811/
[11] http://onlinelibrary.wiley.com/doi/10.1111/j.1525-1497.2004.30415.x/full

Tool	Tool Information
	Additional Notes/Links: Link to survey: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/vsq9/vsq9.pdf Distributor: RAND Health Versions/Languages: Many of the surveys listed are available in other languages Original Publication Date: 1991, Davies & Ware Web Access: http://www.rand.org/health/surveys_tools/vsq9.html Contact/Availability: All of the surveys from RAND Health are public documents. RAND_Health@rand.org Cost: available without charge (for non-commercial purposes).
RAND Patient Satisfaction Questionnaire (PSQ) **IT-6.2.d**	**Summary**: Purpose: Quantifies global satisfaction with medical care as well as satisfaction with six aspects of care: technical quality, interpersonal manner, communication, financial aspects of care, time spent with doctor, and accessibility of care. Overview: The Patient Satisfaction Questionnaire (PSQ), consisting of 80 items, was originally developed by Ware and his colleagues (Ware, Snyder, and Wright, 1976 a, b). The PSQ has been updated to the PSQ-III[12] (50-items). The PSQ-18[13] is a short form version that retains many characteristics of its full-length counterpart. The PSQ sub-scales show acceptable internal consistency reliability. Furthermore, corresponding PSQ-18 and PSQ-III subscales are substantially correlate with one another. The PSQ-18 may be appropriate for use in situations where the need for brevity precludes administration of the full-length PSQ-III. Scores: PSQ-III[14]: All 50-items are accompanied by five response categories (strongly agree, agree, uncertain, disagree, strongly disagree). Items are grouped in Table 1 according to their scale placement. Once scored as shown in Table 2, items within each subscale are simply summed to yield the subscale score. Table 4 presents complete scoring rules for the seven PSQ-III subscales. Subscales are balanced to control for the effects of acquiescent response set on scores. PSQ-18[15]: The PSQ-18 yields separate scores for each of seven different subscales: General Satisfaction (Items 3 and 17); Technical Quality (Items 2, 4, 6, and 14); Interpersonal Manner (Items 10 and 11); Communication (Items 1 and 13); Financial Aspects (Items 5 and 7); Time Spent with Doctor (Items 12 and 15); Accessibility and Convenience (Items 8, 9, 16, and 18). Some PSQ-18 items are worded so that agreement reflects satisfaction with medical care, whereas other items are worded so that agreement reflects dissatisfaction with medical care. All items should be scored so that high scores reflect satisfaction with medical care (see Table 1). After item scoring, items within the same subscale should be averaged together to create the 7 subscale scores (see Table 2). We recommend that items left blank by respondents (missing data) be ignored when calculating scale scores. In other words, scale scores represent the average for all items in the scale that were answered.

[12] http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq3_survey.pdf
[13] http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq18_survey.pdf
[14] http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq3_scoring.pdf
[15] http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq18_scoring.pdf

Tool	Tool Information
	Administration time: The PSQ-18 takes approximately 3-4 minutes to complete. Format: The PSQ-III is a 50-item survey that taps global satisfaction with medical care as well as satisfaction with six aspects of care: technical quality, interpersonal manner, communication, financial aspects of care, time spent with doctor, and accessibility of care.
	Clinical/Quality information: Reliability: PSQ-III: Reliability estimates ranged from 0.77 to 0.89 in the MOS baseline sample, and fell below 0.80 only for the two-item Time Spent with Doctor subscale. As expected, the Access/Availability/Convenience subscale proved to be the most heterogeneous, as reflected in its low homogeneity estimate. As illustrated in Davies et al. (1986), we recommend item-by-item analyses before relying on a summary score where comparing systems of care in terms of satisfaction with accessibility, availability, and convenience. Validity: Validity of the PSQ-III and PSQ-18 can be found at: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq3_scoring.pdf From the literature:
	Availability of Benchmark Data:
	Additional Notes/Links: Link to survey: PSQ-III: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq3_survey.pdf PSQ-18: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq18_survey.pdf Distributor: RAND Health (http://www.rand.org/health/surveys_tools/psq.html) Versions/Languages: Many of the RAND surveys are available in other languages. Original Publication Date: Patient Satisfaction Questionnaire (PSQ), 1976 Copyright: Web Access: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/psq/psq18_survey.pdf Fax: Contact/Availability: RAND_Health@rand.org Cost: All of the surveys from RAND Health are public documents, available without charge.

<table>
<tr><th>Tool</th><th>Tool Information</th></tr>
<tr><td>Assessment of Quality of Life (AQoL) [16]

IT-10.1.a.i</td><td>
Summary:

Purpose: Quantifies patient health-related quality of life as a psychometric and/or utility (index of overall health state utility) measure.

Overview:
As a 'psychometric' measure: Each instrument can be used to derive a simple psychometric score for health related quality of life (HRQoL) and to provide profile scores on the different dimensions or items of the descriptive systems. The score is derived by adding the unweighted response order of each question.
As a 'utility' measure: When utilities are computed, these instruments can provide dimension scores and an overall index of the health state utility which can be used in economic evaluations, and specifically, cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). The 'utilities' are, in effect, preference weights and final utility scores should reflect peoples' preferences more accurately than unweighted aggregates.

Scores:
<table>
<tr><th>Responses</th><th>Multi-Attribute 'Psychometric' Instrument
• Unweighted responses
• Sum responses</th><th>Multi-Attribute 'Utility' Instrument
• Utility weighted responses
• Sum weights to use as additive HRQoL measure</th></tr>
<tr><td>Never</td><td>1</td><td>1</td></tr>
<tr><td>Rarely</td><td>2</td><td>0.86</td></tr>
<tr><td>Some of the time</td><td>3</td><td>0.58</td></tr>
<tr><td>Often</td><td>4</td><td>0.20</td></tr>
<tr><td>Nearly all the time</td><td>5</td><td>0</td></tr>
</table>

Format: Four different instruments measuring different healthcare dimensions:
AQoL-8D: Independent Living, Happiness, Mental Health, Relationships, Self-Worth, Pain, Senses
AQoL-7D: Independent Living, Mental Health, Coping, Relationships, Pain, Senses, Visual Impairment

AQoL-6D: Independent Living, Mental Health, Coping, Relationships, Pain, Senses
AQoL-4D: Independent Living, Mental Health, Relationships, Senses

Administration time:
AQoL-8D: < 6 minutes
AQoL-7D: 3-4 minutes
AQoL-6D: 2-3 minutes
AQoL-4D: 1-2 minutes

Reading Level: The AQoL reading level of 71% (Flesch Reading Ease score) suggests that the instrument should be acceptable to most literate individuals.
</td></tr>
</table>

[16] http://www.aqol.com.au/index.php/choice-of-aqol-instrument

Tool	Tool Information
	Clinical/Quality information: Factors and Norms: **AQoL-4D** Hawthorne, G., Korn, S. and Richardson, J. (2013). Population norms for the AQoL derived from the 2007 Australian National Survey of Mental Health and Wellbeing, *Australian and New Zealand Journal of Public Health* **37**(1): 7–16. **AQoL-6D**[17] **AQoL-7D** – Under Development **AQoL-8D**[18] Reliability/Validity[19]: **Evidence of construct validity**: The AQoL suite of instruments were the only ones to be constructed using psychometric methods developed by psychologists for achieving content and construct validity. *Other MAU instrument descriptive systems have been based upon 'logical' considerations* (face validity) or in the case of the SF6D upon another instrument, the SF36. The need for construct validity was the motivating reason for the commencement of the AQoL program. **Evidence of criterion validity for the combination model**: That is, a demonstration that the combination model predicts the scores of multi attribute health states when they are independently measured. AQoL-6D, 7D, 8D have demonstrated this property. To date other tests of this property do not appear to have been widely conducted for other instruments. **Evidence of the validity of utility (scaling) instruments (TTO, SG, etc.)**: Like other MAU instruments, the AQoL suite of instruments have assumed the validity of a particular scaling instrument, viz. the TTO. However the issue is problematic (and largely ignored in the literature) **Evidence of correlation between MAU utility scores and other instrument scores**: This constitutes the overwhelming majority of studies 'validating' other instruments. It is necessary but far from sufficient for demonstrating validity. As the newest of the MAU instruments, the AQoL suite of instruments has not been included in many such studies (and results from the 100+ studies using an AQoL-4D instrument have not been completed and compiled). However AQoL-4D and 8D have been included in five instrument studies which indicate a sufficiently high correlation between different instruments to confidently predict that this type of evidence will be obtained. As noted elsewhere, however, this type of evidence is 'soft' in the sense that it is easily achieved as even instruments with overall low content validity are likely to correlate with other QoL instruments. **Evidence of overall criterion validity**: No instrument has shown that the percentage change in predicted utility from an MAU instrument corresponds with the same percentage change in real trade-off decisions between the quantity and quality of life. The overall status of MAU instruments, qua predictors of utility, therefore remains problematical. More reliability and validity data can be found at http://onlinelibrary.wiley.com/doi/10.1002/acr.20541/full.
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): http://www.buseco.monash.edu.au/centres/che/pubs/researchpaper66.pdf Clinical Indicator: AQoL: The mean (SD) AQoL utility score was 0.83 (0.20). Gender and age subgroup differences were

[17] http://www.aqol.com.au/documents/AQoL-6D/6D_Population_Norms_021209.pdf
[18] http://www.aqol.com.au/index.php/norms
[19] http://www.aqol.com.au/index.php/validation-faqs

Tool	Tool Information
	apparent; the mean scores for women were consistent until their 50s, when scores declined. Greater variability was observed for males whose scores declined more slowly but consistently between 40–80 years. For both genders, those aged 80+ years had the lowest scores When assessed by health status, those reporting excellent health obtained the highest utility scores; progressive declines were observed with decreasing health status. Effect sizes of 0.13 or greater may reflect important differences between groups A difference in AQoL scores of 0.06 utility points over time suggests a general MID (minimal important differences).[20] The AQoL utility score ranges from −0.04 (health state worse than death) to 0.00 (death) and 1.00 (full health). Normative values, broken down by age (in 10-year age groups) and sex, are available for AQoL-4D from the AQoL web site (http://www.aqol.com.au/documents/AQoL-4D-Population-Norm.pdf). The norms have been derived from a probability sample of 3,010 Australian residents.[21]
	Additional Notes/Links: Link to survey: http://www.aqol.com.au/index.php/aqolquestionnaires Distributor: AQoL Versions/Languages: Spanish, German, Danish, Chinese, Italian Copyright: 2009 Web Access: http://www.aqol.com.au/index.php/contact-aqol-group Contact/Availability: angelo.iezzi@monash.edu Cost: Free
Pediatric Quality of Life Inventory (PedsQL)[22] **IT-10.1.a.ii**	**Summary**: Purpose: The PedsQL™ Measurement Model is a modular approach to measuring health-related quality of life (HRQOL) in healthy children and adolescents and those with acute and chronic health conditions. The PedsQL™ Measurement Model integrates seamlessly both generic core scales and disease-specific modules into one measurement system. Overview: The 23-item PedsQL™ Generic Core Scales were designed to measure the core dimensions of health as delineated by the World Health Organization, as well as role (school) functioning. Two versions are available: PedsQL Child-Self Report (ages 8-12 years) and Parent-Proxy Report for Children (ages 8-12 years). The PedsQL measure 4 Multidimensional Scales (Physical [8 items], Emotional [5 items], Social [5 items], and School Functioning [5 items]), and 3 Summary Scores (Total Scale Score, Physical Health and Psychosocial Health Summary Scores). Scores: 1. On the PedsQL™ Generic Core Scales, for ease of interpretability, items are reversed scored and

[20] http://onlinelibrary.wiley.com/doi/10.1111/j.1467-842X.2005.tb00063.x/abstract;jsessionid=4177E7408411BEB2798A61248D388BD1.d01t03?systemMessage=Wiley+Online+Library+will+be+disrupted+on+31+August+from+10%3A00-12%3A00+BST+%2805%3A00-07%3A00+EDT%29+for+essential+maintenance&userIsAuthenticated=false&deniedAccessCustomisedMessage=

[21] http://onlinelibrary.wiley.com/doi/10.1002/acr.20541/full

[22] http://www.pedsql.org/index.html

Tool	Tool Information
	linearly transformed to a 0-100 scale, so that higher scores indicate better HRQOL (Health-Related Quality of Life). 2. To reverse score, transform the 0-4 scale items to 0-100 as follows: 0=100, 1=75, 2=50, 3=25, 4=0 3. To create the Psychosocial Health Summary Score, the mean is computed as the sum of the items over the number of items answered in the Emotional, Social, and School Functioning Scales. The Physical Health Summary Score is the same as the Physical Functioning Scale Score. 4. To create the Total Scale Score, the mean is computed as the sum of all the items over the number of items answered on all the Scales. Administration time: < 4 minutes Format: 23-item PedsQL Generic Core Scales Reading Level: Ages 2-18; Child Self-Report Ages 5-7, 8-12, 13-18; Parent Proxy-Report Ages 2-4, 5-7, 8-12, 13-18.
	Clinical/Quality information: Factors and Norms: Accordingly, the PedsQL™ Measurement Model consists of developmentally appropriate forms for children ages 2-4, 5-7, 8-12, and 13-18 years. Pediatric self-report is measured in children and adolescents ages 5-18 years, and parent proxy-report of child HRQOL is measured for children and adolescents ages 2-18 years. Reliability: Internal consistency reliability of the PedsQL™ was excellent, with alphas for the generic core scales in both self- and proxy-report greater than the 0.70 standard, and alphas for the full 23-item scale approaching 0.90 for self- and proxy-report. Missing data were minimal. Item response distributions were across the full scale range, with no floor effects, and minimal ceiling effects. Internal consistencies for the total scale score were as follows: child self-report Cronbach's 0.91, parent-proxy report Cronbach's 0.93; physical ealth summary scale score: child self-report Cronbach's 0.87, parent-proxy report Cronbach's 0.89; and psychosocial health summary scale score: child self-report Cronbach's 0.86, parent-proxy report Cronbach's 0.90.[23] Validity: The validity of the PedsQL™ Generic Core Scales was demonstrated through known groups comparisons, and correlations with other measures of disease burden. The PedsQL™ self- and proxy-report distinguished between children with and without a chronic health condition, and within the group of children with a chronic condition, between those who did or did not have an overnight hospital visit in the last 12 months. Further, both child self-report and parent proxy-report correlated significantly with the number of days the child was too ill to pursue normal activities, needed someone to care for him or her, missed school in the last month, the number of days the parent missed from work in the last month, and parent-report of problems pursuing their normal work routine and concentrating at work. The PedsQL™ Generic Core Scales are also responsive to clinical change, as demonstrated in field trials.
	Availability of Benchmark Data Benchmark(s):

[23] http://onlinelibrary.wiley.com/doi/10.1002/acr.20637/pdf

Tool	Tool Information
	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1769359/ Clinical Indicator**:** The range on subscales and the overall scale is 0–100, with lower scores indicating poorer HRQOL and higher scores indicating better HRQOL. When examining the total scale, scores of 4.4 and 4.5 are considered to be minimal clinically meaningful differences on the child self-report and parent-proxy report, respectively.[24]
	Additional Notes/Links: Link to survey: Child-Self Report Ages 8-12: http://www.pedsql.org/pedsql12.html Parent-Proxy Report Ages 8-12: http://www.pedsql.org/pedsql3.html Distributor: James W. Varni, Ph.D., PedsMetrics Versions/Languages: Child-Self Report Ages 8-12, Parent-Proxy Report Ages 8-12/ English Original Publication Date: 1998 Copyright: 1998-2013; James W. Varni, PhD Web Access: http://www.pedsql.org/pedsql12.html Fax: +33 4 72 13 55 73 Contact/Availability: PROinformation@mapi-trust.org; +33 4 72 13 65 75 Cost: Free, but Limited Use License must be agreed too
RAND Medical Outcomes Study: Measures of Quality of Life Survey Core Survey (MOS) **IT-10.1.b.i**	**Summary**: Purpose: The Medical Outcomes Study (MOS) measurement of physical functioning offers an extended ADL scale sensitive to variations at relatively high levels of physical function. It is suitable for use in health surveys and in outcome assessment for outpatient care. Overview: The MOS measures were based on a comprehensive conceptual model that includes two overarching dimensions of health--physical and mental (Stewart, Sherbourne, Hays et al, 1992). Hays and Stewart (1990) provide empirical support for these two health dimensions. Conceptually, the MOS measures were constructed to represent the following: Physical Health (e.g., physical functioning, satisfaction with physical ability, mobility, pain effects, pain severity, role limitations due to physical health), Mental Health (e.g., psychological distress – anxiety and depression, psychological well-being – positive affect and feelings of belonging, cognitive functioning, role limitations due to emotional problems), and General Health (e.g., energy/fatigue, sleep problems, psychophysiological symptoms, social functioning, role functioning – unable to work, role functioning – unable to do housework, current health perceptions, and health distress). Scores: [25]Three scores are derived. A physical function score is formed by averaging non missing items from question 1; the score is transformed to a 0 to 100 scale in which a higher score indicates better

[24] http://onlinelibrary.wiley.com/doi/10.1002/acr.20637/pdf
[25] http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/mos/mos_core_scoring.pdf

Tool	Tool Information
	function. People omitting more than five items receive a missing score. A satisfaction score is based on item 2, transformed to a 0 to 100 scale. Reading Level: 6th grade[26] Format: The MOS has 116 items across 37 domains (see http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/mos/mos_core_scoring.pdf for more information). The MOS is an interviewed questionnaire.
	Clinical/Quality Information: Reliability: Eight of ten physical function items correlated 0.70 or greater with the overall physical scale score; the vigorous activity item correlated 0.62 and the bathing or dressing item showed a lower correlation of 0.48 (1, Table 6-3). Internal consistency for the functioning score was 0.92; for the mobility scale it was 0.71 (1, p98). The alpha internal consistency of a slightly modified version of the scale was 0.92 in a sample of 1,054 elderly respondents; intra-class test-retest reliability was 0.93 on a subset of 52 (3, Table 4). Validity: The physical functioning scale scores correlated 0.58 with the mobility scores and 0.63 with the satisfaction scores (1, Table 6-6). A factor analysis identified a single factor accounting for 70% of the variance.
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): http://www.rand.org/content/dam/rand/pubs/monograph_reports/2008/MR162.pdf http://www.qualitymetric.com/Portals/0/Uploads/Documents/Public/EASD2012_NOVO_PosterBjornerQRV2.pdf Clinical Indicator[27]: All scales are scored so that a high score defines a more favorable health state. The third step in scoring each scale involves recoding item responses for those items that are not asked in a direction consistent with a favorable health state. For example, the item CORE8g asks, "Did you have enough energy to do the things you wanted to do?" If a respondent answers, "none of the time," the precoded response of "6" must be reversed so higher scores will indicate a favorable health state (i.e., more frequent occurrences of having enough energy).
	Additional Notes/Links: Link to survey http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/mos/mos_core_survey.pdf Distributor: RAND Original Publication Date: 1992 Web Access: http://www.rand.org/health/surveys_tools/mos.html Contact/Availability: RAND_Health@rand.org

[26] http://www.qualitymetric.com/Portals/0/Uploads/Documents/Public/QM_Catalog_2011.pdf
[27] http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/mos/mos_core_scoring.pdf

Tool	Tool Information
	Cost: Free
RAND Short Form 12 (SF-12v2) **Health Survey** **IT-10.1.b.ii**	**Summary**: Purpose: To measure functional health and well-being from the patient's point of view. The SF-12v2 is a practical, reliable and valid measure of physical and mental health that is particularly useful in large populations or for applications that combine generic and disease-specific health surveys. Overview: SF-12v2 is an abbreviated version of the SF-36v2 Health Survey that uses 12 of the SF-36v2 items to measure each of the following eight health domains: Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. The SF-12v2 should be in used among individuals 18 years or older. Scores: Each health domain score contributes to the psychometrically-based Physical Component Summary (PCS) and Mental Component Summary (MCS) scores. The SF-12v2 uses T scores (with a mean of 50 and a standard deviation of 10) rather than the original 1-100 for simplified interpretation. When interpreting norm-based scores, one does not have to remember the norms for eight health domain scales; the general population norm is built into the scoring algorithm. For all scales and summary measures, individual respondent scores below 45 and group mean scores below 47 can be interpreted as being below the average range for the general population. And because the standard deviations for each scale are equalized at 10, it is easier to see exactly how far below or above the general population mean a score is in standard deviation units, and comparisons of health domain scale and component summary measure scores. Administration time: 2-3 minutes. Reading Level: 6th grade Format: The SF-12v2 is available in multiple modes of administration and in both standard four-week and acute one-week recall periods. The fixed form is a paper-based version that allows the individual to complete the form themselves. An interviewer script administration provides scripts for the interviewer to follow when the patient is unable to complete the survey on their own. The survey consists of 12 questions with 1-2 questions per health domain.
	Clinical/Quality information: To assess the reliability and validity of the 12-item Health Survey (SF-12), researchers compared component scores with normative levels, examined test-retest reliability, and examined convergent and divergent validity by comparing SF-12 scores to other indexes of physical and mental health. The Department of Psychology, Indiana University-Purdue found the SF-12 distinguished this sample of people with SMI from the general population, was stable over a 1-week interval, consisted of 2 fairly distinct factors, and was related to physical and mental health indexes in expected ways. They concluded the SF-12 appears to be a psychometrically sound instrument for measuring health-related quality of life for people with serious mental illness.[28] Studies from the Health Institute of the New England Medical Center in Boston, MA reported the following: Reliability: Regression methods were used to select and score 12 items from the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) to reproduce the Physical Component Summary and Mental Component Summary scales in the general US population (n=2,333). Twenty cross-sectional and

[28] Med Care. 1996 Mar;34(3):220-33; Health Institute, New England Medical Center, Boston, Massachusetts, USA; Ware J Jr., Kosinski M, Keller SD

Tool	Tool Information
	longitudinal tests of empirical validity previously published for the 36-item short-form scales and summary measures were replicated for the 12-item Physical Component Summary and the 12-item Mental Component Summary, including comparisons between patient groups known to differ or to change in terms of the presence and seriousness of physical and mental conditions, acute symptoms, age and aging, self-reported 1-year changes in health, and recovery for depression. Validity: In 14 validity tests involving physical criteria, relative validity estimates for the 12-item Physical Component Summary ranged from 0.43 to 0.93 (median=0.67) in comparison with the best 36-item short-form scale. Relative validity estimates for the 12-item Mental Component Summary in 6 tests involving mental criteria ranged from 0.60 to 107 (median=0.97) in relation to the best 36-item short-form scale. Average scores for the 2 summary measures closely mirrored those for the 36-item short-form, although standard errors were nearly always larger for the 12-item short-form.[29]
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator: Physical and Mental Health Composite Scores (PCS & MCS) are computed using the scores of twelve questions and range from 0 to 100, where a zero score indicates the lowest level of health measured by the scales and 100 indicates the highest level of health.[30]
	Additional Notes/Links: Link to Survey/Report Demo: http://www.qualitymetric.com/tabid/238/Default.aspx Additional: http://www.sf-36.org/community/SF36v2andSF12v2.shtml Distributor: QualityMetric, Inc. Certified to administer the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Survey in English and Spanish for state Medicaid and managed care organizations across the country. • 28,000 licenses issued to date • 37,000,000 surveys taken Versions and Languages: The SF Surveys are available in 110 languages are published in a variety of scale versions and languages. Version lengths include the SF-36v2, SF-8, SF-12.2, SF-12v2 MH Enhanced and SF-12v2 SET. Original Publication Date: SFv2- 2005 Web Access: http://www.qualitymetric.com/Default.aspx Fax: (401) 334-8801 (800) 572-9394 (U.S. Toll Free) Contact/Availability: Use is by written permission only from the Copyright holder (QualityMetric) and remission of use fees. Cost: Data collection, scoring and reporting the results for any survey requires a license from QualityMetric or one of its authorized resellers. The license fee depends on the survey, the number of uses, the type of report requested, and other considerations. SF-12v2 User's Manual: PDF $150.00

[29] Med Care. 2000 Nov;38(11):1141-50; Salyers MP, Bosworth HB, Swanson JW, Lamb-Pagone J, Osher FC; Department of Psychology, Indiana University-Purdue University, Indianapolis, USA.

[30] http://health.utah.gov/opha/publications/2001hss/sf12/SF12_Interpreting.pdf

Tool	Tool Information
RAND Short Form 20 (SF-20) Health Survey[31] **IT-10.1.b.iii**	**Summary**: Purpose: The SF-20 assesses six important health concepts: number of items included: physical functioning (6 items), role functioning (2 items), social functioning (1 item), mental health (5 items), current health perceptions (5 items), and pain (1 item). Overview: **Physical Functioning:** Six items were selected to assess physical functioning, a dimension measured in the HIE by aggregating twenty items measuring physical limitations and capacities, mobility, and self-care (Stewart, Ware, and Brook, 1978; Stewart, Ware, and Brook, 1982.). The goal was to approximate as closely as possible the 6-level scale constructed in the HIE. Response choices and item wording were modified from the HIE version to capture better specific limitations of interest, to describe more accurately the scale level defined by each item, and to facilitate oral administration. One new item (moderate level of limitation in physical functioning) was added to fill a gap in the HIE scale. **Role Functioning:** Two items were selected to measure limitations in role functioning due to poor health. These are the two best items from the 3-item HIE role functioning scale (Stewart, Ware, and Brook, 1978, 1981, 1982). **Social Functioning:** Social functioning is defined as the ability to develop, maintain, and nurture major social relationships. The single social functioning survey item focuses on whether the respondent's health has limited social activities. **Mental Health:** General mental health was assessed using five items derived empirically from the HIE Mental Health Inventory (MHI). This set is the best 5-item predictor of the summary score based on the full 38-item MHI (Davies, Sherbourne, Peterson, and Ware, 1988). The set represents the four major mental health dimensions (anxiety, depression, loss of behavioral-emotional control, and psychological well- being) as confirmed in factor-analytic studies of the MHI (Veit and Ware, 1983). These five items correlated 0.92 with the MHI total score in the HIE sample used to derive the short-form scale. On cross-validation using another HIE sample, the scale correlated 0.92 with the MHI total score (Davies, Sherbourne, Peterson, and Ware, 1988). **Current Health Perceptions:** The 22-item Health Perceptions Questionnaire (HPQ) (Davies and Ware, 1981; Ware and Karmos, 1976; Ware, 1976) included six subscales that are substantially intercorrelated. The Current Health subscale was the most reliable and empirically valid of these (Davies and Ware, 1981). That subscale also best represents the overall HPQ concept, accounting for the largest amount of variance common to the HPQ subscales. For these reasons, four items were selected from the Current Health subscale (Davies and Ware, 1981). These items had high correlations with the Current Health subscale, had substantial and roughly equal correlations with other physical and mental health measures, and achieved the balance between favorably and unfavorably worded items necessary to control for acquiescent and opposition response sets. **Pain:** The survey includes one measure of pain that asks respondents to rate pain on a scale from none to very severe. Scores:

[31] http://www.rand.org/health/surveys_tools/mos/mos_core_20item_more.html

Tool	Tool Information
	Consistent with previous studies, limitations in physical and role functioning were counted regardless of duration and were scored to reflect the number of limitations present (Stewart et al. 1981; Stewart et al., 1982) Scores were reversed so that a high value indicated better functioning. Mental health scales were scored by summing the item responses, after reversing the scoring of some items, so that a high score indicated better health. Before combining items in the health perceptions scale, the response choices of the overall health item (item 1) were recoded to better reflect the unequal intervals of the item. The single-item measures were scored so that high scores indicated better social functioning and more pain. Finally, for all measures, scores were transformed linearly to 0-100 scales, with 0 and 100 assigned to the lowest and highest possible scores, respectively. Administration time: Self-administration - ~3 minutes; telephone surveys require about three or four minutes. Format: The SF-20 includes 20 items adapted from longer surveys used successfully in the Health Insurance Experiment (HIE). It assesses six important health concepts: number of items included: physical functioning (6 items), role functioning (2 items), social functioning (1 item), mental health (5 items), current health perceptions (5 items), and pain (1 item).
	Clinical/Quality information: Reliability/Validity: Support for the reliability and construct validity of the SF-20 is provided in previously published documents (Stewart et al., 1988, Ware et al., 1992). The document can be purchased from Duke University Press for $84.95 (http://www.dukeupress.edu/Catalog/ViewProduct.php?productid=1017&viewby=title)
	Availability of Benchmark Data/Clinical Indicator: Clinical Indicator: Scores were reversed so that a high value indicated better functioning. Mental health scales were scored by summing the item responses, after reversing the scoring of some items, so that a high score indicated better health. Before combining items in the health perceptions scale, the response choices of the overall health item (item 1) were recoded to better reflect the unequal intervals of the item. The single-item measures were scored so that high scores indicated better social functioning and more pain. Finally, for all measures, scores were transformed linearly to 0-100 scales, with 0 and 100 assigned to the lowest and highest possible scores, respectively.[32]
	Additional Notes/Links: Link to survey: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/mos/mos_core_20item_survey.pdf Versions/Languages: Many of the surveys listed are available in other languages. If you interested in translating any surveys into another language, see translation guidelines

[32] http://www.rand.org/health/surveys_tools/mos/mos_core_20item_more.html

Tool	Tool Information
	Distributor: RAND Original Publication Date: N/A Web Access: http://www.rand.org/content/dam/rand/www/external/health/surveys_tools/mos/mos_core_20item_survey.pdf Availability: Online Cost: Free
RAND Short Form 36[33] (SF-36) Health Survey **IT-10.1.b.iv**	**Summary**: Purpose: SF-36 is a set of generic, coherent, and easily administered quality-of-life measures. These measures rely upon patient self-reporting and are now widely utilized by managed care organizations and by Medicare for routine monitoring and assessment of care outcomes in adult patients. Overview: Measures eight health concepts: physical functioning (10 items), bodily pain (2 items), role limitations due to physical health problems (4 items), role limitations due to personal or emotional problems (3 items), emotional well-being (5 items), social functioning (2 items), energy/fatigue (4 items), general health problems (5 items), and indication of perceived change in health (1 item). Scores: First, use the scoring key to score the responses. Each item is scored on a 0 (lowest) to 100 (highest) range. Scores represent the percentage of total possible score achieved. Second, items in the same scale are averaged together to create the 8 scale scores. Items left blank are not taken into account when calculating the scale score. Hence, scale scores represent the average for all items in the scale that the respondent answered. Scoring Tables; http://www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html Administration time[34]: 5-10 minutes Format: Comprehensive short-form with only 36 questions yields an 8-scale health profile as well as summary measures of health-related quality of life.
	Clinical/Quality information: Factors and Norms[35]: With the release of SF-36v2, norms were updated using data from the 1998 National Survey of Functional Health Status (NSFHS) and norm-based scoring (NBS) algorithms were introduced for all eight scales (Ware et al., 2000). NBS, which employs a linear T-score transformation with mean = 50 and standard deviation = 10, makes it possible to meaningfully compare scores for the eight-scale profile and the physical and mental summary measures in the same graph. SF-36v2 scoring software also yields less biased estimates of missing responses and makes it possible to estimate scores for more respondents with incomplete data (Kosinski, Bayliss, Bjorner, & Ware, 2000). Reliability: The reliability of the eight scales and two summary measures has been estimated using both internal

[33] http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html
[34] http://www.sf-36.org/tools/sf36.shtml
[35] http://www.sf-36.org/tools/sf36.shtml

Tool	Tool Information
	consistency and test-retest methods. With rare exceptions, published reliability statistics have exceeded the minimum standard of 0.70 recommended for measures used in group comparisons in more than 25 studies (Tsai, Bayliss, & Ware, 1997); most have exceeded 0.80 (McHorney et al., 1994; Ware et al., 1993). Reliability estimates for physical and mental summary scores usually exceed 0.90 (Ware et al., 1994). A review of the first15 published studies revealed that the median reliability coefficients for each of the eight scales was equal or greater than 0.80 except for SF, which had a median reliability across studies of 0.76 (Ware et al., 1993). In addition, a reliability of 0.93 has been reported for the MH scale using the alternate forms method, suggesting that the internal-consistency method underestimated the reliability of that scale by about three percent (McHorney & Ware, 1995). Validity: The validity, and therefore the interpretation, of each of the eight scales and the two summary measures has been shown to differ markedly, as would be expected from factor analytic studies of their construct validity (see Figure 2) (McHorney et al., 1993; Ware et al., 1994; Ware, Kosinski, Bayliss, McHorney, Rogers, & Raczek, 1995). Specifically, the MH, RE, and SF scales and the MCS summary measure have been shown to be the most valid of the SF-36 scales as mental health measures. This pattern of results has been replicated in both cross-cultural and longitudinal tests using the method of known-groups validity. The PF, RP, and BP scales and the PCS summary have been shown to be the most valid SF-36 scales for measuring physical health. Criteria used in the known-groups validation of the SF-36, which include accepted clinical indicators of diagnosis and severity of depression, heart disease, and other conditions, are well documented in peer-reviewed publications and in the two users" manuals (Kravitz, Greenfield, Rogers, Manning, Zubkoff, Nelson, Tarlov, & Ware, 1992; McHorney et al., 1993; Ware et al., 1993; Ware et al., 1994; Ware et al., 1995). From the literature: The experience to date with the SF-36 has been documented in nearly 4,000 publications; citations for those published in 1988 through 2000 are documented in a bibliography covering the SF-36 and other instruments in the “SF” family of tools (Turner-Bowker, Bartley, & Ware, 2002). The most complete information about the history and development of the SF-36, its psychometric evaluation, studies of reliability and validity, and normative data is available in the first of three SF-36 user’s manuals (Ware, Snow, Kosinski, & Gandek, 1993). This information was also summarized in the first two peer-reviewed articles about the SF-36 (Ware & Sherbourne, 1992; McHorney, Ware, & Raczek, 1993). A second manual documents the development and validation of the SF-36 physical and mental component summary measures and presents norms for those measures (Ware, Kosinski, & Keller, 1994; Ware, Kosinski, & Dewey, 2000). These user’s manuals have been updated to include more up-to-date norms and other findings and to document the much improved Version 2.0 (SF-36v2), which are discussed below (Ware et al., 2000; Ware & Kosinski, 2001) A fourth manual, first published in 1995 (Ware, Kosinski, & Keller, 1995) and recently updated (Ware, Kosinski, Turner-Bowker, & Gandek, 2002) presents similar information for the SF-12 Health Survey, an even shorter version constructed from a subset of 12 SF-36 items.
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): User’s Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life: http://www.rand.org/pubs/monograph_reports/MR162.html The estimation of a preference-based measure of health from the SF-36: http://www.sciencedirect.com/science/article/pii/S0167629601001308 Clinical Indicator: General population norms provide a much better basis for comparisons across scales. For example, the Physical Functioning (PF) scale averages between 80 and 90 while the Vitality (VT) average score is below 60 (on the 100-point score range) in the general population. In relation to these norms, the impact of asthma appears much larger on the PF scale than on the VT scale, although both are statistically significant. Using the original 0–100 scoring, these differences in norms must be kept in mind when interpreting a profile. Differences in standard deviations, which are also substantial across some

Tool	Tool Information
	scales, must also, be considered for this purpose.[36]
	Additional Notes/Links: Link to survey: http://www.rand.org/health/surveys_tools/mos/mos_core_36item_survey.html Distributor: RAND Versions/Languages: English/Arabic Original Publication Date: N/A Availability: Online Web Access: http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html Contact/Availability: RAND_Health@rand.org Cost: Free
Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) **IT-10.1.c**	**Summary**: Purpose[37]: To assess the degree of enjoyment and satisfaction experienced by subjects in various areas of daily functioning. Overview: The summary scores are found to be reliable and valid measures of dimensions. The Q-LES-Q measures are related to, but not redundant with, measures of overall severity of illness or severity of depression within the sample. Scores: Q-LES-Q-SF (18-item) [38]: The scoring of the Q-LES-Q-SF involves summing only the first 14 items to yield a raw total score. The last two items are not included in the total score but are standalone items. The raw total score ranges from 14 to 70. The raw total score is transformed into a percentage maximum possible score using the following formula: (raw total score −minimum score) / (maximum possible raw score −minimum score). The minimum raw score on the Q-LES-Q-SF is 14, and the maximum score is 70. Thus the formula for % maximum can also be written as (raw score −14)/56. Calculation tables can be found at: https://outcometracker.org/library/Q-LES-Q-SF.pdf Format: The Q-LES-Q is 96 items and the Q-LES-Q-SF is 18 items; the questionnaires are self-administered.
	Clinical/Quality information: Reliability/Validity: The internal consistency and test-retest coefficients were 0.9 and 0.93, respectfully. Almost all items significantly correlated to the total score and other measures used in the study, with the correlations ranging 0.41-0.81. Finally, the responsiveness parameters indicated the Q-LES-Q-SF is 80% sensitive and 100% specific measure.[39] Mick, E., Faraone, S., Spencer, T., Zhang, H., Biederman, J. (2008) Assessing the Validity of the Quality of Life Enjoyment and Satisfaction Questionnaire Short Form in Adults With ADHD. Journal of

[36] http://www.sf-36.org/tools/sf36.shtml
[37] http://www.proqolid.org/instruments/quality_of_life_enjoyment_and_satisfaction_questionnaire_q_les_q
[38] https://outcometracker.org/library/Q-LES-Q-SF.pdf
[39] http://www.ncbi.nlm.nih.gov/pubmed/21896118

Tool	Tool Information
	Attention Disorders, 11, No. 4, 504-509
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): Validity of an abbreviated quality of life and satisfaction questionnaire (Q-LES-Q-18) for schizophrenia, schizoaffective, and mood disorder patients: http://www.ncbi.nlm.nih.gov/pubmed/16119181 Clinical Indicator[40]: The Q-LES-Q scores are converted into a percentile enjoyment/satisfaction. The following is the quartile breakdown: Score 14-28: ≤ 25% Score 29-42: 27 – 50% Score 43-56: 52 – 75% Score ≥ 57: 77+% *Note: No clinical interpretations were identified.*
	Additional Notes/Links: Link to survey: Q-LES-Q-SF (18-item): https://outcometracker.org/library/Q-LES-Q-SF.pdf Distributor: Versions/Languages: Translations available can be found at: http://www.proqolid.org/instruments/quality_of_life_enjoyment_and_satisfaction_questionnaire_q_les_q Original Publication Date: 1993 (Jean Endicott, PhD) Availability: Online Cost: None
Satisfaction with Amplification in Daily Living (SADL) **IT-10.1.d**	**Summary**: Purpose: The SADL (Cox & Alexander, 1999) was designed to quantify hearing aid satisfaction.[41] Overview: Quantifies hearing aid satisfaction using 15-items among four sub-scales: positive effects, service &costs, negative features, and personal image.[42] Scores: http://www.harlmemphis.org/files/4113/5412/2968/SADLScoring.pdf
	Clinical/Quality information: Factors and Norms: The SADL was normed on between 126 and 225 adults, depending on the subscale.[43] The SADL Scale Norms can be found here: http://www.harlmemphis.org/files/7013/5412/2932/SADLPlot.pdf Reliability/Validity: A preliminary evaluation of retest stability was conducted with 104 subjects. Ninety

[40] https://outcometracker.org/library/Q-LES-Q-SF.pdf
[41] http://www.audiologyonline.com/articles/self-report-assessment-hearing-aid-931
[42] http://www.audiologyonline.com/articles/self-report-assessment-hearing-aid-931
[43] http://www.audiologyonline.com/articles/self-report-assessment-hearing-aid-931

Tool	Tool Information
	percent critical differences for the various scores ranged from 0.9 to 2.0 score intervals on a 7 point scale.[44]
	Availability of Benchmark Data/Clinical Indicator: Benchmark(s): Psychometric characteristics of the items were found to be very similar to those reported previously. Thus, the internal validity of the instrument was strongly supported. The assumption that the SADL quantifies satisfaction by assessing its components was evaluated by examining the relationship between SADL scores and scores on a traditional single-item satisfaction measure. A logical and statistically significant relationship was seen between the two measures, thereby supporting the construct validity of both types of data. For private-pay clients, satisfaction scores were very similar to the interim norms published by Cox and Alexander (1999). However, clients whose hearing aids were partly or fully purchased by insurance or benefits programs tended to be more satisfied than interim norms for third-party pay clients derived 5 yrs. ago. For most types of clients, there was a tendency toward more satisfaction in the Negative Features subscale than observed in our previous research.[45] Cox, RM and Alexander, GC. "Measuring satisfaction with amplification in daily life: The SADL Scale", Ear and Hearing, 20: 306-320 (1999). Cox, RM and Alexander, GC. "Validation of the SADL Questionnaire", Ear and Hearing, 22:151-160, 2001
	Additional Notes/Links: Link to survey: http://www.harlmemphis.org/files/4513/5412/2821/SADL15.pdf Distributor: Hearing Aid Research Lab (HARL) at the University of Memphis Original Publication Date: 1999 Availability: Online Copyright: 1999 Cost: Free, but the SADL scoring software can be purchased from the AUSP Software Group for $30.
McGill Quality of Life (MQOL) Index **IT-10.1.e**	**Summary**: Purpose: The McGill Quality of Life Questionnaire (MQOL) was developed by Dr. Robin Cohen and Dr. Balfour Mount of the Division of Palliative Care, Dept. of Oncology, McGill University because existing quality of life questionnaires were not appropriate or valid for use with the terminally ill. An instrument is needed that is valid when used in the setting of any type of terminal illness, from the time of diagnosis to death. MQOL is intended to meet this need. Overview: MQOL has been designed to measure subjective well-being, that is, the patient's experienced quality of life. It may be used in conjunction with other outcome measures when additional health-related outcome variables are of concern. Scores: *POSSIBLE SCORES*

[44] http://www.ncbi.nlm.nih.gov/pubmed/10466567
[45] http://www.ncbi.nlm.nih.gov/pubmed/11324844

Tool	Tool Information
	All MQOL items, MQOL sub measure scores, and MQOL Total Score have a possible range from '0' to '10'. In order for '0' to always indicate the worst situation and '10' the best situation; the following items must have the scores transposed prior to calculating the subscale and Total scores or data analysis. *Prior to calculating MQOL scores or data analysis, transpose the scores for* *Items 1, 2, 3, 5, 6, 7, and 8 by subtracting the raw score from 10 for each subject.* *MQOL SUB-MEASURES* There are 5 MQOL sub measures: Physical Symptoms; Physical Well-being; Psychological; Existential; and Support. They are scored as follows. • *Physical Symptoms.* This is a three-item scale. The score is the mean of the scores for Items 1, 2, and 3 (transposed). • *Physical Well-being.* This is a single-item measure. The score is the score for Item 4. • *Psychological.* This is a four-item scale. The score is the mean of the scores for Items 5, 6, 7, and 8 (all four transposed). • *Existential.* This is a six-item scale. The score is the mean of the scores for Items 9, 10, 11, 12, 13, and 14. • *Support.* This is a two-item scale. The score is the mean of the scores for Items 15 and 16. *MQOL Total Score* The MQOL Total score is the mean of the 5 sub-measure scores. <u>Administration Time:</u> 10-30minutes <u>Format:</u> MQOL comprises five sub-measures relating to: Physical Symptoms; Physical Well-being; Psychological Well-being; Existential Well-being; and Support. MQOL scores reflect subjective well-being in each domain but do not identify the contributing variables. Central goals in MQOL design included brevity and generalizability.
	Clinical/Quality information: Initial MQOL questions were chosen based on a literature review, clinical experience, and the results obtained from a preliminary longitudinal study of quality of life in patients receiving palliative care (n=50). This study used established instruments: the Functional Living Index - Cancer (FLIC) (Schipper et al, *Journal of Clinical Oncology*, 1984), Purpose-in-Life Questionnaire (PIL) (Crumbaugh and Maholick, *Journal of Clinical Psychology*, 1968), and Edmonton Symptom Assessment System (Bruera et al, *Journal of Palliative Care*, 1991), supplemented by additional questions of our own where required for conceptual completeness. MQOL development by us to date has included the following studies: a pilot study involving palliative care patients (n=40)(Cohen et al, *Palliative Medicine*, 1995); a multi-center study, also featuring palliative care subjects (n=150)(Cohen et al, *Palliative Medicine*, 1997); a study of people with cancer at all phases of the disease trajectory (n=247)(Cohen et al, *Cancer*, 1996*)*; a study of people living with HIV/AIDS (n=107)(Cohen et al, *AIDS*, 1996); a study of responsiveness and test-retest reliability (Cohen and Mount, *Cancer*, 2000); and a study of changes in quality of life during the first week of admission to palliative care units (Cohen et al. *Palliative Medicine*, 2001). These studies have enabled a series of evolutionary MQOL modifications to enhance acceptability, validity, and reliability. MQOL includes questions based on those in existing instruments: the FLIC (MQOL items 7 and 8), PIL (MQOL items 9-12), and the Missoula-Vitas Quality of Life Index (items 13-15). Please ensure that those with whom you discuss the MQOL are made aware of the origin of these questions. This study was carried out in eight palliative care services in four Canadian cities. A revised version of The McGill Quality of Life Questionnaire (MQOL) is compared to a single-item scale measuring overall quality of life (SIS), and the self-administered version of the Spitzer Quality of Life Index (SA-QLI), to

<table>
<tr><th>Tool</th><th>Tool Information</th></tr>
<tr><td></td><td>obtain evidence of validity. MQOL total score predicts SIS better than does SA-QLI, although much of the variance remains to be explained. The results of principal components analysis of data using this revised version of MQOL are similar to those from previous MQOL studies with different patient populations. The MQOL subscales. constructed on the basis of principal components analysis, demonstrate acceptable internal consistency reliability. The MQOL measures reflecting physical well-being and existential well-being are important for predicting SIS. Validity of the MQOL. Cohen et al. Pall Med (1997)</td></tr>
<tr><td></td><td>Availability of Benchmark Data/Clinical Indicator:

Benchmark(s):
Means and Standard Deviations for MQOL Scores

<table>
<tr><th>X̄ (SD)</th><th>MQOL total</th><th>Physical well-being</th><th>Physical symptoms</th><th>Psychologic</th><th>Existential</th><th>Support</th><th>MQOL SIS</th></tr>
<tr><td>All subjects</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Good day (n = 63–64)</td><td>7.9 (1.3)</td><td>7.8 (1.8)</td><td>6.8 (2.2)</td><td>8.2 (1.6)</td><td>8.1 (1.2)</td><td>8.6 (1.5)</td><td>8.0 (1.7)</td></tr>
<tr><td>Average day (n = 80)</td><td>6.8 (1.2)</td><td>6.1 (1.8)</td><td>5.4 (2.3)</td><td>7.0 (2.0)</td><td>7.2 (1.3)</td><td>8.5 (1.5)</td><td>6.2 (1.7)</td></tr>
<tr><td>Bad day (n = 66)</td><td>5.3 (1.1)</td><td>3.3 (1.9)</td><td>3.5 (1.9)</td><td>5.3 (2.1)</td><td>6.3 (1.7)</td><td>7.8 (2.0)</td><td>3.3 (1.7)</td></tr>
<tr><td>Oncology Day Centre</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Good day (n = 41)</td><td>8.0 (1.3)</td><td>7.9 (1.8)</td><td>7.1 (2.3)</td><td>8.2 (1.7)</td><td>8.2 (1.1)</td><td>8.6 (1.5)</td><td>8.0 (1.6)</td></tr>
<tr><td>Average day (n = 44)</td><td>7.0 (1.2)</td><td>6.5 (1.5)</td><td>5.8 (2.4)</td><td>6.8 (1.9)</td><td>7.3 (1.3)</td><td>8.4 (1.5)</td><td>6.2 (1.4)</td></tr>
<tr><td>Bad day (n = 34)</td><td>5.3 (1.3)</td><td>3.4 (2.1)</td><td>3.9 (2.2)</td><td>5.0 (2.0)</td><td>6.4 (1.8)</td><td>7.6 (2.1)</td><td>3.5 (1.6)</td></tr>
<tr><td>Palliative home care</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Good day (n = 22–23)</td><td>7.7 (1.1)</td><td>7.7 (1.7)</td><td>6.2 (1.9)</td><td>8.3 (1.9)</td><td>7.8 (1.3)</td><td>8.7 (1.5)</td><td>7.8 (1.9)</td></tr>
<tr><td>Average day (n = 36)</td><td>6.7 (1.3)</td><td>5.7 (2.1)</td><td>4.8 (2.1)</td><td>7.2 (2.1)</td><td>7.1 (1.3)</td><td>8.6 (1.4)</td><td>6.1 (2.0)</td></tr>
<tr><td>Bad day (n = 32)</td><td>5.2 (1.0)</td><td>3.2 (1.7)</td><td>3.1 (1.5)</td><td>5.7 (2.2)</td><td>6.3 (1.6)</td><td>8.0 (1.9)</td><td>3.0 (1.7)</td></tr>
</table>
MQOL: McGill Quality of Life; SD: standard deviation; SIS: single-item scale.

Clinical Indicator: MQOL sub measure scores, and MQOL Total Score have a possible range from ‘0’ to ‘10’. In order for ‘0’ to always indicate the worst situation and ‘10’ the best situation.</td></tr>
<tr><td></td><td>Additional Notes/Links:

Link to survey: http://saph.med.sa/wp-content/uploads/2012/11/mcgill_qol.pdf

Original Publication Date: 1997

Copyright: 1997; Robin Cohen

Contact/Availability:
Contact the author at robin.cohen@mcgill.ca to complete the User's Agreement and obtain a copy of this tool.

Cost: Free</td></tr>
<tr><td>Palliative Care Outcome Scale (POS)[46]

IT-10.1.f</td><td>Summary:

Purpose: The Palliative Care Outcome Study (POS) is a tool to measure patients' physical symptoms, psychological, emotional and spiritual needs, and provision of information and support at the end of life.

Overview: POS was developed in 1999 for use with patients with advanced disease, and to improve</td></tr>
</table>

[46] http://pos-pal.org/maix/

Tool	Tool Information
	outcome measurement by evaluating many essential and important outcomes in palliative care. Scores: POS results can be used to calculate individual item scores and an overall profile score. The overall profile score is useful in understanding the overall experience and status of the patient and their needs and strengths at a specific point in time. The overall profile score is the sum of the scores from each of the 10 questions. The overall profile score can therefore range from zero to 40. Individual item scores are useful when tracking particular dimensions, for example pain or spiritual need. POS has 10 items which assess the following dimensions: physical, emotional, psychological, spiritual and provision of information and support. Administration time: POS is therefore a patient reported outcome measure when the patient version of POS is used. POS takes less than 10 minutes to complete by staff or patients. Format: POS currently consists of ten items which assess physical symptoms, emotional, psychological and spiritual needs, and provision of information and support resulting in individual item scores and overall profile scores. An additional question provides patients with the opportunity to list their main problem/s. A global network of researchers and clinicians continue to collaborate with the creator of POS, Professor Irene J Higginson, to ensure that POS remains an outcome measure of choice.
	Clinical/Quality information: Reliability: Test/re-test reliability was acceptable for seven items. Internal consistency was good (Cronbach's alpha = 0.65 (patients), 0.70 (staff)). Change over time was shown, but did not reach statistical significance. Validity: The measure demonstrated construct validity (Spearman rho = 0.43 to 0.80). The POS consists of two almost identical measures, one of which is completed by staff, the other by patients. Agreement between staff and patient ratings was found to be acceptable for eight out of 10 items at the first assessment. The POS has acceptable validity and reliability. It can be used to assess prospectively palliative care for patients with advanced cancer.[47] From the literature: Generally, individual POS item scores of zero or one require less clinical attention than items that score three or four. For example, if a patient scores a four for question one when rating their pain this means the patient is reporting pain that is overwhelming to them and hinders their ability to think of anything else. A score of zero indicates that the patient isn't affected at all by pain, and a score of one means they are slightly impacted upon by pain but not bothered by it. Changes in scores over time are important to detect as they may indicate disease progression or a change in perception, circumstances or priorities. For example, a pain score that changes from one to two may indicate that something has changed in the patient's profile to the degree that it is now impacting on their activities. A timely clinical response is required when items are scored with a three or four, or when scores change.
	Availability of Benchmark Data/Clinical Indicator: Benchmark(s): PCOC uses established standards of palliative care to develop and support a national benchmarking system that will contribute to improved outcomes. This national initiative allows for the collection,

[47] http://www.ncbi.nlm.nih.gov/pubmed/10847883

Tool	Tool Information
	analysis and reporting of large sets of outcome measurement data. Information about PCOC and examples of their reports can be accessed via their website at http://chsd.uow.edu.au/pcoc/ Clinical Indicator[48]: POS is designed to be responsive to change. It can detect clinically important changes over time that are related to the course of the disease or to an intervention, such as symptom management. What patient says is important to them is a vital consideration when interpreting a POS score. Patients' views can be gained through an informal discussion about the scores, and probes such as "Tell me more about what that score means to you". Discussions with patients about their own priorities and what they perceive to be making a difference to them is helpful. In clinical practice, patient-centered clinical reasoning should be used when interpreting POS scores. Clinical supervision and regular review of POS data aids patient-centered care. Generally, individual POS item scores of zero or one require less clinical attention than items that score three or four. For example, if a patient scores a four for question one when rating their pain this means the patient is reporting pain that is overwhelming to them and hinders their ability to think of anything else. A score of zero indicates that the patient isn't affected at all by pain, and a score of one means they are slightly impacted upon by pain but not bothered by it. Changes in scores over time are important to detect as they may indicate disease progression or a change in perception, circumstances or priorities. For example, a pain score that changes from one to two may indicate that something has changed in the patient's profile to the degree that it is now impacting on their activities. A timely clinical response is required when items are scored with a three or four, or when scores change.
	Additional Notes/Links: Link to Survey: http://pos-pal.org/maix/pos-in-english.php Distributor: Cicely Saunders Institute Versions and Languages: There are currently two versions of POS: POS v1 and POS v2. The main difference between them is on question 7, which asks if life is worthwhile in v1 and about depression in v2. POS-S is an additional module for symptoms which can be used alongside POS to capture more information about symptoms for a range of conditions such as cancer, heart failure and respiratory conditions. Extended versions have also been developed for use with patients living with multiple sclerosis, Parkinson's disease and end-stage renal disease. Original Publication Date: 1999 Web Access: http://pos-pal.org/maix/ Contact/Availability: palliativecare@pos-pal.org Cost: Free, but must register with pos-pal.org
Functional Assessment of Cancer Therapy[49] (FACT-G & FACT-L) **IT-10.1.h**	**Summary**: Purpose: The FACT is a patient-assessed measurement system comprising a core component (the FACTG) that covers general aspects of quality of life (QoL), plus a range of optional condition specific subscales. The system is intended for recording outcomes in clinical trials; it focuses mainly on cancer therapy, but is potentially applicable to other conditions. Overview: The FACT was intended to be broadly applicable yet sensitive to change following treatment; to achieve this, the FACT is conceived as a

[48] http://pos-pal.org/maix/how-to-interpret.php
[49] http://www.facit.org/FACITOrg/Questionnaires

Tool	Tool Information
	measurement system comprising a core set of items applicable to all types of cancer plus disease-specific supplements. The core items are presented as a profile of scores, rather than as an index, because different therapies might be expected to have a differential impact on different dimensions (3, p201). Although developed for use in patients with cancer, the FACT has been tested on those with HIV and rheumatoid arthritis, as well as in the general population (D. Cella, personal communication, 2004). Scores: The answer categories use a Likert format (0 "not at all" to 4 "very much") and allow for administration by telephone (2, p577). A total score sums the subscale scores. A comparison of several ways to adjust for missing data suggested that the best method is to replace a missing item by the mean score on its subscale, as long as at least 50. Administration time: The FACT-G can be completed in five minutes (2, p575), although a questionnaire version took an average of 13.5 minutes to administer in a population with low educational levels (5). Format: The FACT-G forms the core component in a set of instruments originally developed for assessing QoL outcomes for patients with cancer, but it has broadened into an assessment system for Functional Assessment in Chronic Illness Therapy, or FACIT (see www.facit.org). The FACIT organization develops and distributes questionnaires, ensures their standardization, coordinates translations, and provides information on administration, scoring, and interpretation. The FACT-G, described here, contains generic assessments to which supplementary disease specific modules may be added. Items were generated through interviews with patients and oncologists, and the process of test development is described by Cella et al. (2; 4). Based on the results of factor analyses, the items were grouped into four subscales early in the development process: physical well-being (PWB, seven items, score range, 0–28); social and family well-being (SWB, seven items, 0–28); emotional well-being (EWB, six items; 0–24) and functional well-being (FWB, seven items, 0–28).
	Clinical/Quality information: Norms: The FACT-G normative sample data for the general U.S. adult population reference group were collected by Knowledge Networks (KN; Menlo Park, CA), a marketing information and decision support system. KN drew a random sample of 1,400 people, age 18 years and older, from more than 100,000 individuals who were members of an Internet-based survey panel. According to KN, the panel was a demographically representative sample of the general U.S. adult population. Members of the survey panel responded to one survey per month in exchange for free installation of WebTV Internet service. The FACT-G (Version 4)was one such survey thatwas presented electronically to the panel members who completed the survey in their homes (Cella et al., 2003). Reliability: Retest reliability was 0.92 for the FACT-G total score after three to seven days; coefficients for the subscales included 0.88 (PWB), 0.84 (FWB), 0.82 (EWB and SWB), and 0.82 for RWD (4, p208). Validity: During the initial test development, Cella et al. reported results of a factor analysis that produced six dimensions, of which two were merged. FACT-G sub scores distinguished significantly between patients in different stages of their illness, as defined by the NCI criteria (2, Table 4). The physical, functional, and emotional subscales distinguished significantly between patients classified by the ECOG performance rating in terms of whether they changed over time (2, p575). The FACT's physical, functional, and emotional scales, and the overall score, corresponded to independent patient-ratings of meaningful change (4, Table 3; 15). Effect sizes have been reported (15, Tables 5–9). In three samples of patients, the PWB, FWB and overall scores were consistently related to changes in patient condition and

Tool	Tool Information
	severity of illness (11, Tables 4–7).
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): Evidence-based interpretation guidelines for the Functional Assessment of Cancer Therapy-General (FACT-G), a cancer-specific health-related quality of life (HRQOL) instrument, from a range of clinically relevant anchors, incorporating expert judgment about clinical significance. www.dovepress.com/getfile.php?fileID=7764 Clinical Indicator: T-score conversion tables from the FACT-G raw score (US general population and cancer populations) can be found at http://www.facit.org/Default.aspx?PageID=5040800&EID=54463801&CID=7486468. Note: Must register with FACIT to get access to this data. US General Adult Population: Physical Well Being: -1σ= 17/18; mean = 22/23; 1σ= 28 Social/Family Wellbeing: -1σ= 12/13; mean = 19; 1σ= 26 Emotional Well Being: -1σ= 15; mean = 20; 1σ= > 24 Functional Well Being: -1σ= 11/12; mean = 18/19; 1σ= 25/26 Total FACT-G: -1σ= 62; mean = 80; 1σ= 98 Adult Cancer Population: Physical Well Being: -1σ= 15/16; mean = 21/22; 1σ= 27/28 Social/Family Wellbeing: -1σ= 17; mean = 22; 1σ= 27/28 Emotional Well Being: -1σ= 14/15; mean = 19; 1σ= 23 Functional Well Being: -1σ= 12; mean = 19; 1σ= 26 Total FACT-G: -1σ= 64; mean = 81; 1σ= 98
	Additional Notes/Links: Link to survey: The FACT-G and FACT-L can be downloaded as a WORD or PDF document: http://www.facit.org/FACITOrg/Questionnaires Distributor: FACIT Versions and Languages: http://www.facit.org/FACITOrg/Questionnaires Since the original description of the FACT-G in 1993, FACIT has developed numerous questionnaires and coordinated their translation. Besides cancer, subscales have been developed for fatigue, treatment satisfaction, spiritual well-being, and for HIV disease, multiple sclerosis, and other chronic conditions. FACT-G has been translated from English into over 30 languages, including Afrikaans, Bulgarian, Chinese (Taiwan/HongKong), Chinese (Mainland), Czech, Danish, Dutch German, Greek, French, Hebrew, Hungarian, Italian, Japanese, Norwegian, Polish, Portuguese, Russian, Tswana, Pedi, Zulu, Spanish, Swedish, and Thai. Original Publication Date: 1993 Copyright: Complete the FACIT User's Agreement in order to access and use this tool. Web Access: http://www.facit.org/FACITOrg/Questionnaires Cost: English versions are free; however, translations may require fees.

Tool	Tool Information
Missoula-VITAS Quality of Life Index (MVQOLI)[50] **IT-10.1.i**	**Summary**: Purpose: The Missoula-VITAS Quality of Life Index (MVQOLI) is an assessment instrument that gathers patient-reported information about quality of life during advanced illness. Overview: The MVQOLI asks patients about 5 dimensions or domains of quality of life: symptoms; function; interpersonal; well-being; and transcendence. (See definitions below.) The instrument is specifically designed to assess the patients personal experience in each of these dimensions, hence the MVQOLI items are constructed with highly subjective language and no scores appear on the version of the tool seen by patients. The tool seeks to describe the qualitative and subjective experience of quality of life in a way that can be quickly interpreted by professional caregivers. Scores: The MVQOLI scoring manual for the 15 and 25 item surveys can be found at http://www.dyingwell.org/downloads/MVQOLI/Appendix%208%20-%20MVQOLI%20Manual%20Scoring%20Procedure.pdf The MVQOLI items are scored as follows: Assessment: -2 to +2 Satisfaction: -4 to +4 Importance: 1 to 5 (Assessment + Satisfaction) X Importance = QOL in each dimension Reading Level: The newest versions of the tool included with this guide have been revised using simpler language and item formats to make it easier to use for both patients and staff. Format: There are two versions of the MVQOLI – 15 item and 25 item. The instrument was initially designed with 25 items. Clinicians reported that the tool was too long for some patients to complete.
	Clinical/Quality information: Reliability/ Validity: Using data from the original study of reliability and validity, a 15-item version was constructed that has a correlation coefficient of .93 with the 25- item version, indicating that little information is lost when only 15 items are used. From the literature: The MVQOLI has been used with palliative care and hospice patients in a variety of settings including hospice, hospital, home health, long-term care (including assisted living), outpatient palliative care, and pre-hospice programs. It is appropriate for any patient population facing advanced, chronic, progressive illness
	Availability of Benchmark Data: Clinical Indicator[51]: Interpreting the MVQOLI Patient Profile • Zero represents neutral as a rating of a dimension

[50] http://www.dyingwell.org/MVQOLI.htm
[51] http://www.dearborncountyhospital.org/dch/HomeHealth/MVQOLI/Section%205%20-%20Training/3-Independent%20Learning%20Module%20Content%20Pkt%2012-04.doc

Tool	Tool Information
	• Anything above zero is positively impacting a patient's quality of life and anything below zero is negatively impacting a patient's quality of life (see the symptom dimension above as a positive dimension and the function dimension is negative). • The larger the bar, the more important that dimension is to the patient (see the well-being dimension above). • If a bar is large and above the line that dimension is very important to the patient and is positive (well-being). • If a bar is small and below the line it is negatively impacting the quality of life but isn't too important (interpersonal). • The actual scores aren't as important as the profile which shows the dimension scores in relation to each other. A +5 might be a high score for one patient, while for another patient it might be relatively low. • When multiple MVQOLI surveys have been completed by a patient, the scores will be displayed next to each other to allow the team to see the difference in scores over time (see example in the case study on the next few pages).
	Additional Notes/Links: Link to survey: 15- and 25-item versions in English, Spanish, and Greek can be found at http://www.dyingwell.org/MVQOLI.htm Distributor: DyingWell.org Versions and Languages: English, Spanish, and Greek Original Publication Date:1997 Copyright: 1997-2013 Web Access: http://www.dyingwell.org/MVQOLI.htm Contact/Availability: Ira Byock, M.D. The Palliative Care Service, Missoula, MT PH: 406-728-8643 FAX: 406-728-4709 E-mail: IByock@aol.com Cost: Free, but must register with dyingwell.org
CDC Health-Related Quality of Life (HRQoL)[52] Measures **IT-10.1.j**	**Summary**: Purpose: To meet the need for a standard set of valid HRQOL measures that could be used in our national health surveillance system, a collaborative program was initiated in 1989 by the Division of Adult and Community Health (DACH) in the CDC's National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). This HRQOL surveillance program received its initial direction and guidance from several planning meetings that included representatives of state and local chronic disease and health promotion programs, relevant academic disciplines, and survey researchers. Overview: The concept of health-related quality of life (HRQOL) and its determinants have evolved since the 1980s to encompass those aspects of overall quality of life that can be clearly shown to affect health—either physical or mental.[3-6]

[52] http://www.cdc.gov/hrqol/concept.htm

Tool	Tool Information
	Format: [53]The standard 4-item set of Healthy Days core questions (CDC HRQOL– 4) has been in the State-based Behavioral Risk Factor Surveillance System (BRFSS) since 1993 (see BRFSS Website). Since 2000, the CDC HRQOL– 4 has been in the National Health and Nutrition Examination Survey (NHANES) for persons aged 12 and older. Since 2003, the CDC HRQOL– 4 has been in the Medicare Health Outcome Survey (HOS)—a NCQA HEDIS measure. Standard Activity Limitation and Healthy Days Symptoms modules have also been available since January 1995. When used together, these measures comprise the full CDC HRQOL–14 Measure. HRQoL has been proposed as a method to monitor health in the American public health plan, "Healthy People 2010."
	Clinical/Quality information: Reliability:[54] Retest reliability of the HRQoL surveillance measures is moderate to excellent, and internal validity is strong. Validity: http://www.cdc.gov/hrqol/properties/validity.htm From the literature: Several measures have been used to assess HRQOL and related concepts of functional status. Among them are the Medical Outcomes Study Short Forms (SF-12 and SF-36), the Sickness Impact Profile, and the Quality of Well-Being Scale. The SF-36 measures are now used by the Health Care Financing Administration (HCFA) and the National Committee for Quality Assurance's Health Plan Employer Data and Information Set (HEDIS 3.0) to help evaluate the quality of care in managed care plans and other health care applications. While these measures have been widely used and extensively validated in clinical settings and special population studies, their length often makes them impractical to use in population surveillance. Since 1993, CDC, states, and others have demonstrated the usefulness of HRQOL measures in identifying vulnerable population subgroups and in community health assessments (e.g., the federal Department of Health and Human Services' Community Health Status Indicators Project; the University of Wisconsin's Mobilizing Action Toward Community Health Project). Adding HRQOL indicators in community health assessment studies can offer health agencies outcomes that are meaningful to the broad community, identify population disparities in HRQOL, and help prioritize subgroups with unmet needs to improve community quality of life. Because the actions of many groups in a community may affect HRQOL, successful interventions and healthy public policies require active partnerships with multiple community members including the business community, departments of transportation, education, and public safety, health care communities and non-profit groups.
	Availability of Benchmark Data/Clinical Indicator: Benchmark(s): Stratified by state, gender, age and race: http://apps.nccd.cdc.gov/HRQOL/ Clinical Indicator[55]: Frequent mental distress is defined as having 14 or more mentally unhealthy days as measured by the CDC Healthy Days question: Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

[53] http://www.cdc.gov/hrqol/hrqol14_measure.htm
[54] http://www.cdc.gov/hrqol/measurement_properties/andresen1.htm
[55] http://www.cdc.gov/hrqol/faqs.htm

Tool	Tool Information
	The CDC reports there is no real standard set of categories for Frequent Mental Distress or Frequent Physical Distress other than the common use of the 14+ days grouping to indicate a substantial level of impairment. Providers/researchers are free to use any reasonable categorization scheme based on 1) the response distribution to the questions, 2) hypothesized or empirically established levels of severity, 3) or comparability with public domain prevalence data tables or published research. The CDC has not yet done a systematic review of the literature to document what categories have been used, but considers doing that in the future. Fewer groups provide larger sample sizes for testing significant differences and tend to show a more stable dose-response pattern. In general, if you include at least 50 responses per cell, this will provide narrower confidence intervals and more stable prevalence estimates over time.
	Additional Notes/Links: Link to survey: http://www.cdc.gov/hrqol/hrqol14_measure.htm Distributor: CDC Versions and Languages: English/Spanish Original Publication Date: 1993 Web Access: http://calmhsa.org/wp-content/uploads/2013/06/CDC-HRQOL-14.pdf Cost: Free
Child Health Questionnaire Child Form 87 (CHQ-CF-87) and Parent Form 28/50 (CHQ-PF-28/50)[56] **IT-10.1.k**	**Summary**: Purpose: The CHQ assesses a child's physical, emotional, and social well-being from the perspective of a parent or guardian {CHQ-PF50 and PF-28 (short form)} or, in some instances, the child directly (CHQ-CF87, for children ten years of age and older). The child/adolescent's general health, change in health, physical functioning, bodily pain/discomfort, limitations in school work and activities with friends due to physical problems or emotional/behavioral difficulties, behavior, mental health, and self-esteem are measured. Overview: The Child Health Questionnaire™ (CHQ) is a family of generic quality of life instruments that have been designed and normed for children 5-to-18 years of age. The CHQ measures 14 unique physical and psychosocial concepts. The parent form is available in 2 lengths - 50 or 28 items. Scores can be analyzed separately, the CHQ Profile Scores, or combined to derive an overall physical and psychosocial score, the CHQ Summary Scores. In April 2008 HealthActCHQ released the first-ever electronic CHQ Scoring and Interpretation Manual. The 212-page Manual provides information about the conceptual framework and development of the CHQ, the proprietary scoring algorithms, norms and rules for interpretation. The interactive CD-ROM features hyperlinks in both the Table of Contents and the Appendix of Tables for smooth navigation. Users can also click on URLs within the Manual to access the latest updates on translations and the online Bibliography at the HealthActCHQ website. The CHQ surveys and translations are made available upon approval of registration and payment. Scores: Overall means for the individual CHQ scales and items can be derived using a simple summated rating approach. This method yields a profile for each of the 14 health concepts. In addition, the individual scale scores can be aggregated to derive two summary component scores of physical and psychosocial health. Scores transformed to 0 - 100 scale. A higher score always indicates more favorable ratings of health and well-being.

[56] http://www.healthactchq.com/chq.php

Tool	Tool Information
	Administration time: The CHQ-PF28 is estimated at 5-to-10 minutes, the CHQ-PF50 requires 10-to-15 minutes, and the CHQ-CF87 completion times can vary from 16-25 minutes. Format: There are two parent lengths—50 and 28 items. The 28-item form is best for large population studies where many children will be evaluated. The most common parent-completed form is the CHQ-PF50. The child-completed form is the CHQ-CF87, consisting of 87 items. A child short-form is not yet available.
	Clinical/Quality information: Norms: Normed for children from 5-to-18 years of age. Reliability: For CHQ-PF50 - median reliability was 0.84 (US sample) For CHQ-PF28 - median reliability was 0.75 For CHQ-CF87 - range between 0.73 to 0.97 for the different scales[57] Child Health Questionnaire - Parent Form 50 • Test-retest / reproducibility: ICC - 0.37 to 0.84 • Internal consistency: Cronbach's alpha - 0.53 to 0.96 Range of correlations across 4 countries. Alpha reported for each subscale in original citation. All reported within acceptable range. Validity: Construct Validity – Item discriminant validity examined by taking the number of item-scale correlations > 2 SE divided by the total number of tests conducted. Discriminant validity results generally very high, see citation for details. Content & Face Validity – Item development based on conceptual framework. From the literature: Qual Life Res. 1998 Jul;7(5):433-45. **Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results.** Landgraf JM, Maunsell E, Speechley KN, Bullinger M, Campbell S, Abetz L, Ware JE. **Source** HealthAct, Boston MA, USA. **Abstract** Using emerging international guidelines, stringent procedures were used to develop and evaluate Canadian-French, German and UK translations/adaptions of the 50 item, parent-completed Child Health Questionnaire (CHQ-PF50). Multitrait analysis was used to evaluate the convergent and discriminant validity of the hypothesized item sets across countries relative to the results obtained for a representative sample of children in the US. Cronbach's alpha coefficient was used to estimate the internal consistency reliability for each of the health scales. Floor and ceiling effects were also examined. Seventy-nine percent of all the item-scale correlations achieved acceptable internal consistency (0.40 or higher). The tests of the item convergent and discriminant validity were successful at least 87% of the time across all scales and countries. Equal item variance was observed 90% of the time across all countries. The reliability coefficients ranged from a low of 0.43 (parental time impact, Canadian English) to a high of 0.97 (physical functioning index, Canadian French) across all scales (median 0.80). Negligible floor effects were observed across countries. Noteworthy ceiling effects were observed, as expected, for the hypothesized physical scales (mean effect 73%). Conversely, fewer ceiling effects were observed for the

[57] http://brownprojects.wustl.edu/CMHSRMeasures/e18.html

Tool	Tool Information
	psychosocial scales (range 3-17% behavior-parental emotional impact). The item-scaling results obtained in these pilot studies support the psychometric properties of the American-English CHQ-PF50 and its respective translations.
	Availability of Benchmark Data/Clinical Indicator: Clinical Indicator: Scores transformed to 0 - 100 scale. A higher score always indicates more favorable ratings of health and well-being. Broad (high vs. low score by scale domains) scoring interpretation of the CHQ-PF-50 can be found at http://pediatrics.aappublications.org/content/124/5/e942.full.pdf. The link also provides information on the domain z-score and raw score differences between the CHQ-CF and CHQ-PF.
	Additional Notes/Links: Link to survey**:** Questions for all three surveys can be found at: http://www.healthact.com/pdf/chq.PDF Distributor: HealthActCHQ Inc. Versions and Languages: American-Spanish, Canadian-French, Finnish, French, German, Dutch, Italian, Greek, Honduran, Mexican, Norwegian, Portuguese, Swedish Original Publication Date: 1996 Copyright: 2013 Web Access: http://www.healthact.com Contact/Availability: HealthAct Two International Place 800 Boylston Street, 16th Floor Boston, MA 02199 Phone: (857) 453-6665 Fax: (857) 4536501 Cost: CHQ-PF 50: All users are asked to purchase a copy of the User Manual[1] ($250). No additional fee is required if used for research purposes
Family Experiences Interview Schedule (FEIS)[58] **IT-10.1.l**	**Summary**: Purpose: The FEIS measures family experiences (including positive aspects of caregiving, special regards, benefits and gratifications) of caregiving to patients aged 18-64 with serious mental illness such as schizophrenia or bipolar disorders. Overview: The FEIS is a revision of the Family Burden Interview Schedule (FBIS), which was developed in the tradition of a line of research on the family experience going back to the early days of deinstitutionalization in which the family experience of caregiving was conceptualized as "burden." More recently there has been interest in expanding the conceptualization and measurement of family experiences to include the positive aspects of caregiving. Thus the FEIS also includes measures of the special rewards, benefits and gratifications that may derive from caregiving.

[58] http://www.hsri.org/files/uploads/publications/PN_5.pdf

Tool	Tool Information
	Sections Include: 180-item instrument. Financial Expenditures: ADL Care (Objective), ADL Care (Subjective), Supervision (Objective), Supervision (Subjective); Impact of Daily Routines Negative Aspects of the Family Experience: Worry, and Displeasure Positive Aspects of the Family Experience: Benefits, and Gratifications Administration time: 60 minutes to administer in person at baseline and 30 minutes to administer as a follow up by telephone. Format: The FBIS/SF takes a multidimensional approach and distinguishes different aspects of burden from one another. There are 65 items, which include five modules related to the negative aspects of caregiving: (1) assistance with the activities of daily life; (2) supervision of bothersome or troublesome behaviors; (3) impact on daily routines; (4) financial expenditures; and (5) worry on daily routines.
	Clinical/Quality information: Factors and Norms: Caregivers of patients aged 18-64 with serious mental illness such as schizophrenia or bipolar disorders. The FEIS has been used with a variety of client populations, including but not limited to schizophrenia and the affective disorders. It has been administered to a variety of caregivers, including primary care givers as well as other family members. Virtually all types of family relations have been interviewed using the earlier FBIS, including some relationships that while not linked by blood or marriage are nonetheless considered by clients to be "like family." Reliability/Validity: Benefits (Cronbach's alpha = .819) Gratifications (Cronbach's alpha = .87) Objective ADL (Cronbach's alpha = .777) Subjective ADL (Cronbach's alpha = .741) Objective Control (Cronbach's alpha = .65) Subjective control/supervision (Cronbach's alpha = .638) Impact on Daily Routines (Cronbach-'s alpha = .568) Attitudes Towards Professionals (Cronbach's alpha = .815) Worry (Cronbach's alpha = .891) Displeasure (Cronbach's alpha= .852) Stigma (Cronbach.' s alpha = .851) From the literature: The Toolkit for Evaluating Family Experiences with Severe Mental Illness is designed to provide technical assistance to mental health services researchers who wish to include family outcomes when designing evaluation studies. The Toolkit is organized around a particular instrument, The Family Experiences Interview Schedule (FEIS). The FEIS is a revision of the Family Burden Interview Schedule (FBIS), which was developed in the tradition of a line of research on the family experience going back to the early days of deinstitutionalization in which the family experience of caregiving was conceptualized as "burden." More recently there has been interest in expanding the conceptualization and measurement of family experiences to include the positive aspects of caregiving.
	Availability of Benchmark Data/Clinical Indicator: Benchmark(s): Clinical Indicator:
	Additional Notes/Links: Link to survey: http://www.hsri.org/files/uploads/publications/PN_5.pdf (pages 46-89) Distributor: Human Services Research Institute at the University of Massachusetts

Tool	Tool Information
	Original Publication Date: 1993 Contact/Availability: Richard Tessler, Ph.D. & Gail Gamache, Ph.D. Social and Demographic Research Institute Machmer Hall University of Massachusetts Amherst, MA 01003-4830 Cost: Free
Hearing Handicap Inventory for Adults/Elderly -Screening (HHIE-S) **IT-10.1.m**	**Summary**[59] Purpose: The HHIE-S was designed to assess how an individual perceives the social and emotional effects of hearing loss. Overview: The HHIE-S was designed to be used with non-institutionalized older adults in a variety of clinical and community settings. The HHIE-S should be administered yearly during annual well examinations. Individuals should be referred to a hearing specialist for further assessment if the HHIE-S score is > 10 points. Scores: The higher the HHIE-S score, the greater the handicapping effect of a hearing impairment. Possible scores range from 0 (no handicap) to 40 (maximum handicap). Audiologic referral is recommended for individuals scoring 10 points or higher on the HHIE-S. The answers to each question are in a 'Yes', 'No', and 'Sometimes' format. A 'Yes' scores a 4, a Sometimes' scores a 2, and a 'No' scores a 0.[60] Administration Time[61]: 2 minutes Format: 10-item questionnaire. It is usually administered in a face-to-face interview. However, time constraints or a severe-to-profound hearing loss may preclude a face-to-face interview, in which case the HHIE-S may be administered by having the individual do a paper-and-pencil self-report.
	Clinical / Quality Information Reliability[62]: Internal consistency reliability (Cronbach's alpha) was reported as 0.87 in a sample of 162 older adults presenting to a speech and hearing center. Test-retest reliability was reported at 0.84 (P< .0001). Sensitivity when compared to audiogram-defined hearing loss has been reported as 63-80% with a specificity of 67-77% (cutoff score > 10). A cutoff score of > 24 yielded 88-98% specificity with a reduced sensitivity (24-42%). Validity[63]: Bess, Lichtenstein, Logan, & Burger (1989) and Bess, Lichtenstein, & Logan (1991) compared scores on the HHIE-S to scores on the SIP: subjects with severe hearing handicaps (HHIE-S score exceeding 24 points) demonstrated greater effects of hearing impairment in the physical and psychosocial content domains than those subjects classified as having no handicap on the HHIE-S. Additionally, the authors reported that each 10 dB increase in hearing loss was associated with a 2.8 point increase in the SIP physical dimension score. The same 10 dB increase was associated with 2.0 and 1.3 point increases in

[59] http://consultgerirn.org/uploads/File/trythis/try_this_12.pdf
[60] http://teachhealthk-12.uthscsa.edu/curriculum/vision-hearing/hearing08e-handicap.asp
[61] http://teachhealthk-12.uthscsa.edu/curriculum/vision-hearing/hearing08e-handicap.asp
[62] http://consultgerirn.org/uploads/File/trythis/try_this_12.pdf
[63] http://www.asha.org/policy/GL1997-00199.htm#AP3

Tool	Tool Information
	the psychosocial and overall scores of the HHIE-S, respectively. From the literature: Median positive likelihood ratios (LRs) among the screening tests at greater than 25 or 30 dB were in the range of 3.0 to 5.1 for single-question screening, HHIE-S, and whispered voice test at 2 feet (in ascending order). Negative LRs ranged from 0.03 to 0.52 for whispered voice test, single-question screening, and HHIE-S. The median positive LR at greater than 40 dB for the AudioScope audiometer (Welch Allyn, Skaneateles Falls, New York) was 5.8 (range, 1.7 to 4.9), and the median negative LR was 0.05 (range, 0.03 to 0.08) (3, 4). Finger rub and watch tick tests had substantially stronger positive LRs (10 and 70, respectively) compared with other screening methods, but they were only evaluated in a single study (10) and the CIs were very wide (2.6 to 43 and 4.4 to 1120, respectively). Negative LRs for the finger rub and watch tick tests were 0.75 (95% CI, 0.68 to 0.84) and 0.57 (CI, 0.46 to 0.66), respectively.[64]
	Availability of Benchmarks Clinical Indicator: Audiologic referral is recommended for individuals scoring 10 points or higher on the HHIE-S. A response of ‘‘yes’’ is given 4 points, ‘‘sometimes’’ is given 2 points, and ‘‘no’’ is given 0 points. HHIE-S scores range from 0 to 40, with a score of 8 or higher indicative of at least a mild hearing handicap (Ventry & Weinstein, 1983).[65]
	Additional Notes/Links Link to Survey: English: http://teachhealthk-12.uthscsa.edu/curriculum/vision-hearing/pa06pdf/0608E-eng.pdf Spanish: http://teachhealthk-12.uthscsa.edu/curriculum/vision-hearing/pa06pdf/0608-span.pdf Versions/Languages: English and Spanish Original Publication Date: 1986; Ventry and Weinstein
Katz Activities of Daily Living (ADLs) Scale **IT-10.2.b**	**Summary**: Purpose[66]: Assesses functional status as a measurement of the client’s ability to perform activities of daily living independently. Clinicians typically use the tool to detect problems in performing activities of daily living and to plan care accordingly. Overview[67]: The instrument is most effectively used among older adults in a variety of care settings, when baseline measurements, taken when the client is well, are compared to periodic or subsequent measures. The Index ranks adequacy of performance in the six functions of bathing, dressing, toileting, transferring, continence, and feeding. The Katz ADL Index assesses basic activities of daily living. It does not assess more advanced activities of daily living. Katz developed another scale for instrumental activities of daily living such as heavy housework, shopping, managing finances and telephoning. Although the Katz ADL Index is sensitive to changes in declining health status, it is limited in its ability to measure small increments of change seen in the rehabilitation of older adults. A full comprehensive geriatric assessment should follow when appropriate. The Katz ADL Index is very useful in creating a common language about patient function for all practitioners involved in overall care planning and

[64] http://www.uspreventiveservicestaskforce.org/uspstf11/adulthearing/adulthearrs.htm
[65] http://www.clas.ufl.edu/users/mcolburn/Web-links/SPA4321/Impact%20of%20HL%20on%20QOL.pdf
[66] http://consultgerirn.org/uploads/File/trythis/try_this_2.pdf
[67] http://consultgerirn.org/uploads/File/trythis/try_this_2.pdf

Tool	Tool Information
	discharge planning. Scores[68]: Clients are scored yes/no for independence in each of the six functions. A score of 6 indicates full function, 4 indicates moderate impairment, and 2 or less indicates severe functional impairment. Administration time:
	Clinical/Quality information: Reliability/Validity[69]: Reliability: The Katz index has shown good reliability, as evidenced by reliability coefficients ranging from 0.87 to 0.94. Validity: The Katz index has demonstrated accuracy in predicting functional outcomes over time among older adults in short-term care, hospitalized patients, and patients who have had a stroke.1, 3, 17 Hamrin and Lindmark reported convergent (or concurrent) validity as high, with a correlation of 0.95 between the Activity index and the Katz index.
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): Graf, C. (2006). Functional decline in hospitalized older adults. AJN, 106(1), 58-67. Hartigan, I. (2007). A comparative review of the Katz ADL and the Barthel Index in assessing the activities of daily living of older people. International Journal of Older People Nursing, 2(3), 204-212. Katz, S. (1983). Assessing self-maintenance: Activities of daily living, mobility and instrumental activities of daily living. JAGS, 31(12), 721-726. Katz, S., Down, T.D., Cash, H.R., & Grotz, R.C. (1970) Progress in the development of the index of ADL. The Gerontologist, 10(1), 20-30. Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., & Jaffe, M.W. (1963). Studies of illness in the aged: The index of ADL: A standardized measure of biological and psychosocial function. JAMA, 185(12), 914-919. Kresevic, D.M. (2012). Assessment of physical function. In M. Boltz, E. Capezuti, T.T. Fulmer, & D. Zwicker (Eds.), A. O'Meara (Managing Ed.), Evidence-based geriatric nursing protocols for best practice (4th ed., pp. 89-103). NY: Springer Publishing Company, LLC. Clinical Indicator**:** Clients are scored yes/no for independence in each of the six functions. A score of 6 indicates full function, a score of 4 indicates moderate impairment, and 2 or less indicates severe functional impairment.
	Additional Notes/Links: Link to survey: http://www.tuft-healthplans.org/providers/pdf/katz_adl.pdf Distributor: The Hartford Institute for Geriatric Nursing Original Publication Date: 1963 Copyright: The Gerontological Society of America

[68] http://consultgerirn.org/uploads/File/trythis/try_this_2.pdf
[69] http://www.tuft-healthplans.org/providers/pdf/katz_adl.pdf

Tool	Tool Information
Supports Intensity Scale (SIS) **IT-10.2.c**	**Summary** Purpose[70]: The Supports Intensity Scale (SIS) is a tool designed to measure the relative intensity of support each person with developmental disabilities needs to fully participate in community life. The SIS is intended to be used in conjunction with person-centered planning processes to assist planning teams in developing individual support plans that are responsive to the needs and choices of persons with disabilities. Overview[71]: The SIS is comprised of 87 questions across the following subscales: home living, community living, lifelong learning, employment, health and safety, and social. Additionally, the Supplemental Protection and Advocacy Scale measures 8 activities, but is not used to score total Support Intensity Score. The Exceptional Medical and Behavioral Support Needs section measure supports in 15 medical conditions and 13 problem behaviors commonly associated with intellectual disabilities. Additional information can be found at: http://buntinx.org/yahoo_site_admin/assets/docs/SISAdministrationScoringProcedures1.3020358.pdf Scores: A descriptive explanation can be found at: http://buntinx.org/yahoo_site_admin/assets/docs/SISAdministrationScoringProcedures1.3020358.pdf Administration Time: One hour (however, having the patient's support team available can result in 2.5-3 hour administration times). The SIS should be administered by a professional who has completed a 4-year degree program and is working in the field of human services (for example, case manager, psychologist, social worker). Format: A paper and pencil-based test consisting of an 8-page Interview and profile form. Comes with accompanying 128-page User's Manual. SISOnline is an advanced web application system that enables you to score the Supports Intensity Scale online through a standard web browser. The system allows access to a variety of reports and statistics, and maintains a database of historical information and more.
	Clinical / Quality Information Norms/Factors[72]: The normative sample for the SIS standardized scores was made up of 1,306 people with developmental disabilities from 33 states. The SIS normalized standard scale scores have means of 10 and standard deviations of 3, and the composite score is standardized with a mean of 100 and standard deviation of 15. Reliability[73]: SIS has a .87 inter-rater reliability coefficient From the literature: Kuppens S, Bossaert G, Buntinx W, et al. Factorial validity of the Supports Intensity Scale (SIS). Am J Intellect Dev Disabil. 2010 Jul;115(4):327-39. http://aaidd.org/sis/white-papers
	Availability of Benchmark Data/Clinical Indicator[74] Benchmark(s): SIS Mean Scale performance scores by residence type can be found at

[70] http://www.ncdhhs.gov/mhddsas/providers/CAPMRDD/SIS/
[71] http://aaidd.org/docs/default-source/sis-docs/latestsispresentation.pdf?sfvrsn=2
[72] http://www.dshs.wa.gov/pdf/ms/rda/research/5/36.pdf
[73] http://aaidd.org/docs/default-source/sis-docs/latestsispresentation.pdf?sfvrsn=2
[74] http://www.dshs.wa.gov/pdf/ms/rda/research/5/36.pdf

Tool	Tool Information
	http://www.dshs.wa.gov/pdf/ms/rda/research/5/36.pdf Clinical Indicator**:** According to the SIS manual, the accepted criteria for determining exceptional need are: a total score on a support needs scale greater than 5 or at least one item with a response of "extensive support needed" (scored as "2").
	Additional Notes/Links Link to survey: http://aaidd.org/docs/default-source/sis-docs/sis-interview-and-profile-form-(do-not-copy).pdf?sfvrsn=2 Distributor: American Association on Intellectual and Development Disabilities (AAIDD) Versions/Languages: English and French. The SIS-A (Adult) and SIS-C (Child) versions are under development. Original Publication Date: English (1/2004) and French (1/2008) Web Access: http://aaidd.org/sis/product-information Fax: 1 (301) 208-9789 Contact/Availability: 1 (301) 604-1340 Cost: http://aaidd.org/sis/order
Lawton Instrumental Activities of Daily Living (IADLs) Scale **IT-10.2.d**	**Summary**[75] Purpose: The Lawton Instrumental Activities of Daily Living Scale (IADL) is an appropriate instrument to assess independent living skills (Lawton &Brody, 1969). These skills are considered more complex than the basic activities of daily living as measured by the Katz Index of ADLs. Overview: This instrument is intended to be used among older adults, and may be used in community, clinic, or hospital settings. The instrument is not useful for institutionalized older adults. It may be used as a baseline assessment tool and to compare baseline function to periodic assessments. Scores[76]: The most common method is to rate each of the eight items either dichotomously (0 = less able, 1 = more able) or trichotomously (1 = unable, 2 = needs assistance, 3 = independent) and sum the eight responses. The higher the score, the greater the person's abilities. Administration Time: 10-15 minutes Format: Eight items on a paper format.
	Clinical / Quality Information Norms/Factors: Reliability: Reliability was established with twelve subjects interviewed by one interviewer with the

[75] http://consultgerirn.org/uploads/File/trythis/try_this_23.pdf
[76] http://tuftshealthplans.com/providers/pdf/lawton_iadl.pdf

Tool	Tool Information
	second rater present but not participating in the interview process. Inter-rater reliability was established at 0.85. [77] Little is known about the reliability of the Lawton IADL scale, other than the information given in the original report by the developers of the scale. The reproducibility coefficient was 0.96 for men and 0.93 for women (n = 97 and n = 168, respectively). [78] Validity: The validity of the Lawton IADL was tested by determining the correlation of the Lawton IADL with four scales that measured domains of functional status, the Physical Classification (6-point rating of physical health), Mental Status Questionnaire (10-point test of orientation and memory), Behavior and Adjustment rating scales (4-6-point measure of intellectual, person, behavioral and social adjustment), and the PSMS (6-item ADLs). A total of 180 research subjects participated in the study, however, few received all five evaluations. All correlations were significant at the 0.01 or 0.05 level. [79] The correlations between the IADL scale and the other measures of functional status ranged between 0.40 and 0.61. [80]
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator: The total score may range from 0 – 8. A lower score indicates a higher level of dependence. A summary score ranges from 0 (low function, dependent) to 8 (high function, independent) for women, and 0 through 5 for men. [81]
	Additional Notes/Links Link to survey: https://www.abramsoncenter.org/pri/documents/IADL.pdf Versions/Languages: English, Chinese, and Spanish Original Publication Date: 1969 Copyright: The Gerontological Society of America. Permission to needed to reproduce.
Bristol Activities of Daily Living (ADLs) Scale **IT-10.2.e**	**Summary** Purpose: The Bristol ADL scale was specifically designed for use in patients with dementia (Bucks et al, 1996). [82] Overview: The BADL is comprised of 20-items across four components: instrumental activities of daily living (7-items), self-care (6-items), orientation (5-items), and mobility (2-items). Scores: The BADLS has a minimum possible score of 0 (totally independent) and a maximum score of 60 (totally dependent). [83] Format: 20-items across four components.

[77] http://consultgerirn.org/uploads/File/trythis/try_this_23.pdf
[78] http://tuftshealthplans.com/providers/pdf/lawton_iadl.pdf
[79] http://consultgerirn.org/uploads/File/trythis/try_this_23.pdf
[80] http://tuftshealthplans.com/providers/pdf/lawton_iadl.pdf
[81] http://www.healthcare.uiowa.edu/igec/tools/function/lawtonbrody.pdf
[82] http://bjp.rcpsych.org/content/180/2/161.full
[83] http://ageing.oxfordjournals.org/content/25/2/113.full.pdf

Tool	Tool Information
	Clinical / Quality Information Norms/Factors: The final BADL was normed within fifty-nine subjects (25 men and 34 women with dementia) aging 55-91 years. Reliability: The Kappa scores of the 22 items can be divided into the following categories: three have 'fair' Kappa scores (0.21 to 0.40); five have 'moderate' Kappa scores (0.41 to 0.60); 12 have 'good' Kappa scores (0.61 to 0.80); and two have 'very good' Kappa scores (0.81-1.0). Despite the use of carer ratings, therefore, the test-retest reliability of the ADL scale was good. Note: two of the items (medication, and transferring to bed) were removed from the final BADL. Validity: ADL scale scores were correlated both with MMSE and with observed task performance on the Observational Scale. At visit 1, subjects were assessed with the revised ADL scale and the MMSE. The correlation between these two measures was r = —0.55 (p = 0.01, 30.3% variance explained). Ac visit 2 subjects were reassessed with the revised ADL scale and with the Observational scale. The correlation between observed ADL task performance and carer-rated task performance was r = 0.65 (p = 0.004, 42.3% variance explained).
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator: The BADLS has a minimum possible score of 0 (totally independent) and a maximum score of 60 (totally dependent).
	Additional Notes/Links Link to survey: http://www.health.fgov.be/internet2Prd/groups/public/@public/@dg1/@acutecare/documents/ie2divers/19073273_nl.pdf Versions/Languages: English and Dutch Original Publication Date: 1996
Activity Measure for Post-Acute Care (AMPAC) **IT-10.3.a**	**Summary** Purpose: The Activity Measure for Post-Acute Care measures function in three domains: basic mobility (131 items), daily activities (88 items), and applied cognition (50 items). Overview: The AMPAC is used among adults in the inpatient and outpatient rehabilitation, home care, nursing homes and long-term acute care settings. Scores: The following are generic scoring for the inpatient or outpatient short forms: 1. Add the number values associated with the response to each item (For example, items totals yield Raw Score = 10). 2. Match the raw score to the t-Scale scores (t-Scale score = 32.29, SE = 3.42). 3. Find the associated CMS % (CMS % = 76.75%). 4. Locate the correct CMS Functional Modifier Code, or 'G Code' (G code = CL) NOTE: Each AM-PAC Short Form has a separate conversion table. Make sure that you use the correct conversion table. Format: Each short form has different response types. See the AM-PAC Short Form Manual for more: http://www.bu.edu/bostonroc/files/2013/02/AM-PAC-Short-Form-Manual_2013_v2.pdf

<table>
<tr><th>Tool</th><th>Tool Information</th></tr>
<tr><td></td><td>
Clinical / Quality Information

Norms/Factors: Normative data can be found at http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=978

Reliability[84]: Results demonstrate acceptable reliability with the following intraclass correlation coefficients: 1.) test-retest summary scores for each of the three domains ranged between 0.91 and 0.97; 2.) subject proxy summary scores for each of the three domains ranged between 0.68 and 0.90. (Andres P, 2003) Internal consistency reliability was high for the total sample (Cronbach alpha = 0.92 to 0.94), and for specific diagnostic groups (Cronbach alpha = 0.90 to 0.95). (Haley SM, 2004). Additional information can be found at http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=978.

Validity: Coster WJ, et al., examined the dimensional structure and content coverage of a Personal Care and Instrumental Activities item set and compared ADL and IADL items from existing instruments (FIM, MDS, MDS-PAC, OASIS, PF-10) to a set of new items (AM-PAC) as measures of this domain. ADL and IADL items from existing rehabilitation outcomes instruments that depend on skilled upper limb and hand were located along a single continuum, along with the new items from the AM-PAC that addressed gaps in content. Results support the validity of dimension of function as a guide for future development of rehabilitation outcome instruments, such as linked, setting-specific short forms and computerized adaptive testing approaches. Additional information can be found at http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=978.
</td></tr>
<tr><td></td><td>
Availability of Benchmark Data/Clinical Indicator

Clinical Indicator[85]: The Outpatient and Inpatient AM-PAC scores are translated from the raw scores into t-Scale scores (see citation for specific tables). The t-Scale scores then converted into CMS percentages which are linked to a CMS Functional Modifier, or ‘G’Code. Below is a summary of the ‘G’ Codes:
<table>
<tr><th>CMS Modifer ‘G-Code’</th><th>Impairment Limitation Restriction Description</th></tr>
<tr><td>CH</td><td>0 percent impaired, limited or restricted</td></tr>
<tr><td>CI</td><td>At least 1 percent but less than 20 percent impaired, limited or restricted</td></tr>
<tr><td>CJ</td><td>At least 20 percent but less than 40 percent impaired, limited or restricted</td></tr>
<tr><td>CK</td><td>At least 40 percent but less than 60 percent impaired, limited or restricted</td></tr>
<tr><td>CL</td><td>At least 60 percent but less than 80 percent impaired, limited or restricted</td></tr>
<tr><td>CM</td><td>At least 80 percent but less than 100 percent impaired, limited or restricted</td></tr>
<tr><td>CN</td><td>100 percent impaired, limited or restricted</td></tr>
</table>
</td></tr>
<tr><td></td><td>
Additional Notes/Links

Link to survey:
Inpatient – Basic Mobility: http://www.bu.edu/bostonroc/files/2013/02/AM-PAC_SF1a_Basic-Mobility_Inpatient_FORM1.pdf
</td></tr>
</table>

[84] http://www.crecare.com/downloads/AM_PAC_CAT_manual_5.1.07.pdf.pdf
[85] http://www.bu.edu/bostonroc/files/2013/02/AM-PAC-Short-Form-Manual_2013_v2.pdf

Tool	Tool Information
	Inpatient – Daily Activity: http://www.bu.edu/bostonroc/files/2013/05/AM-PAC_SF2a_Daily-Activity_Inpatient_FORM.pdf Outpatient – Basic Mobility: http://www.bu.edu/bostonroc/files/2013/05/AM-PAC_SF3a_Basic-Mobility_Outpatient_FORM.pdf Outpatient - Daily Activity: http://www.bu.edu/bostonroc/files/2013/05/AM-PAC_SF4a_Daily-Activity_Outpatient_FORM.pdf Outpatient - Applied Cognitive: http://www.bu.edu/bostonroc/files/2013/05/AM-PAC_SF5a_Applied-Cognitive_Outpatient_FORM.pdf Distributor: Boston University School of Public Health Versions/Languages: Inpatient and Outpatient short forms Original Publication Date: 2003 Copyright: The AM-PAC short forms are copyrighted and to maintain instrument integrity, the instructions, items and response options cannot be altered. Web Access: http://www.bu.edu/bostonroc/instruments/am-pac/ Contact/Availability: Email: memarino@bu.edu Phone: 617-638-1993 Cost[86]: AM-PAC is free for academic research use; there is a charge for clinical/commercial use. The annual cost for unlimited use of the paper version of the AM-PAC is $250 per year, per site. An annual license for the software is $600.
The Duke Health Profile (Duke) **IT-10.3.b**	**Summary**[87]: Purpose: The Duke Health Profile (Duke) is a 17-item standardized self-report instrument containing six health measures (physical, mental, social, general, perceived health, and self-esteem), and four dysfunction measures (anxiety, depression, pain, and disability). Overview: Most extensive use has been in family practice patients with the broadest spectrum of diagnoses, but it has also been used in patient populations with specific diagnoses such as insulin-dependent diabetes mellitus, end-stage renal disease, ischemic disease, and impotence Scores: The DUKE can be hand-scored using a template for manual scoring. Scoring takes several minutes. Scores on subscales can range from 0 to 100. High scores on the health subscales indicate good health, while high scores on the dysfunction subscales represent high dysfunction or poor health. Administration time: < 5 minutes Format: It can be self-administered by the individual respondent or administered by another person. It is crucial that each question is answered. There are 11 scales. Six scales (i.e., physical health, mental health, social health, general health, perceived health, self-esteem) measure function, with high scores indicating better health. Five scales (i.e., anxiety, depression, anxiety-depression, pain disability) measure dysfunction, with high scores indicating greater dysfunction.

[86] http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=978
[87] http://www.ncbi.nlm.nih.gov/pubmed/9213865

Tool	Tool Information
	Clinical/Quality information[88]: Reliability: Both internal consistency (Cronbach's alpha) and temporal stability (test-retest) testing have supported reliability of the DUKE. Alpha coefficients for the eight multi-item measures ranged from 0.55 to 0.78; test-retest coefficients for the 11 measures ranged from 0.30 to 0.78 and exceeded 0.5 for all except pain and disability health), 0.70 (mental health), and 0.61 (social); the correlation of the overall scores was 0.86 Validity: Validity has been supported for the DUKE scales by: (a) comparison of the DUKE scores with scores of other health measures for the same patients, (b) comparison of DUKE scores between patient groups having different clinical diagnostic profiles and severity of illness, (c) prediction of health-related outcomes by DUKE scores. Convergent and discriminant validity have been shown when comparing with other instruments. The correlations between selected DUKE scales and the equivalent scales from the DUHP instrument they were derived from were 0.72 (physical From the literature: The DUKE has been used primarily for research on health-related outcomes in the clinical setting. Because of its predictive value, the DUKE is one component of a new ambulatory case-mix classification system called the Duke Case-Mix System (DUMIX). The 7-item anxiety-depression scale (DUKE-AD) has been used as an effective screener for DSM-III-R major anxiety and depression. The DUKE-AD can be administered independently, with manual scoring simple enough to be done by the respondent. In addition, the DUKE has been used in health promotion programs to give medical students feedback on their personal health status. The main distinctive feature of[89] the DUKE is its inclusion of the self-esteem category. The early evidence for reliability and validity suggests that the emotional components are sound but that the physical health scale does not perform in the manner expected.
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): Reference score values are available in the *User's Guide* for eight age-gender categories in a sample of normal insurance policyholders (N=3521) and a random sample of primary care patients (N=1997). Patients with painful physical problems had a DUKE physical health mean score of 58.1, while patients with only health maintenance problems had a mean score of 83.9 (scale: 0.0 = poorest health and 100.0 = best health). Patients with mental health problems had a DUKE mental health mean score of 49.2, in contrast to 75.7 for patients with painful physical problems and 79.2 for those with health maintenance.[90] Clinical Indicator[91]: For physical health (items 8, 9, 10, 11, 12), mental health (items 1, 4, 5, 13, 14), social health (2, 6, 7, 15, 16), general health (sum of physical, mental, and social health scores), self-esteem (items 1, 2, 4, 6, 7), and perceived health (item 3), 100 indicates the best health status, and 0 indicates the worst health status. For anxiety (items 2, 5, 7, 10, 12, 14), depression (items 4, 5, 10, 12, 13), anxiety-depression (items 4, 5, 7, 10, 12, 13, 14), pain (item 11), and disability (item 17), 100 indicates the worst health status and 0

[88] http://a4ebm.org/sites/default/files/Measuring%20Health.pdf
[89] http://a4ebm.org/sites/default/files/Measuring%20Health.pdf
[90] http://www.ncbi.nlm.nih.gov/pubmed/2250492
[91] http://healthmeasures.mc.duke.edu/images/DukeForm.pdf

Tool	Tool Information
	indicates the best health status. *Note: If one or more response is missing within one of the eleven scales, a score cannot be calculated for that particular scale.*
	Additional Notes/Links: Link to survey: http://healthmeasures.mc.duke.edu/ Distributor: Department of Community and Family Medicine, Duke University Medical Center, Durham, N.C., U.S.A. Versions and Languages: [92]The DUKE has been translated into seventeen languages; a listing is included in the manual (2, Appendix F) and on the Qolid Web site. Web access: http://healthmeasures.mc.duke.edu/images/DukeForm.pdf Copyright: 1989-2012 Contact/Availability: The developers of the DUKE encourage its use by others. Although the measure is copyrighted to assure quality control, permission to use it for clinical and research purposes is granted upon request, usually without charge. Permission for commercial use must be negotiated with Duke University. Further information is available in the *User's Guide for Duke Health Measures* by George R. Parkerson, Jr., M.D., M.P.H., published by the Department of Community and Family Medicine, Duke University Medical Center, 1999. Telephone: 919-681-3043, E-mail: parke001@mc.duke.edu **Instrument Packet Includes:** ▪ A master copy of the instrument including royalty-free permission to use and reproduce (upon filing a Project Registration Form), as necessary; ▪ DUKE User's Guide for Duke Health Measures; ▪ Reprints of six publications describing the instrument's development, measurement, and/or applications properties; ▪ A Frequently Asked Questions list to aid the instrument's user; and *Order Code: I-DUKE* US$155
Achenbach System of Empirically Based Assessment (ASEBA) **IT-10.3.d**	**Summary**: Assesses adaptive and maladaptive functioning. Purpose: Demonstrates individuals functioning in terms of both quantitative scores and individualized descriptions in respondents' own words. Descriptions include what concerns respondents most about the clients, the best things about clients, and details of competencies and problems that are not captured by quantitative scores alone. Overview: The Achenbach System of Empirically Based Assessment (ASEBA) offers a comprehensive approach to assessing adaptive and maladaptive functioning. Developed through decades of research and practical experience to identify actual patterns of functioning, the ASEBA provides professionals with user-friendly tools. Scores: ASEBA can be hand or computer scored. Respondents complete the CBC and CTR by circling one of three responses and the LDS by circling the words the child uses spontaneously. The behavioral

[92] http://a4ebm.org/sites/default/files/Measuring%20Health.pdf

Tool	Tool Information

	raw scores are derived by summing the response item values (0=Not True, 1=Somewhat True, or 2=Very True or Often True) for the syndrome scale, syndrome groupings, and total score. The raw score for the language development survey is the total number of circled words. The manual provides instructions for converting raw scores into T-scores: normal (under 93 percent), borderline (93 to 97 percent), or clinical (over 97 percent).
	Administration time: 15-20 minutes
	
	Reading Level: Fifth (5^{th}) grade
	
	Format:
	• Child Behavior Checklist for Ages 6-18 (CBCL/6-18)
	  ○ Completed by parents, parent-surrogates, and others who see children in family-like contexts
	• Youth Self-Report (YSR)
	  ○ Completed by youths to describe their own functioning
	• Teacher's Report Form (TRF)
	  ○ Completed by teachers and other school personnel who are familiar with children's functioning in school, such as teacher aides, counselors, administrators, and special educators
	• Adult Self-Report (ASR) and Adult Behavior Checklist (ABCL) – used among individuals aged 18-59 years
	• Older Adult Self-Report (OASR) and Older Adult Behavioral Checklist (OABCL) – adults ≥ 60 years

	Clinical/Quality information: Norms: **National Probability Sample** • To determine which items should be included in the scales, they tested the ability of each CBCL & YSR competence items and TRF Adaptive Functioning item to discriminate between children who were referred for mental health services vs. demographically similar non-referred children • 1,753 children 6-11 • 1,057 youths 11-18 • A "healthy sample" was then derived to create the normative sample Reliability/Validity: Reliability and validity data for preschool, school-age, adults, and older adults can be found at: Preschool: http://www.aseba.org/ordering/ASEBA%20Reliability%20&%20Validity-Pre-school%20.pdf School-Aged: http://www.aseba.org/ordering/ASEBA%20Reliability%20and%20Validity-School%20Age.pdf Adults: http://www.aseba.org/ordering/ASEBA%20Reliability%20and%20Validity%20-%20Adult.pdf Older Adults: http://www.aseba.org/ordering/ASEBA%20Reliability%20and%20Validation-Older%20Adults.pdf From the literature: http://www.aseba.org/ The following is provided from http://archive.acf.hhs.gov/programs/opre/ehs/perf_measures/reports/resources_measuring/res_meas_cdia.html: Reliability: (1) Internal consistency reliability (Cronbach's alpha): the alphas for the CBC scales ranged from .66 to .92 for the syndromes and .63 to .86 for the DSM-oriented scales. The alphas were .89 and .92 for the two broader groupings (internalizing and externalizing syndromes) and .95 for the total score. The alphas for the CTR syndromes ranged from .52 to .96 and for the DSM-oriented scales from .68 to .93. The alphas were .89 and .96 for the internalizing and externalizing groupings and for the total score, .97. (2) Test-retest reliability, with an eight-day interval between tests: the correlations were .85 and .76 for the CBC and CTR, respectively. Test-retest studies on the LDS reported correlations greater or equal to

Tool	Tool Information
	.90. Validity: (1) Concurrent validity: The CBC correctly classified 84 percent of a sample of children (some of whom were diagnosed as having emotional/behavioral problems), and the CTR correctly classified 74 percent of the children. Studies reported correlation coefficients between the CBC problem syndromes and the Toddler Behavior Screening Inventory and the Infant-Toddler Social and Emotional Assessment ranging from .48 to .70. In 11 studies that compared parent LDS scores with those obtained by trained examiners using other measures, the correlations between the parent's score and the trained examiner's ranged from .56 to .87. Other studies found the level of LDS agreement with other measures of language development ranged from .47 to .94. (2) Predictive validity: An 11-year longitudinal study found that children identified by the LDS to have language development problems were more likely to have weak verbal skills at age 13.
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator[93]**:** The ASEBA provides the user with T-scores to compare a child's performance against other children and the scoring forms classify scores as normal (under 93 percent), borderline (93 to 97 percent), or clinical (over 97 percent). The authors recommend that the results be interpreted by someone with some graduate training. Must purchase the ASEBA Scoring License to get more information on translating scores into clinical outcomes: http://www.aseba.org/sitelicense.html
	Additional Notes/Links: Link to Survey: http://www.aseba.org/ordering/howtoorder.html Distributor: ASEBA, or PAR, Inc. 16204 North Florida Avenue Lutz, FL 33549 http://www3.parinc.com/ Versions and Languages: 85 languages. CBCL/1½-5, CBCL/6-18, TRF/6-18, YSR/11-18, ABCL/18-59, ASR/18-59 are available in Spanish (Latino). Original Publication Date: Late 1960's and early 1970's. Copyright: 2013 Web Access: www.aseba.org Email Communication: mail@aseba.org Fax: 802-656-5131 Contact/Availability: ASEBA 1 South Prospect Street Burlington, VT 05401-3456

93 http://archive.acf.hhs.gov/programs/opre/ehs/perf_measures/reports/resources_measuring/res_meas_cdia.html

Tool	Tool Information
	Cost: Paperback surveys can be purchased for $25.00 (50 pack)
Quick Inventory of Depressive Symptomatology[94] **IT-11.22.d**	**Summary**: Purpose: The IDS and QIDS were developed to improve on the available clinician and patient ratings by 1) providing equivalent weightings (0-3) for each symptom item; 2) providing clearly stated anchors that estimate the frequency and severity of symptoms; 3) including all DSM-IV criterion items required diagnosing a major depressive episode; and 4) providing matched clinician and patient ratings. Overview: The 30 item Inventory of Depressive Symptomatology (IDS) (Rush et al. 1986, 1996) and the 16 item Quick Inventory of Depressive Symptomatology (QIDS) (Rush et al. 2003) are designed to assess the severity of depressive symptoms. The IDS and QIDS assess all the criterion symptom domains designated by the American Psychiatry Association Diagnostic and Statistical Manual of Mental Disorders - 4th edition (DSM-IV) (APA 1994) to diagnose a major depressive episode. These assessments can be used to screen for depression, although they have been used predominantly as measures of symptom severity. The seven day period prior to assessment is the usual time frame for assessing symptom severity. The QIDS-C_{30} and QIDS-SR_{16} cover only the nine diagnostic symptom domains used to characterize a major depressive episode, without items to assess atypical, melancholic or their commonly associated symptoms. All 16 items on the QIDS are included within the IDS. The IDS-C_{30} and IDS-SR16 include the criterion symptoms, as well as commonly associated symptoms (e.g. anxiety, irritability) and items relevant to melancholic, or atypical symptom features. Scores: Enter the highest score on any 1 of the 4 sleep items (items 1 to 4). Enter the highest score on any 1 of the 4 weight items (items 6 to 9). Enter the highest score on either of the 2 psychomotor items (15 and 16). There will be one score for each of the nine MDD symptom domains. Add the scores of the of the 9 items (sleep, weight, psychomotor changes, depressed mood, decreased interest, fatigue, guilt, concentration, and suicidal ideation) to obtain the total score. Total scores range from 0-27. Administration Time: 5-7 minutes Format: Both the IDS and the QIDS are available in the clinician (IDS-C_{30} and QIDS-C_{16}) and self-rated versions (IDS-SR_{30} and QIDS-SR_{16}). Pencil and paper or automated telephone-administered format (IVR).
	Clinical/Quality information: Reliability: Cronbach's alpha was 0.85 (QIDS-C_{16}), 0.86 (QIDS-SR_{16}), 0.90 (IDS-C_{30}), 0.92 (IDS-SR_{30}) for the MDD patients, and 0.81 (QIDS-C_{16}), 0.89 (IDS-C_{30}) for the patients with BD. Validity: Both versions of the QIDS have been used with MDD and BD populations. IDS and QIDS total scores were comparable to those obtained by the $HRSD_{17}$ and BDI, Rush et al. (2004) found high correlations between the QIDS-SR_{16}, and the IDS-SR_{30} (c=0.96), $HRSD_{17}$ (c=0.81), $HRSD_{21}$ (c=0.82), and $HRSD_{24}$ (c=0.84) at the exit interview after 12 weeks of acute phase outpatient treatment (n=596). Trivedi et al. (2004) found the QIDS-SR_{16} total score was highly correlated with the IDS-SR_{30} total score for 544 adult outpatients with MDD (c=0.83). They also found robust correlations between the QIDS-C_{16} and IDS-C_{30} total scores for out-patients with MDD (c=0.82, n=544) and Bipolar Disorder (BD) (c=0.81, n=402). From the literature: The IDS_{30} and $QIDS_{16}$ are sensitive to changes in depressive severity in a manner consistent with the HRSD and BDI, with the IDS and QIDS demonstrating a greater sensitivity to change

[94] http://www.ids-qids.org/index2.html#ADMINISTRATION

<table>
<tr><th>Tool</th><th>Tool Information</th></tr>
<tr><td></td><td>in the lower range of scores reported by 434 outpatients patients with MDD and 103 normal controls (Rush et al. 1996). In 68 newly admitted inpatients with a diagnosis of MDD, Corruble et al. (1999) found the IDS-C_{30} and IDS-SR_{30} to be as sensitive to change as the MADRS and the SCL-90 (depression subscale). In 62 patients with MDD, Biggs et al. (2000) found the IDS-C_{30} and IDS-SR_{30} to be significantly more sensitive to detecting change than standard five item visual analogue physician and patient self-report global rating scales. In a retrospective analysis, Rush et al. (2000) found comparable levels of sensitivity to change in depressive severity when comparing $HRSD_{24}$ and the IDS-SR_{30} scores in a sample of 993 outpatients with MDD. In 596 adult outpatients with chronic, nonpsychotic MDD Rush et al. (2003) found the IDS-SR_{30} and QIDS-SR_{16} were equally sensitive to symptom change, when viewed as a discontinuous variable (response or remission), although the QIDS-SR_{16} seemed to be slightly less sensitive to residual symptoms that the longer IDS-SR_{30}. In 544 outpatients with MDD, and 402 outpatients with Bipolar Disorder (BP), Trivedi et al. (2004) reported finding equal sensitivity to symptom change when comparing the IDS-C_{30}, IDS-SR_{30}, QIDS-C_{16}, QIDS-SR_{16}, indicating high concurrent validity for all four scales. High concurrent validity was also documented based on the Medical Outcomes Study 12-item Short Form (SF-12) (Ware et al. 1996) mental health summary score for the population divided in quintiles based on their IDS and QIDS score. In 681 patients with chronic MDD assigned to 3 treatment groups (medication alone, medication and psychotherapy, psychotherapy alone), Rush et al. (2005) compared the $HRSD_{24}$ and the IDS-SR_{30} and QIDS-SR_{16} ratings, finding comparable change scores within groups. In addition, the IDS-SR_{30} and QIDS-SR_{16} confirmed response and remission rates based on the $HRSD_{24}$.</td></tr>
<tr><td></td><td>Availability of Benchmark Data/Clinical Indicator

Clinical Indicator[95]:
Severity of Depression. 0=None, 1=Mild, 2=Moderate, 3=Severe, 4=Very Severe
<table>
<tr><th>Depression Severity</th><th>Quick Inventory of Depressive Symptomatology Score</th></tr>
<tr><td>None</td><td>0-5</td></tr>
<tr><td>Mild</td><td>6-10</td></tr>
<tr><td>Moderate</td><td>11-15</td></tr>
<tr><td>Severe</td><td>16-20</td></tr>
<tr><td>Very Severe</td><td>21-27</td></tr>
</table></td></tr>
<tr><td></td><td>Additional Notes/Links:
Link to survey: http://www.ids-qids.org/index.html

Distributor: University of Pittsburgh, Epidemiology Data Center

Versions and Languages: English and 24 other languages

Original Date of Publication: 2003

Copyright: 2013

Web Access: http://www.ids-qids.org/index2.html#ABOUT

Cost: Free</td></tr>
<tr><td></td><td></td></tr>
</table>

[95] http://www.ids-qids.org/index2.html#table3

Tool	Tool Information
Edinburgh Postnatal Depression Scale[96] **IT-11.22.e**	**Summary**: Purpose: The Edinburgh Postnatal Depression Scale is designed to screen women for symptoms of emotional distress during pregnancy and the postnatal period. The Edinburgh Postnatal Depression Scale (EPDS**)** was developed in 1987 for screening postpartum women in outpatient, home visiting settings, or at the 6-8 week postpartum examination. It has been utilized among numerous populations, including US women and Spanish-speaking women in other countries. The scale has since been validated, and evidence from a number of research studies has confirmed the tool to be both reliable and sensitive in detecting depression. Overview: The EPDS is not a diagnostic tool and must always be used in conjunction with clinical assessment; the EPDS score should not override clinical judgment. A careful clinical assessment should be carried out to confirm the diagnosis. The scale indicates how the mother has felt during the previous week. In doubtful cases it may be useful to repeat the tool after 2 weeks. The scale will not detect mothers with anxiety neuroses, phobias, or personality disorders. Scores: Range of EPDS Scores 0-9 : Scores in this range may indicate the presence of some symptoms of distress that may be short-lived and are less likely to interfere with day to day ability to function at home or at work. However if these symptoms have persisted more than a week or two further enquiry is warranted. 10-12 : Scores within this range indicate presence of symptoms of distress that may be discomforting. Repeat the EDS in 2 weeks-time and continue monitoring progress regularly. If the scores increase to above 12 assess further and consider referral as needed. 13 + (Max: 30): Scores above 12 require further assessment and appropriate management as the likelihood of depression is high. Referral to a psychiatrist/psychologist may be necessary. Administration time: 5 minutes Format: 10-item, clinician administered questionnaire. **Instructions for users:** • The mother is asked to underline the response that comes closest to how she has been feeling in the previous 7 days. All ten items must be completed. • Care should be taken to avoid the possibility of the mother discussing her answers with others. • The mother should complete the scale herself, unless she has limited English or has difficulty with reading. • The EPDS may be used at 6-8 weeks to screen postnatal women. The child health clinic, postnatal check-up or a home visit may provide suitable opportunities for its completion. **Guidelines for Evaluation:** Response categories are scored 0, 1, 2, and 3 according to increased severity of the symptom. Questions 3, 5, 6, 7, 8, 9, 10 are reverse scored (i.e., 3, 2, 1, 0) Individual items are totaled to give an overall score. A score of 12+ indicates the likelihood of depression, but not its severity. The EPDS score is designed to assist, not replace, clinical judgment. Women should be further assessed before deciding on treatment. This scale may be reproduced by users without further permission providing they respect copyright by

[96] http://www.fresno.ucsf.edu/pediatrics/downloads/edinburghscale.pdf

Tool	Tool Information
	quoting the names of the authors, the title and the source of the paper in all reproduced copies. As it reflects the woman's experience of the last 7 days, the EPDS may need to be repeated on further occasions as clinically warranted.
	Clinical/Quality information: Reliability[97]: The split-half reliability of the 10-item EPDS was 0.88 and the standardized α coefficient 0.87. Validity[98]: The sensitivity of the EPDS (the proportion of women with RDC depression who were true positives) was 86%, and the specificity (proportion of RDC non-depressed women who were true negatives) was 78%. The positive predictive value (the proportion of women above the threshold on the EPDS (n = 41) who met RDC for depression (n = 30)) was 73%. These findings suggested that the rate for failing to detect women with depression could be reduced to under 10% by using a lower cut-off, of 9/10. This was the cut-off that we recommended in our initial publication of the EPDS for use as a first-stage screening measure. From the literature: The Edinburgh Postnatal Depression Scale is a useful tool to screen patients for postpartum depression. Reliability for the English (alpha .81) and Spanish (alpha .71) versions of the EPDS was adequate. Mean scores of the EPDS and the Center for Epidemiologic Studies-Depression Scale resulted in a correlation of r=0.75 strengthening the convergent validity of both instruments. A correlation (r=0.43) between index question assessing prenatal mood and EPDS mean scores further enhanced validity.[99] The Edinburgh Postnatal Depression Scale has been developed to assist primary care health professionals to detect mothers suffering from postnatal depression, a distressing disorder more prolonged than the "blues" (which occur in the first week after delivery) but less severe than puerperal psychosis. **Previous studies have shown that postnatal depression affects at least 10% of women and that many depressed mothers remain untreated.** These mothers may cope with their baby and with household tasks, but their enjoyment of life is seriously affected and it is possible that there are long-term effects on the family. The EPDS was developed at health centers in Livingston and Edinburgh. It consists of ten short statements. The mother underlines which of the four possible responses is closest to how she has been feeling during the past week. Most mothers complete the scale without difficulty in less than 5 minutes. The validation study showed that mothers who scored above threshold 92.3% were likely to be suffering from a depressive illness of varying severity. Nevertheless, the EPDS score should not override clinical judgment. A careful clinical assessment should be carried out to confirm the diagnosis. The scale indicates how the mother has felt during the previous week and in doubtful cases, it may be usefully repeated after 2 weeks. The scale will not detect mothers with anxiety neuroses, phobias or personality disorder. Taken from the British Journal of Psychiatry. June, 1987, Vol. 150 by J.L. Cox, J.M. Holden, R. Sagovsky
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator[100]: **Maximum score:** 30

[97] http://www.rcpsych.ac.uk/files/samplechapter/81_1.pdf
[98] http://www.rcpsych.ac.uk/files/samplechapter/81_1.pdf
[99] http://www.resourcenter.net/images/snrs/files/sojnr_articles2/vol10num01art03.html
[100] http://www.fresno.ucsf.edu/pediatrics/downloads/edinburghscale.pdf

<table>
<tr><th>Tool</th><th>Tool Information</th></tr>
<tr><td></td><td>Possible depression: 10 or higher
Always look at Question #10: The EPDS includes one question (Item 10) about suicidal thoughts and should be scored before the woman leaves the office in order to detect whether this item has been checked. Further enquiry about the nature of any thoughts of self-harm is required in order for the level of risk to be determined and appropriate referrals made where indicated to ensure the safety of the mother and baby.</td></tr>
<tr><td></td><td>Additional Notes/Links:

Link to survey: http://www.fresno.ucsf.edu/pediatrics/downloads/edinburghscale.pdf

Distributor: UCSF

Versions and Languages:
Validated
1. Arabic
2. Chinese
3. Dutch
4. French
5. German
6. Igbo
7. Italian
8. Japanese
9. Malay
10. Maltese
11. Norwegian
12. Portuguese
13. Punjabi
14. South African - English
15. Spanish
16. Swedish
17. Turkish
18. Vietnamese

Not Validated
1. Afaan Oromo - Ethiopia
2. Amharic
3. Czech
4. Farsi/Persian
5. Filipino/Tagalog
6. Greek
7. Hebrew
8. Hindi
9. Indonesian
10. Khmer/Cambodian
11. Korean
12. Macedonian
13. Myanmar/Burmese
14. Serbian
15. Slovenian
16. Somali
17. Thai
18. Urdu

Original Publication Date: 1987

Copyright: Users may reproduce the scale without further permission providing they respect copyright by quoting the names of the authors, the title and the source of the paper in all reproduced copies.

Cost: Free</td></tr>
<tr><td>Experience of Care and Health Outcomes (ECHO) 3.0[101]

IT-11.24.a</td><td>Summary:

Purpose: Consumers, clinicians, MBHOs, health care plans, purchasers, States, and Federal agencies can collect consumers' ratings of their behavioral health treatment, including mental health and alcohol, drug, and other substance abuse services.

Overview: ECHO™ assesses specific aspects of care, including getting care quickly, communication with clinicians, information provided by clinicians on medication side effects, family involvement in care, information about self-help groups and treatment, cultural competence, perceived improvement in functioning, patient rights, and health plan or MBHO administrative services.</td></tr>
</table>

[101] http://www.hcp.med.harvard.edu/echo/

Tool	Tool Information
	Scores: There are a number of statistical methods for creating the composite measures. We recommend following the algorithm developed by the CAHPS Team. Information about the CAHPS methodology is available in the CAHPS Survey and Reporting Kit available at www.CAHPS-SUN.org (note: you may need to register to access these documents on the website). The CAHPS Team has also developed an analysis program in SAS that can be used to calculate the reporting measures. Please see Volume 7 of HEDIS 2003, Specifications for the ECHOTM 1.1H Survey for MBHOS, for information about computing reporting measures for the NCQA MBHO version of ECHO. Format: The current version of the ECHOTM health plan survey contains 63 items and the MBHO survey contains 51 items. Twenty-eight items on the MBHO survey and 33 items on the health plan survey are used to evaluate the quality of care provided. Many of these items can be grouped into composite measures with other items that measure similar concepts for reporting
	Clinical/Quality information: Full Field test report of methods, criteria for item analysis, criterion validity, factor analysis and loading, reliability, and discriminant validity can be found here: http://www.hcp.med.harvard.edu/echo/ NQF endorsed: http://www.qualitymeasures.ahrq.gov/content.aspx?id=26651
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): The National CAHPS Benchmarking Database (NCBD) is developing a national database of ECHO™ survey results. Survey sponsors who submit data to the NCBD will receive a free benchmarking report. For more information about the NCBD see http://ncbd.cahps.org/Home/index.asp or contact NCBD1@Westat.com Clinical Indicator: Not applicable - Patient Satisfaction tool. Tool scoring methodology can be found at: http://www.hcp.med.harvard.edu/echo/ECHO.measures.scoring.version.3.0.pdf
	Additional Notes/Links: Link to survey: The surveys can be downloaded for free from the Web site: http://www.hcp.med.harvard.edu/echo/home.html Distributor: CAHPS Original Publication Date: 1999 Versions and Languages: **ECHO 3.0 - Experience of Care and Health Outcomes has two surveys**—one for MBHOs and one for health plans. Both surveys ask the same questions, but the health plan survey includes questions about administrative services, such as filling out paperwork and finding information in written materials Child and Adult versions are available. Each survey also offers supplemental questions. Also available in Spanish. Web Access: http://www.hcp.med.harvard.edu/echo/ Contact/Availability: NCBD1@Westat.com Cost: Free

Tool	Tool Information
Generalized Anxiety Disorder (GAD-7) **IT-11.25.a**	**Summary**[102] Purpose: Developed to diagnose generalized anxiety disorder, but can be used as a screening tool for panic, social anxiety, and post-traumatic stress disorder. Overview: Scores: Seven items, each of which is scored 0 to 3, providing a 0 to 21 severity score. Cut points of 5, 10, and 15 represent mild, moderate, and severe levels of depressive, anxiety, and somatic symptoms, on the GAD-7. Reading Level[103]: 6-9th grade Format: Seven items, each of which is scored 0 to 3, providing a 0 to 21 severity score. Clinician administered questionnaire.
	Clinical / Quality Information Norms/Factors: Validated within 2740 primary care patients.[104] Reliability: The internal consistency of the GAD-7 was excellent (Cronbach α = .92). Test-retest reliability was also good (intraclass correlation = 0.83).[105] Internal consistency was identical across all subgroups (alpha = 0.89).[106] Validity: Comparison of scores derived from the self-report scales with those derived from the MHP-administered versions of the same scales yielded similar results (intraclass correlation = 0.83), indicating good procedural validity.[107] Intercorrelations with the PHQ-2 and the Rosenberg Self-Esteem Scale were r = 0.64 (P < 0.001) and r = -0.43 (P < 0.001), respectively.[108] From the literature: Sensitivity = 89%; Specificity=82%[109] Anxiety disorders: panic disorder (sensitivity 74%, specificity 81%), social anxiety disorder (sensitivity 72%, specificity 80%), and posttraumatic stress disorder (sensitivity 66%, specificity 81%).[110] GAD-7 and GAD-2 benchmarks within the primary care population can be found here http://www.goodmedicine.org.uk/files/assessment,%20phq9,%20gad7,%20etc.pdf
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator[111]: Thresholds of anxiety scores as follows: 0-4: None

[102] http://www.phqscreeners.com/instructions/instructions.pdf
[103] https://www.ncbi.nlm.nih.gov/books/NBK82544/
[104] http://www.phqscreeners.com/instructions/instructions.pdf
[105] http://archinte.jamanetwork.com/article.aspx?articleid=410326
[106] http://www.ncbi.nlm.nih.gov/pubmed/18388841
[107] http://archinte.jamanetwork.com/article.aspx?articleid=410326
[108] http://www.ncbi.nlm.nih.gov/pubmed/18388841
[109] Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092-1097.
[110] http://m.psychiatrictimes.com/clinical-scales/DepressionAnxiety/GAD-7.html
[111] http://www.phqscreeners.com/instructions/instructions.pdf

Tool	Tool Information
	5-9: Mild Anxiety 10-14: Moderate Anxiety 15-21: Severe Anxiety
	Additional Notes/Links Link to survey: http://www.mpho.org/resource/d/34008/GAD708.19.08Cartwright.pdf Distributor: Pfizer Versions/Languages: GAD-7 has been translated into numerous languages. Original Publication Date: 2006 Copyright: No permission is required to reproduce, translate, display or distribute them. Web Access: http://www.phqscreeners.com/overview.aspx?Screener=03_GAD-7 Contact/Availability: questions@phqscreeners.com Cost: Free
Daily Living Activities (DLA-20) **IT-11.26.a**	**Summary**: Purpose: The Daily Living Activities functional assessment tool is designed to assess what daily living areas are impacted by mental illness or disability. Overview: The assessment tool quickly identifies where outcomes are needed so clinicians can address those functional deficits on individualized service plans. Use of this tool ensures valid scores and consistent utilization for healthcare report cards. Study findings provide evidence of the usefulness of the DLA to support the functional assessment data needs of service providers. Scores: DLA-20 Youth Mental Health: Scoring Instructions: If all 20 DLAs are rated, sum column and take ½ for estimated CGAS or Step 1. Add scores from applicable column. Step 2. Divide sum by number of activities actually rated. This is the average DLA score. Step 3. To estimate CGAS, multiply the average DLA score by 10. Compare to Axis V and Lower GAF if consumer is symptomatic. Step 4. +/- Change Score: subtract initial average DLA score (R1) from most recent rating (R2-R5). Administration time: The tool has been shown to take approximately 6 to 10 minutes to complete at the conclusion of an assessment. Format: The DLA is intended to be used by all disabilities and ages. Developmental Disabilities and Alcohol/Drug Abuse forms are personalized for daily functional strengths and problems associated with those diagnoses. An Adult form exists for SMI and SPMI consumers over the age of 18 and a Youth form for consumers between the ages of 6 and 18.
	Clinical/Quality information:

Tool	Tool Information
	Reliability:[112] The DLA is a reliable and valid measure for the purposes of level of care consideration, treatment planning around outcomes, and to correlate and predict DSMIV, Axis V. Two studies with 971 consumers over repeated measures will be reviewed with the results reflecting a satisfactory treatment plan time-saver for case coordinators. The tool is published in the *Research on Social Work Practice* (Abstract and other reference articles are in Appendix B). Please note, however, that since 2005, the DLA has been copyrighted to protect reliability and validity, not for additional monetary remuneration beyond training fees. Validity: Criterion-related validity was evidenced by the ability of DLA scores to differentiate consumers in different levels of care and by diagnostic categories.
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator[113]: Performance across the 20 DLAs is measured with the following criteria: 1: None of the time; extremely severe impairment of problems in functioning; pervasive level of continuous paid supports needed 2: A little of the time; severe impairment or problems in functioning' extensive level of continuous paid supports needed 3: Occasionally moderate severe impairment or problems in functioning; moderate level of continuous paid supports needed 4: Some of the time; moderate impairment or problems in functioning; low levels of continuous paid supports needed 5: A good bit of the time; mild impairment or problems in functioning; moderate level of intermittent paid supports needed 6: Most of the time; very mild impairment or problems in functioning; low level of intermittent paid supports needed 7: All of the time; independently managed DLA in community; no impairment or problem in functioning requiring paid supports DLA score can be converted to the GAF (Global Assessment of Functioning) for comparisons of outcome changes. Multiply the average DLA score by 10 to get GAF score. Below is the GAF Scale[114]: **91 - 100** No symptoms. Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities. **81 - 90** Absent or minimal symptoms (e.g., mild anxiety before an exam), good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, no more than everyday problems or concerns (e.g., an occasional argument with family members). **71 - 80** If symptoms are present, they are transient and expectable reactions to psychosocial stressors (e.g., difficulty concentrating after family argument); no more than slight impairment in social, occupational, or school functioning (e.g., temporarily falling behind in schoolwork). **61 - 70** Some mild symptoms (e.g., depressed mood and mild insomnia) or some difficulty in social, occupational, or school functioning (e.g., occasional truancy, or theft within the household), but generally functioning pretty well, has some meaningful interpersonal relationships. **51 - 60** Moderate symptoms (e.g., flat affect and circumstantial speech, occasional panic attacks) or moderate difficulty in social, occupational, or school functioning (e.g., few friends, conflicts with peers or co-workers). **41 - 50** Serious symptoms (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) or any serious impairment in social, occupational, or school functioning (e.g., no friends, unable to keep a job, cannot work). **31 - 40** Some impairment in reality testing or communication (e.g., speech is at times illogical, obscure, or

[112] http://www.thenationalcouncil.org/galleries/resources-services%20files/DLA%20Sample.pdf
[113] http://www.thenationalcouncil.org/wp-content/uploads/2012/11/DLA-Sample.pdf
[114] http://en.wikipedia.org/wiki/Global_Assessment_of_Functioning

Tool	Tool Information
	irrelevant) or major impairment in several areas, such as work or school, family relations, judgment, thinking, or mood (e.g., depressed adult avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school). **21 - 30** Behavior is considerably influenced by delusions or hallucinations or serious impairment, in communication or judgment (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) or inability to function in almost all areas (e.g., stays in bed all day, no job, home, or friends) **11 - 20** Some danger of hurting self or others (e.g., suicide attempts without clear expectation of death; frequently violent; manic excitement) or occasionally fails to maintain minimal personal hygiene (e.g., smears feces) or gross impairment in communication (e.g., largely incoherent or mute). **1 - 10** Persistent danger of severely hurting self or others (e.g., recurrent violence) or persistent inability to maintain minimal personal hygiene or serious suicidal act with clear expectation of death.
Positive Symptom Rating Scale (PSRS) **IT-11.27.a**	**Additional Notes/Links:** Link to Survey: Adult: http://www.thenationalcouncil.org/galleries/resources-services%20files/DLA%20Sample.pdf Child: http://dmh.mo.gov/docs/mentalillness/DLA20Youth.pdf Copyright: W.S. Presmanes, M.A., M.Ed., and R.L. Scott, PhD.; 2005 Web Access: http://www.thenationalcouncil.org/mtm-services/DLA20-FAQ.html Contact/Availability: MTM Services Willa Presmanes, M.ED., MA Senior Outcomes Consultant Phone: (770) 396-6615 E-mail: MTMWilla@aol.com Website: http://www.mtmservices.org/ Cost: Free
	Summary: Purpose: The PSRS is a 4-item measure of positive symptom severity for schizophrenia Overview: The 4-item PSRS assesses the following positive symptoms of schizophrenia: suspiciousness, unusual thought content, hallucinations, and conceptual disorganization. Developed as part of the Texas Medication Algorithm Project, the PSRS is based on the Brief Psychiatric Rating Scale. Scores: The PSRS is a 4-item measure of positive symptom severity for schizophrenia. Each item is scored on a 1 to 7 scale, where higher scores indicate higher severity. PSRS scores range from 4 to 28 points. Format: Clinicians can use the same scoring sheet to record scores from both the PSRS and the BNSA. Administration Time: To administer the PSRS, the provider should read the anchor descriptions for each dimension and then record the appropriate rating. The interview takes no more than 5 minutes for the clinician to complete.
	Clinical Information Norms: According to the Texas Medication Algorithm Project's manual for schizophrenia, an

Tool	Tool Information
	improvement of at least 20% over intake is clinically significant (Argo et al., 2008). This 20% threshold is slightly lower, however, than the recommended lower threshold for other positive and negative symptom assessments for schizophrenia. In a review of quantitative assessments for schizophrenia, Correll et al. (2011) suggest a score reduction of 25% for very chronic or treatment-resistant patients diagnosed with schizophrenia, and a reduction of 50% for acutely ill patients diagnosed with schizophrenia (see also Leucht et al., 2009; Leucht et al., 2005a, and Leucht et al., 2005b). Validity: These items are from the Brief Psychiatric Rating Scale and the expanded version of the Brief Psychiatric Rating Scale, both of which have been shown to be valid and reliable.
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): http://schizophreniabulletin.oxfordjournals.org/content/13/2/261.long MHMRA of Harris County's analysis of FY 2012 adult TRAG data shows that there is statistically significant and clinically meaningful change in PSRS scores since intake. It would probably be safe to expect a 26% to 68% improvement at the 6-month mark, a 20% to 71% improvement at the 12-month mark, and 34% to 100% improvement at the 18-month mark. Based on the analysis, it would be feasible to use PSRS scores as DSRIP outcome measures, so long as the appropriate percent change calculations are used and the aforementioned percentages considered. The score improvements over intake are mostly statistically significant, appear to be clinically meaningful, and have a generally upward trend. Clinical Indicator: The questions are scored as following: 1: Not present 2: Very Mild 3: Mild 4: Moderate 5: Moderately Severe 6: Severe 7: Extremely Severe The symptoms measured are: 1: Suspiciousness 2: Unusual Thought Content 3: Hallucinations 4: Conceptual Disorganization More on the individual rankings of the positive symptoms can be found at: http://www.valueoptions.com/northstar/providers/handbook/PSRS_for_Schizophrenia_Algorithm-Section3-Adult_UA.pdf
	Other information: The 4-Item Positive Symptom Rating Scale was adapted from the Expanded Version of the BPRS developed by: Ventura J, Lukoff D, Nuechterlein KH, Liberman RP, Green MF, and Shaner, *A Manual for the expanded Brief Psychiatric Rating Scale*. International Journal of Methods Psychiatry Research 1993; 3:227-244. http://www.sccp.sc.edu/centers/scorxe/protected/downloads/45107%20padforproofing.pdf

Tool	Tool Information
Aberrant Behavior Checklist (ABC) **IT-11.27.c**	**Summary** Purpose: The Aberrant Behavior Checklist (ABC) is a symptom checklist for assessing problem behaviors of children and adults with mental retardation. Used to assess problem behaviors of children and adults with mental retardation at home, in residential facilities, ICFs/MR, and work training centers. Can be used within children and adolescents with mental retardation in educational settings, residential and community-based facilities, and developmental centers. Overview: The ABC asks for degree of retardation, the individual's medical status and current medical condition. Then 58 specific symptoms are rated and an extensive manual provides comprehensive descriptions for each assessed behavior. The checklist can be completed by parents, special educators, psychologists, direct caregivers, nurses, and others with knowledge of the person being assessed. Scores: 5-Factor structure: 1) Irritability, agitation, crying (15 items); 2) Lethargy, social withdrawal (16 items); 3) Stereotypic behavior (7 items); 4) Hyperactivity, non-compliance (16 items); and 5) Inappropriate speech (4 items). Each item rated from 0 (not at all a problem) to 3 (the problem is severe in degree). Administration time: 10-15 minutes Format: The ABC is a symptom checklist for assessing problem behaviors in individuals ages 6 to 54 with mental retardation.
	Clinical/Quality information: Norms: Developed on a sample of 927 institutionalized, profoundly developmentally delayed adolescents and adults in New Zealand. Reliability: Internal consistency: Aman et al. reported internal consistencies of 0.86-0.94 in the original development study. Generally, other studies have confirmed this range of internal consistencies. However, some studies have found internal consistencies as low as 0.19 (Freund, teacher form). Test-retest: The original development study reported test-retest reliabilities of 0.96-0.99. However, the subsequent studies failed to validate these findings. Generally, have been fairly good, ranging from 0.50-0.67 (Freund, teacher form) to 0.80-0.95 (Freund, parent form). Inter-rater: The original development study reported inter-rater reliabilities of 0.17-0.90, with a mean of 0.60. Subsequent studies have found a wide variability of inter-rater reliabilities, ranging from 0.12 to 0.95 (both in Schroeder). Validity: There has been extensive validation of the 5-factor structure. The original development study found that the ABC demonstrated moderate discriminative validity with a number of instruments, as well as convergent validity with behavioral observation reports. It also demonstrated adequate predictive validity. Subsequent studies have provided further evidence of predictive, convergent and discriminative validities From the literature: Aman, M.G., Burrow, W.H., & Wolford, P.L. (1995). The Aberrant Behavior Checklist-Community: Factor validity and effect of subject variables for adults in group homes. American Journal on Mental Retardation, 100(3), pp. 283-292. Aman, M.G., Richmond, G., Stewart, A.W., Bell, J.C., & Kissell, R.C. (1987). The Aberrant Behavior Checklist: Factor structure and the effect of subject variables in American and New Zealand facilities. American Journal of Mental Deficiency, 91(6), pp. 570-578. Aman, M.G., Singh, N.N., Stewart, A.W., & Field, C.J. (1985a). The Aberrant Behavior Checklist: A

Tool	Tool Information
	behavior rating scale for the assessment of treatment effects. American Journal of Mental Deficiency, 89(5), pp. 485-491. Aman, M.G., Singh, N.N., Stewart, A.W., & Field, C.J. (1985b). Psychometric characteristics of the Aberrant Behavior Checklist. American Journal of Mental Deficiency, 89(5), pp. 492-502. Aman, M.G., Singh, N.N., & Turbott, S.H. (1987). Reliability of the Aberrant Behavior Checklist and the effect of variations in instructions. American Journal of Mental Deficiency, 92(2), pp. 237-240. Bihm, E.M., & Poindexter, A.R. (1991). Cross-validation of the factor structure of the Aberrant Behavior Checklist for persons with mental retardation. American Journal on Mental Retardation, 96(2), pp. 209-211. Brinkley, J., Nations, L., Abramson, R. K., Hall, A., Wright, H. H., Gabriels, R. et al. (2007). Factor analysis of the aberrant behavior checklist in individuals with autism spectrum disorders. Journal of Autism and Developmental Disorders, 37(10), 1949-1959. Brown, E. C., Aman, M. G., & Havercamp, S. M. (2002). Factor analysis and norms for parent ratings on the Aberrant Behavior Checklist-Community for young people in special education. Research in Developmental Disabilities, 23(1), 45-60. Freund, L. S., & Reiss, A. L. (1991). Rating problem behaviors in outpatients with mental retardation: use of the Aberrant Behavior Checklist. Research in Developmental Disabilities, 12(4), 435-451. Hill, J., Powlitch, S., & Furniss, F. (2008). Convergent validity of the aberrant behavior checklist and behavior problems inventory with people with complex needs. Research in Developmental Disabilities, 29(1), 45-60. Karabekiroglu, K., & Aman, M. G. (2009). Validity of the aberrant behavior checklist in a clinical sample of toddlers. Child Psychiatry and Human Development, 40(1), 99-110. Marchburn, E.C., & Aman, M.G. (1992). Factor validity and norms for the Aberrant Behavior Checklist in a community sample of children with mental retardation. Journal of Autism and Developmental Disorders, 22(3), 357-373. Miller, M. L., Fee, V. E., Jones, C. J., Miller, M. L., Fee, V. E., & Jones, C. J. (2004). Psychometric properties of ADHD rating scales among children with mental retardation. Research in Developmental Disabilities, 25(5), 477-492. Miller, M. L., Fee, V. E., & Netterville, A. K. (2004). Psychometric properties of ADHD rating scales among children with mental retardation I: Reliability. Research in Developmental Disabilities, 25(5), 459-476. Newton, J.T., Sturmey, P. (1988). The Aberrant Behavior Checklist: A British replication and extension of its psychometric properties. Journal of Mental Deficiency Research, 32(2), 87-92. Ono, Y. (1996). Factor validity and reliability for the Aberrant Behavior Checklist-Community in a Japanese population with mental retardation. Research in Developmental Disabilities, 17(4), 303-309. Paclawskyj, T. R., Matson, J. L., Bamburg, J. W., & Baglio, C. S. (1997). A comparison of the Diagnostic Assessment for the Severely Handicapped-II (DASH-II) and the Aberrant Behavior Checklist (ABC). Research in Developmental Disabilities, 18(4), 289-298. Rojahn, J., & Helsel, W. J. (1991). The Aberrant Behavior Checklist with children and adolescents with dual diagnosis. J Autism and Developmental Disorders, 21(1), 17-28. Schroeder, S. R., Rojahn, J., & Reese, R. M. (1997). Brief report: Reliability and validity of instruments for assessing psychotropic medication effects on self-injurious behavior in mental retardation. Journal of Autism and Developmental Disorders 27[1], 89-102. Walsh, K. K., & Shenouda, N. (1999). Correlations among the Reiss Screen, the Adaptive Behavior Scale Part II, and the Aberrant Behavior Checklist. American Journal of Mental Retardation, 104(3), 236-248.
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator: Must purchase the Aberrant Behavior Checklist Manual from Slosson Educational Publications, Inc. to access the clinical indicator/threshold information. This data can be purchased at http://www.slossonnews.com/ABC.html.
	Additional Notes/Links:

Tool	Tool Information
	Distributor: Stoelting (http://www.stoeltingco.com/aberrant-behavior-checklist-abc-residential-kit.html) and Slosson Educational Publications, Inc. (http://www.slosson.com/onlinecatalogstore_i1002727.html?catId=51452) Versions and Languages: ABC-Residential, ABC-Community; Teacher and parent forms for each version. Afrikaans, Chinese, Czech, Danish, Dutch, English, Finnish, Frereanch (Canadian), French (International), German, Hebrew, Hungarian, Italian, Japanese, Korean, Lithuanian, Norwegian, Portuguese, Romanian, Russian, Slovak, Slovenian, Spanish, Turkish, Telugu, and Vietnamese Copyright: Permission required to use Web Access: http://www.stoeltingco.com/aberrant-behavior-checklist-abc-residential-kit.html Email Communication: Info@StoeltingCo.com Fax: 1-630-860-9775 Contact/Availability:. Stoelting Co. 620 Wheat Lane Wood Dale, IL 60191 Phone: 1-800-860-9775 Cost: ABC residential kit (including manual and 50 residential & community forms/score sheets) costs $109.00. The ABC community kit (including manual, 50 forms/sheets and supplemental community manual) costs $125. The manual alone costs $54.00. A pack of 50 forms/sheets costs $56.00. All prices are US dollars and are accurate as of 2011.
Adult Needs and Strength Assessment (ANSA) **IT-11.27.d**	**Summary**: Purpose: The ANSA is an effective assessment tool for used in either the development of individual plans of care or for use in designing and planning systems of care for adults with mental health challenges. To administer the ANSA assessment tool found at the end of this manual, the care coordinator or other service provider should read the anchor descriptions for each dimension and then record the appropriate rating on the ANSA assessment form. One ANSA assessment tool is completed for each case reviewed. Aids service planning processes and allows for the monitoring of outcomes. Overview: The Adult Needs and Strengths Assessment (ANSA) is a multi-purpose tool developed for adult's behavioral health services to support decision making, including level of care and service planning, to facilitate quality improvement initiatives, and to allow for the monitoring of outcomes of services. The ANSA is currently used in a number of hospitals, emergency departments, psychosocial rehabilitation programs, and ACT programs. The ANSA was developed from a communication perspective so as to facilitate the linkage between the assessment process and the design of individualized service plans including the application of evidence-based practices. ANSA is comprised of five required sections – Life Domain Functioning, Strengths, Acculturation, Behavioral Health Needs, and Risk Behaviors – and one optional section (Caregiver Strengths and Needs). Scores: Needs: A **'0'** indicates *no evidence, no need for action*,

Tool	Tool Information
	A **'1'** indicates a *need for watchful waiting to see whether action is needed* (i.e., flag it for later review to see if any circumstances change or refer for assessment), A **'2'** indicates a *need for action*, and A **'3'** indicates a dangerous or disabling need or the *need for either immediate or intensive action*. Strengths: A rating of **'0'** reflects *a significant strength that is present*, A rating of **'1'** reflects *that a moderate level of the strength is present*, A rating of **'2'** reflects that *a mild level of the strength is present*, and A rating of **'3'** reflects that *the strength is not present*. For Strengths, a rating of **'0'** or **'1'** reflects *a strength that can be used to build around*, while a rating of **'2'** or **'3'** reflects *a strength that needs to be developed or identified*. Format: When the ANSA is administered, each of the dimensions is rated on its own 4-point scale after the initial intake interview, routine service contact, or following the review of a case file.
	Clinical/Quality information: Norms: As with all survey measures, careful observance of recommended sampling and survey administration protocols is essential to using the ANSA. That is, valid, reliable, and comparable-results rely on utilizing the specific protocol outlined for the ANSA. If these specifications are followed, the results will be suitable for comparisons using pre- post-test formats or cost reduction models The ANSA will replace the TRAG to assess needs, strengths, and level of care beginning September 1, 2013. Certification is required to perform the CANS. This will provide comparison groups and baseline data from participants across the state that can be utilized for evaluation purposes. http://www.dshs.state.tx.us/mhsa/trr/ansa/ Reliability: The ANSA has demonstrated reliability and validity. With training, anyone with a bachelor's degree can learn to complete the tool reliably, although some applications require a higher degree. The average reliability of the ANSA is 0.75 with vignettes, 0.86 with case records, and can be above 0.90 with live cases. The ANSA is auditable and audit reliabilities demonstrate that the ANSA is reliable at the item level. http://behavioralhealthreform.com/outcomes/cans-ansa Validity: Validity is demonstrated with the ANSA relationship to level of care decisions and other similar measures of symptoms, risk behaviors, and functioning. http://behavioralhealthreform.com/outcomes/cans-ansa From the literature: The CANS and ANSA are comprehensive assessments that are holistic and strengths-based. Each of these assessments contains core modules related to: mental health problems, life functioning, risk behaviors and strengths. Additionally, the CANS includes a core module on caregiver's strengths and needs and optional modules on transition, trauma, substance use, legal issues, acculturation and sexual aggression. The ANSA includes optional modules on caregiver's strengths and needs, cognition, trauma, substance use, employment, acculturation, danger to self/others and sexual aggression. All items contained in the CANS and ANSA will be tracked overtime using a web-based interface. Clinicians, supervisors, agency leadership and NHBBH will have access to on-demand customized reports that will allow them to monitor outcomes by item and domain at the client, clinician, program and agency levels. http://behavioralhealthreform.com/outcomes/cans-ansa
	Availability of Benchmark Data/Clinical Indicator Benchmark(s):

Tool	Tool Information
	The ANSA will replace the TRAG for assessing needs, strengths, and level of care beginning September 1, 2013. Certification is required to perform the ANSA. After this date, statewide data will be available for comparison purposes. http://www.dshs.state.tx.us/mhsa/trr/ansa/ Clinical Indicator[115]: Individual item scores are used to identify patient needs to be addressed or strengths that need to be developed. Scores are not added to develop a composite score, but individually identify areas for healthcare providers to intervene with the patient.
Children and Adolescent Needs and Strengths Assessment (CANS-MH) **IT-11.27.e**	**Additional Notes/Links** Link to Survey: http://www.praedfoundation.org/ANSA%20Form%202.0.pdf Distributer: The Praed Foundation Versions/Languages: 2.0; English Original Publication Date: 2003 Copyright:2003 Web Access: http://www.praedfoundation.org/About%20the%20ANSA.html Fax: 312.503.0466 Contact/Availability: John S. Lyons, Ph.D. Mental Health Services and Policy Program Institute for Health Services Research and Policy Studies Northwestern University 339 East Chicago Avenue, Wieboldt Bldg. 717 Chicago, Illinois 60611 312.503.0425 Cost: Free
	Summary: Purpose: The CANS–MH is designed to be used either as a prospective assessment tool for decision support during the process of planning services or as a retrospective assessment tool based on the review of existing information for use in the design of high quality systems of services. This flexibility allows for a variety of innovative applications. Overview: The CANS was developed based on communication theory. The background of the CANS comes from prior work in modeling decision-making for psychiatric services. This measure was developed to assess those dimensions crucial to good clinical decision-making for expensive mental health service interventions. We have demonstrated its utility in reforming decision making (Lyons, Mintzer, Kisiel, & Shallcross, 1998) and for quality improvement (Lyons, Kisiel, Dulcan, Chesler & Cohen, 1997; Leon, Uziel-Miller, Lyons, Tracy, 1998).

[115] http://www.praedfoundation.org/ANSA%20Manual%202.0.pdf

Tool	Tool Information
	Scores: In terms of quality improvement activities, a number of settings have utilized a fidelity model approach to look at service/treatment/action planning based on the CANS assessment. A rating of '2' or '3' on a CANS need suggests that this area must be addressed in the plan. A rating of a '0' or '1'. identifies a strength that can be used for strength-based planning and a '2' or '3' a strength that should be the focus on strength-building activities. Finally, the CANS tool can be used to monitor outcomes. This can be accomplished in two ways. First, items that are initially rated a '2' or '3' are monitored over time to determine the percent of individuals who move to a rating of '0' or '1' (resolved need, built strength). Or, dimension scores can be generated by summing items within each of the dimensions (Symptoms, Risk Behaviors, Functioning, etc.). These scores can be compared over the course of treatment. CANS dimension (domain) scores have been shown to be valid outcome measures in residential treatment, intensive community treatment, foster care and treatment foster care, community mental health, and juvenile justice programs. Format: 41 items, completed by a mental health professional, child welfare workers, teachers, or other service providers. The CANS-MH may also be used retrospectively based on archival case data. http://www.cebc4cw.org/assessment-tool/child-and-adolescent-needs-and-strengths-mental-health/ http://www.praedfoundation.org/About%20the%20CANS.html Administration: Time depends on the extent of information gathered administration time but usually approximately 10 minutes.
	Clinical/Quality information: Norms: While Child Summary Scores may theoretically range from 0 to 30, they typically fall in the range from 1 – 15, and any rating over 8 (which is the average Child Summary Score for children at discharge from psychiatric hospitalizations) will generally represent a child with very significant and complex behavioral health issues. http://www.mass.gov/eohhs/docs/masshealth/cbhi/cans-analysis-report-june2009-july2010.pdf The CANS will replace the TRAG to assess needs, strengths, and level of care beginning September 1, 2013. Certification is required to perform the CANS. This will provide comparison groups and baseline data from participants across the state that can be utilized for evaluation purposes. http://www.dshs.state.tx.us/mhsa/trr/cans/ As with all survey measures, careful observance of recommended sampling and survey administration protocols is essential to using the CANS. That is, valid, reliable, and comparable-results rely on utilizing the specific protocol outlined for the CANS. If these specifications are followed, the results will be suitable for comparisons using pre- post-test formats or cost reduction models. Reliability: There is substantial research involving the CANS. Reliability studies have demonstrated that the CANS is reliable at the item level. Training and certification is required for the use of the CANS and the recommended minimum for certification is a reliability of 0.70 using an intraclass correlation coefficient on a test vignette. Average reliability after training is approximately 0.80. Reliability on case record reviews has been demonstrated to be 0.85 while reliability with live interview strategies is above 0.90. http://www.praedfoundation.org/About%20the%20CANS.html A number of reliability studies have been accomplished using the CANS-MH including studies with a variety of practitioners and researchers. A total sample of more than 300 subjects have been included in these reliability studies. When clinical vignettes are used as the source of ratings, the average reliability

Tool	Tool Information
	across studies is 0.74. When case records or current cases are used as the source of ratings, the average reliability across studies is 0.85. In a study in Iowa, the reliability of individual items was assessed between clinicians and researchers. The average reliability of individual items of the CANS-MH was 0.73 across 40 cases. A number of different types of individual have been trained to use the CANS-MH reliably including mental health providers, child welfare case workers, probation officers, and family advocates (parents of children with difficulties). A minimum of a bachelor's degree with some training or experience with mental health is needed to use the CANS-MH reliably after training. Validity: The validity of the CANS-MH has been studied in a variety of ways. In a study in Allegheny County, Pennsylvania, the CANS was found to be significantly correlated with an independently assessed CAFAS (Rautkis & Hdalio, 2001). In this study, the Caregiver Needs & Strengths total was found to be correlated with an independent measure of burden. In a sample of more than 1700 cases in 15 different program types across the State of New York, the total scores on the dimensions of the CANS-MH (e.g. Problems, Risk Behaviors) reliably distinguished level of care . In a comparison of CANS-MH level of care guidelines to clinical judgment, staff at Multnomah County, Oregon found that the CANS-MH informed level of care criteria agreed with the expert panel decision 91% of the time. It has also been used to distinguish needs of children in rural and urban settings (Anderson & Estle, 2001). Validity has been demonstrated through the relationship of the CANS to other measures of similar constructs such as the CAFAS and CBCL. In addition, validity has been demonstrated through the relationship of the CANS to service use and outcomes. A bibliography of CANS research can be found on this website. Validity is demonstrated with the CANS relationship to level of care decisions and other similar measures of symptoms, risk behaviors, and functioning. Validity has been demonstrated through the relationship of the CANS to other measures of similar constructs such as the CAFAS and CBCL. In addition, validity has been demonstrated through the relationship of the CANS to service use and outcomes. http://www.praedfoundation.org/About%20the%20CANS.html From the literature: The CANS is a 'communimetric' measure, developed from communication theory rather than psychometric theory. Most other measures used for outcomes management purposes were developed from psychometric theories. The CANS was developed from a communication perspective so as to facilitate the linkage between the assessment process and the design of individualized service plans. The way the CANS works is that each item suggests different pathways for service planning. There are four levels of each item with anchored definitions. For needs: No evidence, Watchful waiting/prevention, Action, Immediate/Intensive Action For strengths: Centerpiece strength, Strength that you can use in planning, Strength has been identified-must be built, No strength identified. Decision support applications include the development of specific algorithms for levels of care including treatment foster care, residential treatment, intensive community services, and traditional outpatient care. Algorithms can be localized for sensitivity to varying service delivery systems and cultures. The applications of CANS-based decision algorithms have documented dramatic impacts on service system. In Illinois, use of a simple decision model for residential treatment resulted in savings of approximately $80 million per year in residential treatment in the late 1990's. In Philadelphia, their use of a decision model for Treatment Foster Care reduced lengths of stay dramatically and saved the city $11 million in the first year of use. http://www.praedfoundation.org/About%20the%20CANS.html The strength of the measurement approach has been that it is face valid and easy-to-use yet provides comprehensive information regarding the clinical status of the child or youth.

Tool	Tool Information
	Availability of Benchmark Data/Clinical Indicator Benchmark(s): The CANS will replace the TRAG to assess needs, strengths, and level of care beginning September 1, 2013. After this date, statewide data will be available for comparison purposes. http://www.dshs.state.tx.us/mhsa/trr/cans/ Clinical Indicator[116]: Individual item scores are used to identify patient needs to be addressed or strengths that need to be developed. Scores are not added to develop a composite score, but individually identify areas for healthcare providers to intervene with the patient.
	Additional Notes/Links: Link to survey: http://www.praedfoundation.org/CANS-MH%20Form.pdf Distributor: The Praed Foundation Versions and Languages: Currently, ANSA Form 2.0 is being used. Original Publication Date: The original version, the Severity of Psychiatric Illness (SPI), was created in the 1990's to study decision-making in psychiatric emergency systems. Copyright: 1999 Web Access: www.praedfoundation.org Contact/Availability: Buddin Praed Foundation at praedfoundation@yahoo.com Cost: Free
CAGE Questionnaire **IT-11.27.f**	**Summary**[117, 118,119] Purpose: Designed as an assessment tool to identify alcoholics. Overview: Four clinical interview questions, the CAGE questions, have proved useful in helping to make a diagnosis of alcoholism. The questions focus on Cutting down, Annoyance by criticism, Guilty feeling, and Eye-openers. The acronym "CAGE" helps the physician to recall the questions. Scores: Item responses on the CAGE are scored 0 or 1, with a higher score an indication of alcohol problems. A total score of 2 or greater is considered clinically significant. Administration Time: 2 minutes Format: The CAGE questions can be used in the clinical setting using informal phrasing. It has been demonstrated that they are most effective when used as part of a general health history and should NOT be preceded by questions about how much or how frequently the patient drinks
	Clinical / Quality Information[120]

[116] http://www.praedfoundation.org/CANS-MH%20Manual.pdf
[117] http://www.integration.samhsa.gov/clinical-practice/sbirt/CAGE_questionaire.pdf
[118] http://jama.jamanetwork.com/article.aspx?articleid=394693
[119] http://www.iprc.unc.edu/longscan/pages/measures/Baseline/CAGE%20Questionnaire.pdf

Tool	Tool Information
	Reliability: In a general population survey of 703 drinkers over the age of 18, a factor analysis of the CAGE indicated that the four items measure a single dimension ($\chi 2 = 1.22$, 2 df, $p = 0.54$), presumably problem drinking, and exhibit good internal reliability, factor loadings ranging from .55 to .92 (Smart et al., 1991). Validity: The CAGE has been validated in both the US (Beresford et al., 1982) and the UK (Barrison et al., 1982; Bernadt et al., 1982; King, 1986), and research using clinical samples has indicated that the CAGE identifies most alcoholics. Early studies found that the instrument functioned most effectively at a cut-off point of two or more affirmative replies with a sensitivity of 84%, a specificity of 95%, and a positive predictive value of 45% (King, 1986; Mayfield, McLeod, & Hall, 1974). A more recent critical review of the literature has suggested that lower thresholds be used for a positive screening result for women (Bradley et. al, 1998, see Score Interpretation above). Although the test is reported to reliably distinguish groups diagnosed as alcoholic from other clinical samples, it may produce a high number of false positives, especially individuals who formerly had drinking problems (Strang, Bradley, & Stockwell, 1989). In a comparison between the CAGE and the Clyde Bank Questionnaire (CBQ) (a disguised alcohol questionnaire), the CAGE correlated well with the CBQ for subjects who were chemically dependent on alcohol ($r = .77$), but poorly for problem drinkers with no symptoms of physical dependence ($r = .25$). Very few studies have examined the use of the CAGE in a general population survey and compared results with other aggregate estimates of alcohol problems (Smart et al., 1991). Saunders and Kershaw (1978, 1980) used the CAGE in a community sample and reported poor validity, with hospital records as the source of validation. However, this study was limited in that not all alcoholics or problem drinkers have clinical records confirming their behavior. It has been noted that there is probably no absolute method of validation of an alcoholism screening test, short of direct observation (King, 1986).
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator: A higher score indicates alcohol problems. A total score of 2 or greater is considered clinically significant.
	Additional Notes/Links Link to survey: https://www.pmrts.samhsa.gov/PrevCourses/ViewFile.aspx?filename=elab_supps_pg10.pdf Versions/Languages: CAGE (Alcohol specific), and the CAGE-AID (Adapted to include Drug Use)[121] Original Publication Date: 1984 Copyright: Dr. John Ewing Web Access: http://www.integration.samhsa.gov/clinical-practice/sbirt/CAGE_questionaire.pdf Cost: No other permission is necessary unless it is used in any profit-making endeavor in which case this Center would require to negotiate a payment.
	Summary[122,123]

[120] http://www.iprc.unc.edu/longscan/pages/measures/Baseline/CAGE%20Questionnaire.pdf
[121] http://www.hopkinsmedicine.org/johns_hopkins_healthcare/downloads/CAGE%20Substance%20Screening%20Tool.pdf
[122] http://whqlibdoc.who.int/hq/2001/WHO_MSD_MSB_01.6a.pdf

Tool	Tool Information
Alcohol Use Disorders Identification Test (AUDIT) **IT-11.27.h**	Purpose: Designed to identify persons with hazardous and harmful patterns of alcohol consumption, and to aid providers in screening for excessive drinking and to assist in brief assessment. Overview: The AUDIT is comprised of three domains: Hazardous Alcohol Use (3 items), Dependence Symptoms (3 items), and Harmful Alcohol Use (4 items). Scores: Responses to each question are scored from 0 to 4, giving a maximum possible score of 40. Administration Time: 2-4 minutes Format: 10 questions that are scored 0-4. Interview or self-report format with healthcare provider.
	Clinical / Quality Information[124] Norms/Factors: 2000 patients were recruited from a variety of health care facilities, including specialized alcohol treatment centers. Sixty-four percent were current drinkers, 25% of whom were diagnosed as alcohol dependent. Reliability[125]: Indices of internal consistency, including Cronbach's α and item-total correlations are generally in the 0.80's. Validity: A cut-off value of 8 points yielded sensitivities for the AUDIT for various indices of problematic drinking that were generally in the mid 0.90's. Specificities across countries and across criteria averaged in the 0.80's.
	Availability of Benchmark Data/Clinical Indicator Clinical Indicator[126]: A score of 8 or more is associated with harmful or hazardous drinking, a score of 13 or more in women, and 15 or more in men, is likely to indicate alcohol dependence.
	Additional Notes/Links Link to survey: http://www.integration.samhsa.gov/AUDIT_screener_for_alcohol.pdf Distributor: Department of Mental Health and Substance Dependence, World Health Organization Versions/Languages: Spanish, Slavic, Norwegian, French, German, Russian, Japanese, Swahili, and several other languages. Original Publication Date: 1993 Copyright: 1993 World Health Organization Web Access: http://whqlibdoc.who.int/hq/2001/who_msd_msb_01.6a.pdf

[123] http://www.ncbi.nlm.nih.gov/pubmed/8329970
[124] http://whqlibdoc.who.int/hq/2001/WHO_MSD_MSB_01.6a.pdf
[125] http://onlinelibrary.wiley.com/doi/10.1111/j.1530-0277.1997.tb03811.x/abstract
[126] http://www.agencymeddirectors.wa.gov/Files/aas.pdf

Tool	Tool Information
	Contact/Availability: 1211 Geneva 27, Switzerland Cost: Free
Patient Health Questionnaire (PHQ) **IT-11.XX**	**Summary**[127] Purpose: The Patient Health Questionnaire was designed as a self-administered method of measuring the 5 most common types of patient mental disorders: depressive, anxiety, somatoform, alcohol, and eating disorders. The PHQ-9 assesses and monitors depression severity. The PHQ-15 was created to assess somatic symptom severity and the potential presence of somatization and somatoform disorders. The PHQ-SADS assesses depressive or anxiety disorders present with somatic complaints and co-occurrence of somatic, anxiety, and depressive symptoms within primary care patients. PHQ-4: Combination for the PHQ-2 and GAD-2; brief depression and anxiety screener. Overview: PHQ: Five modules covering 5 common types of mental disorders: depression, anxiety, somatoform, alcohol, and eating. PHQ-15: Somatic symptom scale from PHQ. PHQ-9: Depression scale from PHQ. PHQ-4: PHQ-2 and GAD-2 PHQ-SADS: PHQ-9, GAD-7, and PHQ-15 measures, plus panic measure from original PHQ. Scores: Diagnostic algorithms for the PHQ can be found at http://www.phqscreeners.com/instructions/instructions.pdf PHQ-4: Scores range between 0-12. PHQ-9: This is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of ―not at all,‖ ―several days,‖ ―more than half the days,‖ and ―nearly every day,‖ respectively. PHQ-9 total score for the nine items ranges from 0 to 27. PHQ-15: This is calculated by assigning scores of 0, 1, and 2 to the response categories of ―not at all‖ , ―bothered a little‖ , and ―bothered a lot‖ , for the 13 somatic symptoms of the PHQ (items 1a-1m). Also, 2 items from the depression module (sleep and tired) are scored 0 (―not at all‖), 1 (―several days‖) or 2 (―more than half the days‖ or ―nearly every day‖). PHQ-SADS: Combines the PHQ-9, GAD-7, and PHQ-15. The GAD-7 is Seven items, each of which is scored 0 to 3, providing a 0 to 21 severity score. Administration Time: PHQ: 8 minutes Format: Clinician-administered or Self-administered.
	Clinical/Quality Information Norms/Factors: PHQ: The PHQ resulted from two large studies enrolling 6000 patients (3000 from general internal medicine and family practice clinics and 3000 from obstetrics-gynecology clinics[128]).

[127] http://www.phqscreeners.com/instructions/instructions.pdf

Tool	Tool Information
	PHQ-4: Data were analyzed from 2,149 patients drawn from 15 primary-care clinics in the United States.[129] PHQ-9: 3,000 primary care patients reported here (1,422 from 5 general internal medicine clinics and 1,578 from 3 family practice clinics).[130] PHQ-15: The PHQ-15 norms can be found at http://www.psychosomaticmedicine.org/content/64/2/258/T2.expansion.html Reliability: PHQ-9: Internal consistency of the PHQ-9 has been shown to be high. A study involving two different patient populations produced Cronbach alphas of .86 and .89.[131] PHQ-15: The internal reliability of the PHQ-15 was excellent, with a Cronbach's α of 0.80 in both the primary care and obstetrics-gynecology samples.[132] PHQ-SADS: Maintains the reliability of the GAD-7, PHQ-9 and PHQ-15. [133] Validity: PHQ-4: Factor analysis confirmed two discrete factors (Depression and Anxiety) that explained 84% of the total variance. Increasing PHQ-4 scores were strongly associated with functional impairment, disability days, and healthcare use. Anxiety had a substantial effect on functional status that was independent of depression.[134] Construct validity of the PHQ-4, PHQ-2, and GAD-2 was supported by intercorrelations with other self-report scales and with demographic risk factors for depression and anxiety. PHQ-2 and GAD-2 scores of 3 corresponded to percentile ranks of 93.4% and 95.2%, respectively, whereas PHQ-2 and GAD-2 scores of 5 corresponded to percentile ranks of 99.0% and 99.2%, respectively.[135] PHQ-9: Scores > 10 had a sensitivity of 88% and a specificity of 88% for Major Depressive Disorder. Criteria validity was established by conducting 580 structured interviews by a mental health professional. Results from these interviews showed that individuals who scored high (≥ 10) on the PHQ-9 were between 7 to 13.6 times more likely to be diagnosed with depression by the mental health professional. On the other hand, individuals scoring low (≤ 4) on the PHQ-9 had a less than a 1 in 25 chance of having depression (Kroenke et al, 2001).[136] PHQ-15: The 15 individual symptoms showed moderate associations with one other: the majority of item-item correlations in both samples were in the 0.20-to-0.29 (45%) or the 0.10-to-0.19 range (33%). Only 6% of the item-item correlations were less than 0.10, and 9% exceeded 0.40, with the highest being the correlation between trouble sleeping and fatigue (0.55). The stepwise decrements in SF-20 scores with

[128] http://www.nri-inc.org/projects/SDICC/WorkGroups/spitzer2.pdf
[129] http://www.ncbi.nlm.nih.gov/pubmed/19996233
[130] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/
[131] http://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health.aspx
[132] http://www.psychosomaticmedicine.org/content/64/2/258.long
[133] http://www.ncbi.nlm.nih.gov/pubmed/20633738
[134] http://www.ncbi.nlm.nih.gov/pubmed/19996233
[135] http://www.ncbi.nlm.nih.gov/pubmed/19616305
[136] http://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health.aspx

Tool	Tool Information
	increasing PHQ-15 scores show a consistent pattern across all six domains. Third, most pairwise comparisons within each SF-20 scale between successive PHQ-15 levels of severity were highly significant ($p < .001$). [137] PHQ-SADS: Maintains the validity of the GAD-7, PHQ-9 and PHQ-15.[138]
	Additional Notes/Links Link to survey: Patient Health Questionnaire: http://www.phqscreeners.com/pdfs/01_PHQ/English.pdf PHQ-9: http://www.phqscreeners.com/pdfs/02_PHQ-9/English.pdf PHQ-15 (Physical Symptoms): http://www.phqscreeners.com/pdfs/04_PHQ-15/English.pdf PHQ-SADS: http://www.phqscreeners.com/pdfs/05_PHQ-SADS/English.pdf PHY-4: http://www.phqscreeners.com/pdfs/08_PHQ-4/English.pdf Distributer: Pfizer Versions/Languages: Multiple languages are available at www.phqscreeners.com. If a translation is not available for a language you are interested in using, and you have the interest and resources to develop a linguistically valid translation, please send an e-mail to questions@phqscreeners.com for instructions on how to proceed. Original Publication Date: PHQ: 1999 PHQ-4: 2009 PHQ-9: 2001 Copyright: No permission is required to reproduce, translate, display or distribute. Web Access: www.phqscreeners.com Fax: Contact/Availability: questions@phqscreeners.com Dr. Spitzer at rls8@columbia.edu Dr. Kroenke at kkroenke@regenstrief.org Cost: Free

[137] http://www.psychosomaticmedicine.org/content/64/2/258.long
[138] http://www.ncbi.nlm.nih.gov/pubmed/20633738