HEALTH-RELATED QUALITY OF LIFE AFTER LIVER TRANSPLANTATION: A META-ANALYSIS

by

DENNA M. BRAVATA
INGRAM OLKIN
AMBER E. BARNATO
EMMET B. KEEFFE
DOUGLAS K. OWENS

Stanford University

Technical Report No. 1999–2
March 1999

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Health-Related Quality of Life After Liver Transplantation: A Meta-Analysis

Dena M. Bravata, M.D.\textsuperscript{1,2}  
Ingram Olkin, Ph.D.\textsuperscript{3,4}  
Amber E. Barnato, M.D., M.P.H\textsuperscript{2}  
Emmet B. Keeffe, M.D.\textsuperscript{5}  
Douglas K. Owens, M.D., M.Sc.\textsuperscript{1,2,6}

\textsuperscript{1}Department of Veterans Affairs Health Care System, Palo Alto, California. \textsuperscript{2}Center for Primary Care & Outcomes Research, Department of Medicine, Stanford University School of Medicine, Stanford, California. \textsuperscript{3}Department of Statistics, Stanford University, Stanford, California. \textsuperscript{4}School of Education, Stanford University, Stanford, California. \textsuperscript{5}Division of Gastroenterology, Department of Medicine, Stanford University School of Medicine, Stanford, California. \textsuperscript{6}Stanford Medical Informatics and Department of Health Research & Policy, Stanford University School of Medicine, Stanford, California.

Grant support: Dr. Bravata is a VA Ambulatory-Care Fellow, VA Health Care System, Palo Alto, CA. Dr. Barnato is a Agency for Health Care Research and Policy Fellow supported by a National Research Service Award. Dr. Owens is supported by a Career Development Award from the VA Health Services Research and Development Service. Dr. Olkin is supported by National Science Foundation grant DMS 96 26 265.

Correspondence: Dena M. Bravata, M.D., Section of General Internal Medicine (111A), VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304, FAX 650 852-3487, Email: bravata@smi.stanford.edu

Reprint requests: Douglas K. Owens, M.D., M.Sc., Section of General Internal Medicine (111A), VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304, FAX 650 852-3487

Running Title: Quality of Life After Liver Transplantation
ABSTRACT

Goal: To assess health-related quality of life after liver transplantation.

Materials and Methods: Structured MEDLINE and Embase literature searches identified 5473 potentially relevant articles. Thirty-two additional references were collected from articles' bibliographies. Of the total 5505 identified articles, 49 studies reporting data on 3576 transplant recipients met our inclusion criteria: They included an assessment of quality of life in adult patients, reported either pre- and post-transplantation data or had a comparison group, and were written in English. We combined post-transplantation quality of life scores from 15 studies that reported data from the same quality of life scales to assess the magnitude of the effect of liver transplantation on quality of life scales. We also performed a sign-test on all 49 studies to evaluate the direction (positive or negative) of the effect of transplantation on quality of life.

Findings: Transplantation resulted in improvement of 32% in Karnofsky scores, 11% in Sickness Impact Profile scores, and a 20% to 50% in the domains of the Nottingham Health Profile. The sign test demonstrated significant improvement in post-transplantation physical health ($p < 0.0004$), sexual functioning ($p < 0.008$), daily activities ($p < 0.02$), general HRQL ($p < 0.02$), and social functioning ($p < 0.05$); but not psychological health ($p < 0.08$).

Conclusions: In general, the health-related quality of life of the 3576 patients was impaired pre- and improved post-transplantation. Recipients reported large gains in those aspects of quality of life most affected by physical health and smaller improvements in areas affected by psychological functioning.
1. INTRODUCTION

During the past 15 years, orthotopic liver transplantation (OLT) has emerged as the treatment of choice for end-stage liver disease of various etiologies (1). Approximately 7500 transplants are performed annually at 200 international centers (2, 3). Most centers now report 1-year survival rates for their adult recipients of 80 to 90%, and 9-year survival rates of 55% (2). As the clinical outcomes of liver transplantation improve (fewer postoperative complications, more effective immunosuppression, etc.), other outcomes such as health-related quality of life (HRQL) are increasingly important targets of evaluation. Whereas considerable research has been done on the morbidity and mortality of OLT, much less is known about the influence of this procedure on transplant recipients’ HRQL (4).

Comprehensive measurement of post-transplantation HRQL is important for several reasons. First, many patients are more concerned about quality of life (QOL) than longevity (5, 6). Therefore, assessing interventions exclusively in terms of length of life, survival, or mortality does not characterize the health outcomes about which patients care (7). Second, liver-transplant recipients may suffer from debilitating fatigue, bone pain, loss of self-esteem, depression, and complications of immunosuppression. Although some of these conditions have obvious physical findings that are easily measured by traditional clinical outcomes, others have no bodily manifestations and are poorly evaluated by conventional clinical testing (8). Measuring HRQL can provide a more complete estimate of the overall health status of liver transplant recipients (8).

Third, a full accounting of an intervention’s effect on HRQL is a key component of an evaluation of both effectiveness and cost effectiveness. Given the considerable cost and scarcity of the resources involved in liver transplantation, these health-economic analyses are of particular importance in the field of liver transplantation. Finally, from a practical perspective, an understanding of HRQL post-transplantation can help health-care professionals to inform prospective transplant recipients’ decision on whether or not to have the procedure, and to design post-transplantation treatment programs that focus on those areas of recipients’ lives most affected postoperatively.
The two most commonly used methods for measuring HRQL are generic instruments for use in the general population (including utility measures and health profiles) and specific instruments for use in patients who have particular diseases. Utility measures are derived from economic and decision theories and reflect how patients value health outcomes. Utility scores are the primary means of including QOL in cost-effectiveness analyses (9, 10). The other category of generic instruments for the measurement of HRQL are health profiles. These instruments typically explore one or more of six broad domains: physical health, psychological health, social functioning, sexual functioning, ability to perform daily activities, and general well-being. (8) Health profiles vary considerably with respect to the particular aspect of HRQL they are designed to measure. Some provide scores that are dimension specific (e.g., physical functioning or emotional functioning), whereas others give an aggregate score. Although these generic profiles are appropriate for measuring elements of HRQL in the general population and between groups of patients, specific profiles may be more sensitive than generic measures to clinical changes in a particular cohort of patients (8, 11). In spite of the numerous techniques for assessing utilities and the many psychometric scales for assessing HRQL, the use of generic health profiles dominates the literature of HRQL after OLT (11).

Portions of this literature have been reviewed qualitatively (12–18); our study is the first to synthesize quantitatively the entire available English language literature on HRQL after adult liver transplantation. We sought to characterize the overall effect of liver transplantation on various domains of HRQL including patients' physical health, psychological health, social functioning, sexual functioning, ability to perform daily activities, and general well-being; and to assess the effect of transplantation on scores from the most commonly used HRQL instruments.

2. MATERIALS AND METHODS

2.1. ELIGIBILITY CRITERIA. We considered trials eligible for this meta-analysis if they (1) included an assessment of one or more of the HRQL domains of physical health,
psychological health, social functioning, sexual functioning, ability to perform daily activities, and
general well-being; (2) included pre-transplantation and post-transplantation data or compared
liver-transplant recipients to a control group; (3) were written in English; (4) and studied adult
patients.

2.2. DATA SOURCES. One of us (DMB) and a professional librarian with extensive
experience in medical literature searches independently developed search strategies to identify
studies the met the eligibility criteria. Using both strategies, we performed structured searches of
MEDLINE and Embase databases for articles published in English between January 1966 and
October 1998 that were indexed with the MeSH terms “adult” and “liver transplantation” (see
Appendix). We also read the bibliographies of retrieved articles and conference proceedings to
obtain further citations.

2.3. STUDY SELECTION AND DATA ABSTRACTION. One of us (DMB)
reviewed all titles and abstracts identified in the search for potentially relevant articles. Two of us
(DMB and AEB) independently abstracted from each relevant study data including patient
characteristics (e.g., age, gender, and indication for transplantation), study design, HRQL
instrument used, and the scores on pre- and post-transplantation HRQL scales. We resolved
disagreements between abstracters by repeated review and discussion. We then entered abstracted
data into an electronic database (Microsoft Excel 5.0, Redmond, OR, 1995).

2.4. STATISTICAL ANALYSIS. We used Microsoft Excel 5.0 for database
management and combined the data in two ways.

2.4.1. Combination by HRQL Instrument. In the first part of our analysis, we
evaluated studies that used the same HRQL instrument. We calculated an individual effect size for
each study from a difference in proportions or means between pre- and post-transplantation
groups. We weighted these individual effect sizes by the number of subjects in each study, and
used both random- and fixed-effects models to estimate an overall effect size. We performed tests
of homogeneity on all summary effects. Fifteen studies reported data from the following four
generic scales that could be combined by differences pre- and post-transplantation: the Karnofsky
Performance Status Scale, the Sickness Impact Profile, the Nottingham Health Profile, and the Medical Outcomes Survey.

The Karnofsky Performance Status Scale (KPS) was designed in 1948 as a measure of functional performance for evaluating the efficacy of cancer-chemotherapy trials (19). Today, the KPS (which requires only a short time to administer) enjoys widespread use and is often employed as a measure of HRQL (20). The KPS score derives from observable characteristics of physical functioning, as rated by a health-care professional. It ranges from 0 to 100, rated at 10-point intervals. A score of 40 to 49 indicates that the patient is disabled and requires special care and assistance; a score of 60 to 69 indicates that the patient is mostly able to perform self-care but needs occasional assistance; a score of 90 to 99 indicates that the patient is able to perform normal activities but has minor complaints; and a score of 100 indicates that the patient can perform normal activities without complaints. The major disadvantages of the KPS as a measure of HRQL are its reliance on an observer, its ceiling effect, and its lack of psychosocial or sexual dimensions (20, 21).

The Sickness Impact Profile (SIP) measures changes in behavior due to sickness (22). It is broadly applicable and was designed for measuring the outcomes of care and for monitoring patients’ clinical progress (11). The SIP is among the longest of the generic measures (136 questions measuring 12 dimensions of health). Its complicated scoring is based on a scale from 0 (best) to 15 (worst) for each quality-of-life domain. The advantages of the SIP are its widespread use, its inclusion of physical and psychological dimensions, and its internal consistency, which has been well established. It has been criticized for being too long, for omitting pain measures, for not having an established sensitivity of the functional-limitations profile to clinical change, and for using preset weights on each item that are inappropriate for certain illnesses (21, 23).

The Nottingham Health Profile (NHP) was designed to give an indication of a person’s perceived physical, social, and emotional health problems (24, 25). Unlike the SIP, on which it was based, the NHP asks about feelings and emotional states directly, rather than seeking
to assess them via changes in behavior. The emphasis is on the respondent’s subjective assessment of their health status. The multidimensional NHP includes 38 items that yield standardized scores from 0 (best) to 100 (worst) in six domains; healthy subjects should score 15 or less in each domain. The NHP is a frequently used measure (especially in Europe), it is simple to administer, and it covers a broad range of health domains. The NHP has been criticized for its bias toward severe ill health, its weighting scheme, its complexity of scoring, and its lack of a single overall score (21, 23).

The Medical Outcomes Survey (SF-36) was designed as a generic indicator of health status for use in population surveys and evaluative studies of health policy; it is applicable to a wide range of types and severities of conditions (26). It includes 36 questions that cover eight domains, with each domain yielding a score ranging from 0 (worst) to 100 (best). The acquisition of a chronic medical condition is associated with a decrement of 20 to 25 points on the physical-functioning scale; whereas serious depressive symptoms would be expected to decrease the mental-health scale by 25 to 30 points (27). Advantages of the SF-36 include its widespread use in various contexts, its validated internal consistency, and its short length. Its disadvantages include that it provides no overall score, may have floor and ceiling effects in some health domains, may not be sensitive to within subject differences over time, and does not include items about cognitive function (21).

2.4.2. Combination by Sign Test. Given that only 15 of the 49 included studies reported data from HRQL scales that could be combined using effect sizes, the second part of our analysis used a nonparametric sign test to combine all studies that presented pre-transplantation and post-transplantation data for the same HRQL domain (e.g., daily activities or sexual functioning) (28). The sign test provides an indication of the direction (positive or negative) of the effect of transplantation on that HRQL domain, but is limited in that it neither estimates the magnitude of the effect nor incorporates sample size. We classified as positive any study that reported a statistically significant improvement in an HRQL domain after transplantation, and as nonpositive any that
reported statistically insignificant improvement or statistically significant worsening. We then calculated the ratio of positive studies to all studies in a given domain. An equal number of positive and nonpositive studies indicated that transplantation has no effect on HRQL. By comparing the observed proportion of positive studies to the binomial distribution, we obtained a level of significance for each HRQL domain (29, 30).

We performed sensitivity analyses on the sign-test results to determine whether the ratio of positive studies to total studies was sensitive to particular patient-population or study-design characteristics. The variables selected for sensitivity analyses were gender, age, underlying liver disease, number of subjects, percent of eligible subjects actually enrolled, and year of publication.

3. RESULTS

Results reported in the text and tables follow the format of mean ± one standard deviation per study unless otherwise noted.

3.1. IDENTIFIED STUDIES. Our literature search identified 5473 titles of potentially relevant articles. We identified 32 additional references by manually searching the bibliographies of retrieved articles. Of the total 5505 studies, 58 met our inclusion criteria (39 from MEDLINE, 11 from Embase, and eight from manual searches) (4, 31–87). We combined multiple reports of the same patient populations; the Chicago-Hamburg group (49–51); the National Institute of Diabetes and Digestive and Kidney Diseases Liver transplantation database (NIDDK) (33, 34, 72, 87); the Vanderbilt group (46, 62); the Baylor group (52–54); and the Toronto group (32, 56). This reduced the total number of included trials to 49. The included studies present HRQL data for 3576 patients who received their liver transplants between 1963 and 1996 (Table 1).

3.2. INCLUDED STUDIES. The study designs differed considerably. Thirty-seven studies compared pre- and post-transplantation HRQL data (Chicago-Hamburg, NIDDK, Vanderbilt, Baylor, Toronto, 31, 35-36, 39, 41, 42, 44-45, 48, 57, 59-61, 63-69, 73, 75-85). Twelve of these studies and 19 others compared liver-transplant recipients with a control group
(Chicago-Hamburg, NIDDK, Toronto, 4, 37, 38, 40, 43, 45, 47-48, 55, 58, 60, 63-66, 68, 70-71, 74, 76-79, 82-86). The selection of a control group varied widely across studies (Table 2).

Of the 49 included studies, 43 reported eligibility criteria; of these, 8 used age greater than 16 years, another 8 required fluency in a particular language, and 18 used survival, with the mean minimal survival across all studies of 12 months (range 3 to 36 months). Patients were ineligible to be included in the studies for a variety of reasons, including having received a previous liver transplant (5 studies), a history of alcohol use (3 studies), a history of a psychiatric disorder (3 studies), or acute hepatic failure (4 studies).

Another source of design diversity comes from the enormous variety of HRQL instruments used. Fifty-two unique scales were used by the 49 studies. Thirty-nine studies (80%) used a validated HRQL measure. The mean number of scales used per study was 3 (range 1 to 6). Two studies did not report which HRQL instrument was used. The most frequently used measures were the Karnofsky Index (11 studies), Sickness Impact Profile (7 studies), Stait–Trait Anxiety Inventory (7 studies), Medical Outcomes Survey/SF-36 (6 studies), Nottingham Health Profile (5 studies), National Institute of Diabetes and Digestive and Kidney Disease Liver transplantation QOL survey (5 studies), Index of Well-Being (5 studies), European Organization for Research and Treatment of Cancer QOL Questionnaire (5 studies), and Psychosocial Adjustment to Illness Scale-Self Rate (5 studies). Sixteen studies used self-developed questionnaires.

Thirty-nine studies reported both the number of patients eligible to participate and the number of patients from whom they collected HRQL data. This response rate varied widely across the studies (81% ± 23%; range 28–100%). The reasons reported for not collecting data on all patients included death, loss to follow up, and failure to return mailed surveys. None of the included studies performed an intention-to-treat analysis.

3.3. PATIENT CHARACTERISTICS. The included studies presented HRQL data for an international group of 3576 patients (Table 1). Of whom, 2438 patients received transplants in 18 U.S. centers; 897 patients in 13 European centers; 180 patients in two Canadian centers; 54 patients in an Australian and 7 in a Taiwanese center.
The mean age of the patients included in the meta-analysis was 45 years (± 6.6 years). Fifty-four percent of them were men (± 21%). The patients differed considerably with respect to the length of time since their transplantation (mean 27.5 months ± 21.6 months). In general, the underlying liver diseases necessitating their transplant follow the trends of the general adult liver-transplantation population (Table 3). Among the included studies, primary biliary cirrhosis or primary sclerosis cholangitis were the most commonly reported (23% for both), with alcoholic cirrhosis (19%) and chronic active hepatitis (11%) the next most frequent. The large number of miscellaneous and unspecified diseases (44%) is a result of those studies that compared one type of transplant patients to "all others" without specifying the distribution of other diseases.

3.4. Results of Statistical Combination

3.4.1. Effect Sizes by Specific HRQL Instruments. If more than one study reported data from a particular HRQL scale, we combined those results to estimate the effect of liver transplantation on that instrument. Of the 52 HRQL measures used by the studies, pre- and post-transplantation data were reported for four scales.

3.4.1a. Karnofsky Performance Scale. Seven studies reported pre- and post-transplantation data for the KPS (31, 34-36, 39, 46, 53, 62) (Figure 1). The mean pre-transplant KPS score for 745 patients was 49 (± 15) corresponding to a functional level described as "disabled, requires special care and assistance" (20). For the 406 patients evaluated at 1 year after transplantation, the mean score was 88 (± 5) corresponding to a functional level described as "able to perform normal activities with some effort, has minor symptoms"(20). For the 216 patients evaluated more than 1 year post-transplantation, the mean score was 94 (± 5) corresponding to a functional level described as "able to perform normal activities but has minor complaints" (20).

We calculated the percent change in KPS from pre-transplantation to 1 year post-transplantation for the six studies that reported 1-year follow-up data, then combined these values using a random effects model to obtain a summary effect of transplantation on Karnofsky score at
1 year. The improvement of 31.6% (±32.1%) was not statistically significant (p ≥ 0.05). Four studies evaluated Karnofsky scores at follow-up greater than 1 year. The combined changes in these scores yielded a summary effect of 34% (± 7.8%) improvement over pre-transplantation scores (p < 0.00001).

3.4.1b. Sickness Impact Profile. Seven studies reported SIP data on 232 pre- and 378 post-transplantation patients (4, 37, 58, 62-64, 72). The mean pre-transplantation SIP score was 21.5% (± 6.1%). This reflects impairment similar to patients with oxygen dependent chronic obstructive pulmonary disease, chronic low back pain, and physically disabled adults (88). The mean SIP score improved to 8.7% (± 5.4%) post-transplantation (p < 0.01) (Table 4). This reflects impairment similar to groups of patients with hypothyroidism, Crohn’s disease, angina, and end-stage renal disease on hemodialysis (88). Four studies provided sufficient data for us to calculate a summary effect of transplantation on SIP score of 10.5% (± 4.3%) improvement (p <0.008).

3.4.1c. Nottingham Health Profile. Five studies provided raw scores from Part I of the NHP for 72 pre- and 104 post-transplantation patients (Table 5) (35, 36, 41, 58, 63). Prior to receiving their transplants, patients were impaired in all domains with the greatest impairments in energy and ability to sleep. Pre-transplantation NHP scores were similar to scores from the chronically ill elderly (92). Patients had significant improvement in all NHP domains except emotional reaction (p >0.18). Post-transplantation NHP scores were similar to those of women in the second trimester of pregnancy (92). We combined the differences of the means pre- and post-transplantation of three studies that provided sufficient NHP data (35, 36, 63). The summary effects for all domains were highly significant (p < 0.001), with the greatest improvements in energy (51±8.4, p <0.00001) and sleep (39±9, p <0.00001).

3.4.1d. Medical Outcomes Survey (SF-36). Seven studies used the SF-36 (38, 47, 48, 56, 89-91). None reported pre-transplantation data; five reported post-transplantation scores
for a total of 468 patients (Figure 2). The means for the subscales show that patients performed best on social functioning (79±12), mental health (72±8.7), role emotional (74±22), and physical functioning (67±12), and worst on role physical (50±24), vitality (54±9), bodily pain (54±7), and general health (58±11). Their scores were not significantly different from either those of the general U.S. population \((p > 0.1)\) or those of patients who had both minor and severe medical illnesses, except in the pain domain, where the transplant recipients scored lower \((p < 0.0001)\) (11, 27).

3.4.2. Combination by HRQL Domain. The results of the sign test (Figure 3) demonstrate that a statistically significant number of studies report improvement post-transplantation for physical health \((p < 0.0004)\), sexual functioning \((p < 0.008)\), daily activities \((p < 0.02)\), general HRQL \((p < 0.02)\), and social functioning \((p < 0.05)\). However, the ratio of positive to nonpositive studies was not significant for psychological health \((p < 0.08)\).

We recalculated the sign test to evaluate the sensitivity of these results to patient and study-design characteristics. The relatively small number of included studies \((n = 49)\) precluded detailed sensitivity analyses. In spite of this limitation, we found that HRQL as combined by the sign test was mildly sensitive \((p < 0.05)\) to mean age at time of transplantation and to percent of eligible subjects actually enrolled, but was insensitive \((p > 0.05\) for all domains) to gender, underlying liver disease, year of report publication, and number of subjects.

DISCUSSION

We synthesized the published literature on adults’ HRQL after liver transplantation. In general, the HRQL of the 3576 patients analyzed was impaired pre-transplantation and improved post-transplantation. The most significant improvements were in physical health, sexual
functioning, daily activities, and general QOL. Less progress was demonstrated in psychological and social functioning.

Pre-transplantation impairments in HRQL were reported in every study. The mean pre-transplantation KPS score of 49 corresponds to a functional level described as “disabled, requires special care and assistance” (20). The mean pre-transplantation SIP score of 21.5 (± 6.1) reflects greater impairment in HRQL than found among patients hospitalized for community acquired pneumonia (mean SIP 12.6), outpatients with rheumatoid arthritis (mean SIP 9.2), and the general population (mean SIP 2.6) (92). The mean pre-transplantation NHP scores were worse in all domains than the general population, the “fit elderly”, women in the first trimester of pregnancy and patients with minor non-acute conditions; they were similar to patients with peripheral vascular disease, the chronically ill elderly, and patients with osteoarthritis (25, 93).

At every level of analysis, the elements of QOL most affected by physical health showed the greatest improvement after transplantation (the 32% improvement in KPS; the energy and sleep subscores of the NHP; the role physical, vitality, and general-health subscores of the SF-36; and the sign test, $p < 0.0004$). Given the enormous influence that end-stage liver disease has on patients’ physical conditions, it is not surprising that those elements of HRQL most affected by the recipients’ physical state (e.g., the ability to perform activities of daily living or sexual activity) show the greatest improvements post-transplantation. The smaller gains in general quality of life may be secondary to these improvements in physical health. Even modest improvements in the ability to perform daily activities can result in a person feeling that her general QOL has improved, even though she has not been restored to her premorbid condition (94). For example, mitigation of sexual dysfunction (common among patients who have end-stage liver disease due to abnormalities of sex hormones) can represent a major improvement in general QOL for many patients (62, 95-98).

The psychosocial aspects of QOL consistently showed the smallest improvements post-transplantation (the social-isolation subscore of the NHP; the social functioning and role emotional subscores of the SF-36; and the sign test $p < 0.05$ for social functioning and $p < 0.08$ for
psychological functioning). Given that the mean time since transplantation for the 3576 patients was 27.9 months, most of the patients included in this analysis had long since passed the initial phase of excitement after having receiving their transplants and were living with the realities of life with a transplanted organ, which include the effects of immunosuppression, dependence on the health-care system, and potential discrimination at the workplace (60, 80, 99). In addition, several studies have attributed the persistence of psychological impairment post-transplantation to incomplete recovery from neurophysiologic damage caused by the preceding liver disease (70, 100-102).

We must consider the design limitations of both the meta-analysis and the included studies when we interpret the results of our analyses. The two major weaknesses of our meta-analysis are that we studied only articles published in English and, given that so few studies reported pre- and post-transplantation data, we were limited in our ability to combine results by mean differences and used the sign test, which, being a ratio of positive studies to all studies, is sensitive to publication bias. None of the traditional methods for assessing potential publication bias are applicable to the sign test (29, 30).

The HRQL measures used affected the potential outcomes of our meta-analysis in three ways. First, in general, the greatest benefit of the meta-analytic method of research synthesis is the ability to combine the effects of like studies and to use the power of the total number of subjects to narrow confidence intervals on a summary effect size (29, 30). Given the heterogeneity among the scales used and the paucity of studies that reported HRQL scores, we were able to perform summary-effect-size calculations for only the KPS, SIP, and NHP. Second, most of the included studies used generic (e.g., SF-36) rather than specific (e.g., QOL Index—Liver Transplant Version) HRQL instruments. Given that the HRQL of pre-transplantation patients is so profoundly disturbed, we should expect generic HRQL instruments to show a significant change in HRQL with transplantation (8, 11). However, in those HRQL domains in which no significant change occurred post-transplantation, it may be that the generic measure was insensitive to changes that would have been appreciated with a disease specific measure. Specifically, the ability of the SIP
and the NHP to identify within-patient changes over time has been controversial, with some authors demonstrating correlations between improvements in these scales and clinical change in their patients (25, 103, 104), and others showing no such correlation (105-108). Third, unfortunately, none of the included studies reported utility measures. Given that utilities are the only HRQL measures in a format useable in cost-effectiveness analyses, none of the HRQL data reported to date can be used for those important analyses. Unlike the scores from psychometric scales, utilities are based on patients' preferences. Given that individual transplant recipients' preferences about their QOL probably differ considerably, and that such differences might lead to differences in the preferred therapy, a post-transplantation program that does not consider patients' preferences may provide inappropriate treatment (9).

Some of the methodological characteristics of the included studies should be considered when attempting to generalize the results of this meta-analysis to the general liver transplant population. Certain of the inclusion criteria may have biased the results to show a more positive effect. For example, by excluding patients who were in fulminant hepatic failure, had had more than one organ transplant, had not yet survived beyond the first year, or had a history of a psychiatric disorder or alcohol use, investigators may have eliminated the pool of potentially more seriously ill patients whose HRQL scores probably would be lower than those for the general liver-transplant population. Additionally, given that 18 studies excluded patients who had not survived an average of 12 months since their transplantation, the results of this meta-analysis should be applied with caution to patients who are less than 1 year post-transplantation. The included patients are similar to the general liver-transplant population with respect to age and gender. However, given the large proportion (44%) of patients for whom no specific underlying liver disease was reported, it is difficult to know how representative the included patients were in that regard. A patient's underlying disease independently predicts his transplant-associated morbidity and mortality (9-year survival after transplantation is 74% for primary biliary cirrhosis, 70% for primary sclerosing cholangitis, 47% for hepatitis B, and 26% for malignancies) (2). Among the 2024 patients for whom the etiology of liver failure was reported, they had more of the diseases
with the best survival and fewer of the diseases with the poorest survival, a distribution that may have resulted in greater post-transplant improvements in HRQL. However, the sign-test results were insensitive to changes in underlying disease.

It is difficult to know how the variability in the proportion of eligible patients who were actually enrolled affected the results. Most articles did not provide information about the patients who were eligible for inclusion but were lost to follow-up. Given that the sign test was mildly sensitive to this response rate, the lack of information about excluded patients may represent a major limitation among the included studies.

Although we found considerable improvement in HRQL from the aggregated data of the 3576 transplant recipients, not every individual patient experienced an improvement: A subset of recipients experienced a decrement (33, 34). Knowledge of the factors related to decreased HRQL after liver transplantation in these and all other recipients would help us to develop new intervention strategies, as well as would potentially affect future allocation decisions. Thus, identification of these characteristics should be a research priority (13).

We conclude that HRQL is profoundly affected by end-stage liver disease and that liver transplantation tends to improve all domains post-transplantation, with greatest improvements in those areas most affected by physical functioning and the smallest gains in the psycho-social areas. We recommend that transplantation treatment programs expand the psychological and social support available to patients both before and after transplantation. Pre-transplantation, we hope that the results of this meta-analysis will help transplant teams to inform potential recipients about what their QOL after transplantation is likely to be. As patients are accepted onto the transplant list, programs should particularly emphasize treatments to improve disturbed sleep and low energy. Post-transplantation, programs should provide intensive support emphasizing tools for improved social functioning, pain management, and mental health.

As the first attempt to synthesize quantitatively the complete published English language literature on post-transplantation HRQL, our results were influenced by the methodologies employed by the included studies. As the demand for rigorous HRQL data increases for use in
designing comprehensive transplant programs and in cost-effectiveness analyses, future studies of liver transplant recipients should include utility-based measures of HRQL; well validated, commonly used health profiles that are disease specific and have been demonstrated to correlate with clinical change over time; and report sufficient data as to facilitate comparisons among groups of recipients.
Acknowledgments: We thank Christopher Stave and Heather Varughese for their assistance with literature searching, Richard Olshen for his review of the statistics, and Lyn Dupré for her editorial assistance.

Appendix: A professional research librarian and one of the authors (DMB) performed a literature searches to identify pertinent published data. For articles published from January 1996 to October 1998, we searched MEDLINE and Embase databases according to the following strategy:

1. S1 Liver Transplantation (All Fields); yield 16,018
2. S2 Adult (All Fields); yield 2,715,610
3. S3 S1 AND S2; yield 6,264
4. S4 English (Language); yield 7,330,096
5. S5 S3 AND S4; yield 5,473
References


44. Esquivel CO, Van Thiel DH, Demetris AJ. Transplantation for primary biliary cirrhosis. Gastroenterology. 1988;94(5 Pt 1):1207-16.


<table>
<thead>
<tr>
<th>Location</th>
<th>Reference</th>
<th>Number</th>
<th>Study Site</th>
<th>Number</th>
<th>Reference</th>
<th>Stadiumpbw</th>
<th>Number</th>
<th>Reference</th>
</tr>
</thead>
</table>

Table 1. Summary of Study Characteristics.
<table>
<thead>
<tr>
<th>Location</th>
<th>Reference</th>
<th>Number</th>
<th>Study Dates</th>
<th>Mean (months)</th>
<th>Follow-up</th>
<th>Age % Male</th>
<th>Eligible</th>
<th>Included</th>
<th>Enrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiscosin</td>
<td>Kuchel</td>
<td>(48)</td>
<td>1976 - 1984</td>
<td>56</td>
<td>30</td>
<td>12.0</td>
<td>166</td>
<td>83%</td>
<td>73%</td>
</tr>
<tr>
<td>Michigan</td>
<td>Neves</td>
<td>(49)</td>
<td>1976 - 1984</td>
<td>56</td>
<td>48</td>
<td>30.0</td>
<td>32</td>
<td>73%</td>
<td>69%</td>
</tr>
<tr>
<td>Detroit</td>
<td>L-J</td>
<td>(4)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>29.0</td>
<td>30.0</td>
<td>30</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Chicago</td>
<td>L-J</td>
<td>(27)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>41.0</td>
<td>36.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>San Francisco</td>
<td>Chin</td>
<td>(16)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>39.0</td>
<td>36.0</td>
<td>42</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>VA Mississippi</td>
<td>Vincent</td>
<td>(69)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>30.0</td>
<td>38.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Cook County</td>
<td>Drexel</td>
<td>(3)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>0.0</td>
<td>36.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Cleveland University</td>
<td>Cleveland</td>
<td>(46, 27)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>36.0</td>
<td>42.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Minneapolis/St. Paul</td>
<td>Mason</td>
<td>(89)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>30.0</td>
<td>31.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>San Francisco</td>
<td>Drexel</td>
<td>(70)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>21.0</td>
<td>28.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Dallas</td>
<td>Pavone</td>
<td>(68)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>28.0</td>
<td>26.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Houston</td>
<td>Pavone</td>
<td>(68)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>24.0</td>
<td>24.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>San Francisco</td>
<td>Chung</td>
<td>(5)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>29.0</td>
<td>26.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>New York</td>
<td>Robinson</td>
<td>(5)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>30.0</td>
<td>30.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>Robinson</td>
<td>(5)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>30.0</td>
<td>30.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>Kumer</td>
<td>(48)</td>
<td>1976 - 1984</td>
<td>30</td>
<td>30.0</td>
<td>30.0</td>
<td>12</td>
<td>47%</td>
<td>47%</td>
</tr>
</tbody>
</table>

Table 1: Continued. Summary of Study Characteristics.
<table>
<thead>
<tr>
<th>Type of comparison group used</th>
<th>Reference(s)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transplanted for alcoholic liver disease vs. for other causes of liver failure</td>
<td>7 (40, 48, 60, 71, NIDDK, 77, 84)</td>
</tr>
<tr>
<td>Immunosuppression by cyclosporine vs. tacrolimus</td>
<td>2 (65, 108)</td>
</tr>
<tr>
<td>Transplanted for cancer vs. other causes of liver failure</td>
<td>2 (37, 82)</td>
</tr>
<tr>
<td>Recipients less than 60 years old vs. 60 years or older</td>
<td>1 (87)</td>
</tr>
<tr>
<td>Transplanted less than 2 years prior to evaluation vs. more than 2 years prior to evaluation</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Patients who required an intensive-care unit vs. those who did not</td>
<td>1 (66)</td>
</tr>
<tr>
<td>Transplanted for hepatitis B vs. hepatitis C</td>
<td>1 (43)</td>
</tr>
<tr>
<td>OLT* recipients vs. the general U.S. population</td>
<td>10 (38, 45, 47, Chicago-Hamburg, 55, 58, 70, 74, 76, 89)</td>
</tr>
<tr>
<td>OLT recipients vs. patients with end-stage liver disease</td>
<td>4 (45, Chicago-Hamburg, 55, 78)</td>
</tr>
<tr>
<td>OLT recipients vs. patients with gastrointestinal or other medical illness</td>
<td>2 (38, 70)</td>
</tr>
<tr>
<td>OLT recipients vs. recipients of other organ transplants</td>
<td>2 (56, 64)</td>
</tr>
<tr>
<td>OLT recipients vs. patients who were either rejected from the OLT list or are waiting for a transplant</td>
<td>1 (63)</td>
</tr>
</tbody>
</table>

**Total** 31 (34)†

*OLT = orthotopic liver transplantation
†The following studies compared OLT recipients to two comparison groups (Chicago-Hamburg, 37,69).
Table 3. Underlying liver disease in the included studies and in the general U.S. adult liver-transplant population.

<table>
<thead>
<tr>
<th>Primary liver disease</th>
<th>Percent of general transplant population*</th>
<th>Number of patients</th>
<th>Percent of included patients</th>
<th>Number of patients</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-1-antitrypsin</td>
<td>4.4</td>
<td>900</td>
<td>0.2</td>
<td>7</td>
<td>ns</td>
</tr>
<tr>
<td>Fulminant hepatic failure</td>
<td>6</td>
<td>1,227</td>
<td>0.9</td>
<td>32</td>
<td>ns</td>
</tr>
<tr>
<td>Malignancy</td>
<td>4.2</td>
<td>852</td>
<td>1.5</td>
<td>54</td>
<td>ns</td>
</tr>
<tr>
<td>Metabolic disorders</td>
<td>6.5</td>
<td>1329</td>
<td>1.6</td>
<td>57</td>
<td>ns</td>
</tr>
<tr>
<td>Chronic active hepatitis</td>
<td>25.8</td>
<td>5277</td>
<td>11.2</td>
<td>401</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Alcoholic liver disease</td>
<td>21.4</td>
<td>4,382</td>
<td>18.8</td>
<td>672</td>
<td>ns</td>
</tr>
<tr>
<td>PBC/PSC§</td>
<td>19.1</td>
<td>3,898</td>
<td>22.4</td>
<td>801</td>
<td>&lt;0.02</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>12.6</td>
<td>2,577</td>
<td>43.4</td>
<td>1,552</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>20,452</strong></td>
<td><strong>100</strong></td>
<td><strong>3,576</strong></td>
<td></td>
</tr>
</tbody>
</table>

*As reported by the Pitt-United Network of Organ Sharing Liver transplant registry for patients who received liver transplants in the United States between 1988 and 1996 (2).

†These p values reflect the level of significance for test of difference between the proportions of the general liver-transplant population and the included studies.

§PBC = Primary biliary cirrhosis; PSC = primary sclerosing cholangitis
<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Mean*</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dickson</td>
<td>30</td>
<td>12.4</td>
<td>7.9</td>
<td>30</td>
<td>12.4</td>
<td>7.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hicks</td>
<td>35</td>
<td>7</td>
<td>9</td>
<td>35</td>
<td>7</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moore 1997</td>
<td>32</td>
<td>25.3</td>
<td>10</td>
<td>32</td>
<td>10</td>
<td>10</td>
<td>10.2</td>
<td>10</td>
</tr>
<tr>
<td>Riether</td>
<td>17</td>
<td>15.35</td>
<td>12.4</td>
<td>10</td>
<td>15.35</td>
<td>12.4</td>
<td>-5.7</td>
<td>13.2</td>
</tr>
<tr>
<td>Tarter 1984</td>
<td>10</td>
<td>14</td>
<td></td>
<td>10</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarter 1991</td>
<td>53</td>
<td>13.6</td>
<td>12</td>
<td>53</td>
<td>13.6</td>
<td>12</td>
<td>-16.2</td>
<td>7.9</td>
</tr>
<tr>
<td>Tarter 1992</td>
<td>130</td>
<td>7</td>
<td>7.5</td>
<td>130</td>
<td>7</td>
<td>7.5</td>
<td>-12.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Total/Mean</td>
<td>307</td>
<td>11.2</td>
<td>8.9</td>
<td>300</td>
<td>11.2</td>
<td>8.9</td>
<td>10.5</td>
<td>4.3</td>
</tr>
</tbody>
</table>

* A negative value indicates a worse score post-transplantation.
†This mean is weighted by the number of patients in each study.
§The summary effect was obtained by combining the individual change in SIP scores by a random effects model (p < 0.008).
<table>
<thead>
<tr>
<th>Domain</th>
<th>Pre-transplantation (n = 72)</th>
<th>Post-transplantation (n = 54)</th>
<th>Change in scores $^\S$</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>$P$ value $^\dagger$</th>
<th>Mean</th>
<th>SD</th>
<th>$P$ value $^\S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility</td>
<td>29.6</td>
<td>16</td>
<td>13.3</td>
<td>18</td>
<td></td>
<td>&lt;0.005</td>
<td></td>
<td>20.9</td>
<td>5.1</td>
<td></td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Pain</td>
<td>24.4</td>
<td>14</td>
<td>8.2</td>
<td>21</td>
<td></td>
<td>&lt;0.028</td>
<td></td>
<td>20.4</td>
<td>6.8</td>
<td></td>
<td>&lt;0.002</td>
</tr>
<tr>
<td>Energy</td>
<td>59.9</td>
<td>23</td>
<td>19.7</td>
<td>24</td>
<td></td>
<td>&lt;0.000</td>
<td></td>
<td>51.3</td>
<td>8.4</td>
<td></td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Sleep</td>
<td>44.0</td>
<td>20</td>
<td>17.6</td>
<td>16</td>
<td></td>
<td>&lt;0.002</td>
<td></td>
<td>39.4</td>
<td>6.6</td>
<td></td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Social Isolation</td>
<td>18.1</td>
<td>13</td>
<td>11.9</td>
<td>10</td>
<td></td>
<td>&lt;0.023</td>
<td></td>
<td>14.7</td>
<td>4.7</td>
<td></td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Emotional Reaction</td>
<td>22.3</td>
<td>13</td>
<td>8.9</td>
<td>13</td>
<td></td>
<td>ns</td>
<td></td>
<td>21.0</td>
<td>6.0</td>
<td></td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>

*$SD$ = standard deviation  
$^\dagger$Two tailed Student's $t$-test  
$^\S$Difference between pre- and post-transplantation scores weighted by the number of subjects at each time period
FIGURE LEGENDS

Figure 1 Karnofsky scores pre- and post-transplantation. The Karnofsky scores at pre-transplantation, 1 year post-transplantation and more than 1 year post-transplantation are shown. Pre-transplantation scores are denoted by a square, 1 year scores by a triangle and scores beyond one year by a circle. The mean score is weighted by the number of patients for whom Karnosky scores were available at each time point.

Figure 2 Medical Outcome Survey (SF-36) scores post-transplantation. The post-transplantation SF-36 scores for individual studies are denoted by black squares with a 95% confidence interval noted by the top and bottom of the error bar. SF-36 data for post-transplantation patients are presented for each HRQL domain. Also shown are scores of the general U.S. population and patients who had minor and serious medical problems (11, 27).

Figure 3 Summary results of sign test. The numbers of positive studies for each quality-of-life domain are represented by black bars; those for nonpositive studies are represented by white bars.