

Differences in Quality of Life Across Renal Replacement Therapies: A Meta-Analytic Comparison

Jill I. Cameron, MSc, Catherine Whiteside, MD, PhD, Joel Katz, PhD, and Gerald M. Devins, PhD

• A meta-analysis compared emotional distress and psychological well-being across renal replacement therapies (RRTs) and examined whether differences could be explained by: (1) treatment modalities, (2) case mix, or (3) methodologic rigor. Standard meta-analytic procedures were used to evaluate published comparative studies. Successful renal transplantation was associated with: (1) lower distress (effect size, $d = -0.43$ SD) and greater well-being ($d = 0.62$ SD) than incenter hemodialysis (CHD) and (2) lower distress ($d = -0.29$ SD) and greater well-being ($d = 0.53$ SD) than continuous ambulatory peritoneal dialysis (CAPD). CAPD was characterized by greater well-being ($d = 0.18$ SD) than CHD and CHD was associated with greater distress ($d = 0.16$ SD) than home hemodialysis. Although methodologic rigor and case-mix differences did not correlate with the magnitude of psychosocial differences across RRTs, 10 of the 12 comparisons (83%) were threatened by publication bias (ie, that nonsignificant comparisons may have been underrepresented in the published literature). Thus, although significant quality-of-life differences were evident across treatment groups, the types of patients representative of the various RRTs also differed significantly in terms of case-mix variables relevant to psychosocial well-being and emotional distress. Published findings indicating differential quality of life across RRTs may thus be attributable to: (1) valid differences in effective renal replacement, reduced medical complications, and lifestyles afforded by these treatment modalities; (2) case-mix differences in the patient samples selected to represent them in research comparisons; or (3) both of these alternative explanations.

INDEX WORDS: Quality of life; renal replacement therapy (RRT); meta-analysis; comparative studies; psychosocial; case mix.

SUCCESSFUL RENAL transplantation (RT) is generally accepted as the preferred treatment for end-stage renal disease (ESRD). Among all renal replacement therapies (RRTs), successful transplantation most closely replicates the normal process of waste removal and precludes known shortcomings of maintenance dialysis (eg, medical complications, such as hyperparathyroidism and transient azotemia). In addition, it eliminates the constraints imposed by maintenance dialysis and, as a result, many assert that successful transplantation affords a better quality of life.¹ Although this position is supported by a growing consensus in the published literature, findings are inconsistent (eg, Bremer et al,² Evans et al,³ and Simmons et al⁴ versus Devins et al^{5,6} and Johnson et al⁷). It is also unclear whether this quality-of-life superiority is due to genuine benefits of the treatment or preexisting nontreatment differences between groups. Patients are not randomly assigned to RRTs and, as a result, treatment groups frequently differ with respect to many characteristics relevant to quality of life (eg, age, socioeconomic status, and nonrenal health).⁸ Research findings are ambiguous when

investigators do not take such case-mix differences into consideration.⁹ Other method weaknesses that threaten the validity of quality-of-life comparisons across RRTs include reliance on samples of convenience and the use of new and unvalidated measurement instruments.¹⁰

A comprehensive review of this literature can establish the extent to which quality of life differs across alternative RRTs. The traditional approach to reviewing empirical literature is the narrative review (eg, De-Nour¹¹). Unfortunately, the breadth and comprehensiveness of narrative reviews are limited by difficulties inherent in the challenges of identifying and integrating the findings of independent studies, especially when the

question of interest has been studied for many years and by numerous investigators. Narrative reviews often focus on a select or nonrepresentative sample of published articles, potentially reflecting the reviewer's unique perspective and/or biases.¹² They typically exclude unpublished studies, especially those that report statistically nonsignificant results (because these are difficult to locate); non-English-language publications are often underrepresented; and relevant articles are often overlooked despite powerful computerized literature search engines, such as MEDLINE. Given their impressionistic integration and interpretation of research findings, it is difficult to evaluate the extent to which the validity of observations incorporated into a narrative literature review may have been compromised by extraneous influences (eg, case-mix differences). Finally, reliance on significance tests to indicate the reliability of treatment-group differences allows studies with large sample sizes to exert a disproportionate effect on the interpretation of results and thus the conclusions reached by the narrative review.

Meta-analysis¹³ provides a quantitatively more precise research-synthesis strategy that overcomes many of the weaknesses intrinsic to traditional narrative reviews. Meta-analysis endeavors to be comprehensive, and reviewers are encouraged to include all research examining the question of interest. If limited to published studies, it is possible that statistically significant differences may be overrepresented because of editorial biases to accept only positive findings (this phenomenon has been labeled "publication bias"¹⁴). Strategies exist (eg, calculating fail-safe numbers and tolerance levels^{13,14}) to estimate whether and to what extent the validity of results may be threatened by this problem, and methodologists have emphasized the need to incorporate such considerations in the interpretation of meta-analyses.^{12,15} Quantitative effect-size (ES) estimates facilitate direct comparisons between studies by standardizing results (eg, differences between group means) to facilitate comparisons based on different measurement instruments and varying statistical techniques.¹³ In addition, meta-analysis can provide quantitative estimates of the degree to which study characteristics may be related to the magnitude of treatment effects, examining, for example, whether and to what extent differences in research de-

signs, methodologic rigor, or case-mix differences are related to quality-of-life differences between dialysis patients and transplant recipients.

THE PRESENT STUDY

A meta-analysis was conducted to examine differences in quality of life across alternative treatment modalities for ESRD. All published research comparing RRTs in terms of emotional distress and/or psychological well-being were systematically retrieved from the literature. The analysis was designed to address the following research questions: Do psychological well-being and/or emotional distress differ across alternative RRTs and, if they do, to what extent can differences be explained by (1) differences in case mix across treatment groups or (2) differences in methodologic rigor across studies?

METHODS

Study Retrieval

Relevant studies were identified through MEDLINE, PsychINFO, CINAHL, personal reference databases, and the bibliographies of retrieved articles. Studies were included in the analysis if they met the following criteria: (1) they reported at least one quantitative comparison between at least two modes of treatment, ie, RT, hospital-based incenter hemodialysis (CHD), continuous ambulatory peritoneal dialysis (CAPD), and home hemodialysis (HHD); (2) included at least one measure of psychological well-being or emotional distress; (3) used a prospective research design, including repeated-measures, an equivalent-comparison group, or an independent-comparison group; (4) involved adult patients (aged ≥ 18 years); and (5) reported the information necessary to calculate directly or to estimate ESs. Studies were excluded if they met one or more of the following criteria: (1) quality of life on one's current mode of treatment was compared retrospectively with an earlier modality and no other comparative data were reported, (2) treatment included something other than conventional RRT (eg, drug trials or multiorgan transplantation), and (3) data concerning psychosocial outcomes were provided by someone other than the patient (eg, nurse or family member).¹⁶

In meta-analysis, data from research participants can only be included once for each dependent variable. We therefore evaluated the independence of research samples by reviewing articles with similar or overlapping investigators to determine whether the same groups of subjects might have been used in more than one publication. This was achieved by determining when, where, and how many subjects were included in a study. If more than one published article reported findings based on the same subjects, the articles were treated as a single study and all the publications in the series were used to obtain the required information.

Dependent Variables

Two complementary psychosocial constructs were examined to represent quality of life in ESRD: emotional distress

(eg, depression, anxiety, negative affect) and psychological well-being (eg, life satisfaction, happiness, positive affect). Current consensus in the psychology of emotion maintains that positive and negative affect are statistically independent and are best represented by distinct unipolar measures. Consistent with this assertion is that nonoverlapping determinants have been identified for each of these states.^{6,17-20} Although positive and negative affect frequently show complementary patterns of association, their distinctiveness has been observed in quality-of-life research comparing alternative RRTs in which dialysis and posttransplantation patients differed significantly in psychological well-being but not emotional distress.^{5,6,21}

Instruments commonly used to measure emotional distress include the Center for Epidemiologic Studies Depression Scale,²² the Beck²³ Depression Inventory, and the Profile of Mood States.²⁴ Common measures of psychological well-being include the Bradburn¹⁸ Affect Balance Scale, the Campbell²⁵ Index of Well-Being, and the Cantril²⁶ Ladder (Life Satisfaction).

Independent Variables

The meta-analysis was designed to investigate the extent to which emotional distress and psychological well-being were related to the following independent variables: (1) mode of RRT, (2) case mix, (3) methodologic rigor, and (4) research design. Case mix refers to preexisting differences across treatment groups that are independent of RRT-related differences (eg, age, sex, education, socioeconomic status, marital status, physical health, duration of treatment and/or illness). Methodologic rigor relates to the validity of the research method. It was evaluated using an 11-item checklist developed for the present study (eg, use of valid and reliable measurement instruments, use of sampling controls, and examination of and/or control for case-mix differences). We also included the Science or Social Science Citation Index impact score associated with the scholarly outlet in which the study was published as a complementary index of methodologic rigor (this score indicates the frequency with which articles published in a particular outlet are cited in other published papers).

Meta-Analytic Methods

A standardized data extraction protocol obtained all information necessary to calculate ESs for each article included in the meta-analysis. Two independent reviewers evaluated the precision of the data extraction form and obtained 79% agreement. Reliability for each of the items used to indicate methodologic rigor was very good; κ coefficients ranged from 0.82 to 1.0.

The random-effects model was the primary statistical model adopted because study characteristics are not consistent across publications. The fixed-effects model was used to examine the impact of study characteristics, methodologic rigor, and case-mix differences on ESs because of the availability of automated standardized procedures using conventional statistical computer packages (eg, SPSS²⁷ [SPSS Inc, Chicago, IL]). The two models only differ when there is a large amount of between-study variability, in which case the random-effects model takes this variability into consideration and produces more conservative results.

ESs were represented by Cohen's *d* to estimate the standardized mean difference between treatment groups expressed in SD units.^{13,28-30} Adjusted ESs were calculated to control for potential overestimation caused by small sample sizes.²⁸

Because information is not reported uniformly across studies, we calculated ESs using (1) the standardized mean difference method, (2) Cohen's *d* for categorical data, (3) significance tests, or (4) significance levels. For repeated-measures and equivalent-group designs, significance tests or levels were not used to avoid overestimation of the ES.³¹ When all the information necessary to calculate an individual ES was not available, we calculated an estimate using standard methods.³⁰ Because each treatment comparison can contribute only one ES for each dependent variable, ESs were averaged for studies using more than one measure of the same dependent variable (ie, multivariate measurement).³²

Statistical Analyses

We calculated summary ESs and 95% confidence intervals (CIs) for each dependent variable for which there were five or more studies comparing particular treatment modalities.¹³ We calculated fail-safe numbers and tolerance levels to evaluate the threat of publication bias (that studies showing statistically nonsignificant treatment differences may not have been published), following established methods.¹⁴ The fail-safe number indicates how many unpublished studies with null results would be necessary to render statistically significant meta-analytic ESs nonsignificant. The tolerance level (five times the number of studies included in the meta-analysis plus 10) is a method of evaluating the strength of the fail-safe number and is the number of unpublished studies that might reasonably be expected to exist. According to recognized criteria, the validity of meta-analytic results is threatened by publication bias whenever the tolerance level exceeds the fail-safe number.¹⁴ Such circumstances leave open the possibility that publication bias threatens the validity of meta-analytic findings because the plausible number of unpublished negative findings exceeds the number required to reduce a statistically significant ES to nonsignificant (this problem has also been referred to as the file-drawer problem).

We undertook sensitivity analyses³³ using weighted analysis of variance and correlation to determine whether observed treatment-group differences were systematically related to study characteristics, such as case mix, research design, or methodologic rigor. These analyses were intended to indicate whether the magnitude of treatment-group differences were related systematically to characteristics of the studies, including group differences in case-mix variables. First, we examined research characteristics (eg, research design, total sample size, year of publication, and method of ES estimation), common sources of variability,³⁴ to determine, for example, if one type of research design showed larger or smaller group differences in the dependent variables. Second, we examined case-mix differences through sensitivity analyses that compared ESs across studies in which the treatment groups differed or did not differ with

respect to each of the case-mix variables. Third, we correlated degree of method rigor with ESs (eg, might larger group differences be evident in methodologically weaker studies?).

To facilitate the interpretation of results, we determined their associated percentile ranks by referring the ES statistic to the normal distribution. This transformation can be interpreted to indicate how highly a person scoring at the median in one treatment group would rank in the other treatment group for a given variable.¹²

RESULTS

Identification of Published Reports

The literature search strategy identified 3,267 studies published before August 1998, and we reviewed titles and abstracts to identify publications satisfying the inclusion and exclusion criteria. Sixty-one published articles (1.9%) met these conditions and were included in the meta-analysis. We excluded publications for the following reasons: (1) 2,761 articles (84.5%) did not measure a psychosocial variable, although they had quality-of-life implications; (2) 95 articles (2.9%) did not report new data (eg, literature reviews, editorials); (3) 61 articles (1.9%) were drug studies (eg, investigating erythropoietin or cyclosporine); (4) 25 articles (0.8%) involved children; (5) 25 articles (0.8%) concerned survival or physical symptoms; (6) 190 articles (5.8%) reported psychosocial measurements for a single or combined treatment group; (7) 13 articles (0.4%) overlapped with another published article; and (8) 36 articles (1.1%) were excluded for a variety of other reasons (eg, retrospective design).

For studies not providing all the information necessary to calculate an ES, we contacted the authors by mail and requested the missing information. Of the 44% who responded, 41% (or 18% overall) provided additional information.

As a result, it was necessary to exclude 13 of the 61 retrieved studies (21%) from the meta-analysis because of insufficient information to estimate ESs. The final number of studies included in the analyses was thus 49. This included 77 treatment comparisons involving emotional distress and 66 involving psychological well-being.

Treatment Comparisons of Emotional Distress and Psychological Well-Being

The meta-analytic results for the treatment comparisons involving emotional distress and psychological well-being are listed in Tables 1 and 2. These tables present random-effects model summary results for each treatment comparison, including summary ESs, 95% CIs, percentile ranks, fail-safe numbers, and tolerance levels to evaluate the threat of publication bias. ES estimates are significantly different from 0 (indicating a reliable difference between the treatment groups) if the corresponding 95% CI does not include 0. The percentile rank indicates how an individual scoring at the median in one group (ie, 50th percentile) would score were he or she a member of the other treatment group. For example, the corresponding percentile rank for an ES of 1.0 would be 84, indicating that an individual reporting quality of life at the median for one treatment group would have scored at a level greater than 84% of the patients in the other group.

Emotional distress. Table 1 lists the treatment comparisons regarding emotional distress. Successful renal transplant recipients reported significantly less emotional distress than patients undergoing CHD (mean ES = -0.43; n = 25; 95% CI, -0.53 to -0.33) and CAPD (mean ES =

Table 1. Summary Results of Treatment Comparisons of Emotional Distress

Comparison	Mean ES	95% CI	Percentile Rank	Fail-Safe Number	Tolerance Level
RT v CHD (25)	-0.43*	-0.53 to -0.33	33	79	135
RT v CAPD (10)	-0.29*	-0.55 to -0.03	39	2	60
RT v HHD (7)	-0.21	-0.49 to 0.07	42	0	45
CAPD v CHD (17)	-0.09	-0.29 to 0.11	46	0	95
CAPD v HHD (7)	0.09	-0.25 to 0.43	54	7	45
CHD v HHD (11)	0.16*	0.07 to 0.24	56	70	65

NOTE. Number in parentheses indicates number of studies summary results based on.

*Effect sizes significantly different from 0.

Table 2. Summary Results of Treatment Comparisons of Psychological Well-Being

Comparison	Mean ES	95% CI	Percentile Rank	Fail-Safe Number	Tolerance Level
RT vCHD (16)	0.62*	0.46 to 0.78	73	201	90
RT vCAPD (11)	0.53*	0.30 to 0.76	70	49	65
RT vHHD (7)	0.66	-0.09 to 1.41	75	0	45
CAPD vCHD (18)	0.18*	0.07 to 0.28	57	28	100
CAPD vHHD (7)	-0.06	-0.45 to 0.34	48	0	45
CHD vHHD (7)	-0.19	-0.73 to 0.35	42	0	45

NOTE. Number in parentheses indicates number of studies summary results based on.

*Effect sizes significantly different from 0.

-0.29; $n = 10$; 95% CI, -0.55 to -0.03). CHD patients reported significantly more emotional distress than HHD patients (mean ES = 0.16; $n = 11$; 95% CI, 0.07 to 0.24). No other group comparisons were statistically significant. All but one of the comparisons, however, may be threatened by publication bias. Because the tolerance level (65) was lower than the fail-safe number (70), the comparison of CHD versus HHD may not be threatened by publication bias.

Psychological well-being. Table 2 lists the treatment comparisons for psychological well-being. Successful renal transplant recipients reported significantly more psychological well-being than patients undergoing CHD (mean ES = 0.62; $n = 16$; 95% CI, 0.46 to 0.78) and CAPD (mean ES = 0.53; $n = 11$; 95% CI, 0.30 to 0.76). CAPD patients reported significantly more psychological well-being than CHD patients (mean ES = 0.18; $n = 18$; 95% CI, 0.07 to 0.28). No other significant group differences were observed. As noted for the comparisons involving emotional distress, however, all but one of the treatment comparisons were threatened by publication bias. The fail-safe number (201) exceeded the tolerance level (90) for the comparison of renal transplant versus CHD patients.

Sensitivity Analyses

Research characteristics. We examined several characteristics of the published literature to determine whether study characteristics might be related systematically to the magnitude of reported results. None of these comparisons, however, indicated a significant relation between study characteristics and the magnitude of treatment-group differences in psychosocial outcomes. Across all studies comparing one or both

dependent variables, the research designs included 89 studies (88.1%) with an independent-group design, 6 studies (5.9%) with equivalent- or matched-group designs, and 6 studies (5.9%) with a prospective repeated-measures design. Results indicated that neither emotional distress nor psychological well-being ESs differed significantly across these three categories. ESs did not differ significantly whether they were calculated directly or estimated because of incomplete reporting of statistical results. The methodologic rigor of the study, assessed by the checklist developed for this analysis and the Citation Index impact score, did not significantly relate to the ES estimates when examined across all treatment comparisons. Additionally, the total sample size and year of publication were not significantly associated with the ES estimates.

Case-mix differences. Case-mix differences between treatment groups were examined as potential sources of variability across ESs. To facilitate the sensitivity analyses, case-mix variables were coded 0 when the treatment comparison did not significantly differ with respect to the variable, 1 when they differed significantly, and 2 when the case-mix variable was not assessed. Sensitivity analyses were conducted to determine whether ESs varied with respect to the categorized case-mix treatment-group differences. Six of the 120 sensitivity analyses (5.8%) were significant at P less than 0.05, but because of the very large number of comparisons, these must be considered skeptically. We conclude, therefore, that ESs did not vary systematically across studies depending on whether they differed with respect to case mix.

These results are limited, however, because of the relatively small numbers of studies compar-

Table 3. Summary of Differences Between Treatment Groups on Case-Mix Variables, Emotional Distress Treatment Comparisons

Comparison*	Age†	Physical Indicators	Employment	Education	Sex (men)
RT vCHD (25)	-0.39‡ (19)	-0.49‡ (14)	0.28‡ (13)	0.27‡ (9)	0.03 (20)
RT vCAPD (10)	-0.88‡ (6)	-0.46‡ (7)	0.54‡ (4)	0.11 (4)	0.17‡ (8)
RT vHHD (7)	-0.54‡ (6)	-0.41‡ (6)	0.14 (3)	0.66‡ (4)	0.04 (6)
CAPD vCHD (17)	-0.05 (12)	-0.21‡ (10)	0.16 (8)	0.23‡ (7)	0.07 (15)
CAPD vHHD (7)	0.28‡ (6)	0.12 (5)	-0.58 (2)	0.07 (4)	-0.12 (6)
CHD vHHD (11)	0.19 (8)	0.22‡ (7)	-0.26 (3)	-0.37‡ (5)	-0.11 (7)

*Treatment comparison (number of studies examining emotional distress or psychological well-being).

†Mean effect sizes (number of studies included in mean).

‡Mean effect size is significantly different from 0; $P < 0.05$.

ing RRTs with respect to emotional distress and psychological well-being (ie, the number of treatment comparisons for each dependent variable ranged between 7 and 25). Moreover, few studies provided the information necessary to calculate ESs for the case-mix variables. Nevertheless, significant case-mix differences were evident across the treatment groups. Tables 3 and 4 list the differences between groups regarding fundamental case-mix variables (age, physical indicators, employment status, educational achievement, and sex). The tables present mean ESs, their statistical significance, and the number of studies reporting information for each variable. In addition to the already reported differences in psychological well-being and emotional distress, comparisons between RT and CHD were confounded by significant group differences in age, physical indicators (eg, comorbid conditions, serum electrolyte levels), educational attainment, and employment status. Transplant recipients were significantly younger, healthier, more highly educated, and more likely to be employed than

CHD patients. Similar trends were evident for other treatment comparisons.

DISCUSSION

The present meta-analysis corroborates the growing consensus that the potential for a high quality of life differs across RRTs. Synthesizing data from 49 published comparative studies, including 77 comparisons involving emotional distress and 66 involving psychological well-being, renal transplant recipients reported significantly less emotional distress and more psychological well-being than patients on either hospital-based CHD or CAPD. CAPD patients reported more psychological well-being than those on hospital-based CHD, and those on CHD reported more emotional distress than those receiving HHD.

Although meta-analysis strives to be comprehensive in integrating available studies, reliance on published comparisons can render results vulnerable to publication bias. This refers to the general problem that studies showing statisti-

Table 4. Summary of Differences Between Treatment Groups on Case-Mix Variables, Psychological Well-Being Treatment Comparisons

Comparison*	Age†	Physical Indicators	Employment	Education	Sex (men)
RT vCHD (16)	-0.73‡ (11)	-0.55‡ (12)	0.36‡ (8)	0.31‡ (4)	0.06 (13)
RT vCAPD (11)	-0.89‡ (6)	-0.48‡ (9)	0.59‡ (5)	0.13 (4)	0.17‡ (9)
RT vHHD (7)	-0.68‡ (5)	-0.40‡ (6)	0.14 (2)	0.22 (3)	-0.06 (6)
CAPD vCHD (18)	-0.18‡ (10)	-0.32 (13)	0.17‡ (9)	0.27‡ (7)	-0.03 (13)
CAPD vHHD (7)	0.21 (5)	0.23‡ (6)	-0.58‡ (2)	0.04 (3)	-0.07 (6)
CHD vHHD (7)	0.21 (5)	0.33‡ (7)	-0.39 (2)	-0.25 (3)	-0.25‡ (6)

*Treatment comparison (number of studies examining emotional distress or psychological well-being).

†Mean effect sizes (number of studies included in mean).

‡Mean effect size is significantly different from 0; $P < 0.05$.

cally significant effects are more likely to be accepted for publication than are those with nonsignificant results, the file-drawer problem.¹⁴ This problem was highlighted with regard to the literature concerning quality of life across alternative RRTs. Fail-safe numbers exceeded tolerance levels for four of six statistically significant treatment comparisons. ESs estimated on the basis of published studies may thus be inflated because nonsignificant comparisons may have been underrepresented. As a result, one cannot definitively rule out the competing hypothesis that the observed superiority of some RRTs may be attributable not to valid differences in the quality of life afforded by a particular treatment modality, but rather to a limitation of the literature. The number of published quality-of-life comparisons is insufficient to confirm the prevailing consensus (eg, that successful RT affords a superior quality of life relative to maintenance hemodialysis or CAPD).

The inability to control for case-mix differences across groups presents additional challenges to the synthesis and interpretation of this literature. As suspected, many key variables differed significantly across RRT groups. Renal transplant recipients, for example, were significantly younger, healthier, more highly educated, and more likely to be employed than their counterparts on maintenance dialysis. These differences were similar in magnitude, moreover, to the RRT-associated ESs observed for emotional distress and psychological well-being. Socioeconomic and other demographic characteristics are importantly related to quality of life in ESRD and across physically healthy and chronic disease populations more generally. Quality-of-life outcomes in ESRD and other chronic disease populations are reliably associated with sex,^{35,36} comorbid illnesses,^{5,6,9} and age.^{37,38}

Two studies, highly regarded for their methodologic rigor,^{2,3} may help illustrate the points we have asserted in interpreting the present meta-analytic results. Both used large sample sizes, psychometrically sound instruments, and directly examined the issue of case mix. Bremer et al² compared psychological well-being, represented by measures of positive affect and life satisfaction, across RRT modalities, reporting group differences similar to those we have observed based on the larger literature. Differences

across modalities were muted, however, when statistical analyses controlled for case mix. Although a number of potentially relevant variables were included (age, sex, race, education, primary diagnosis, duration of treatment, and comorbid conditions), employment status and income were not, despite their importance to psychological well-being.³⁹ Relevant to the present emphasis on the explanatory power of such factors, moreover, is the finding that case-mix variables were significantly related to psychological well-being. Comorbid conditions did not differ significantly across RRT groups in this study, however, and although the study sample was large ($n = 489$), it represented only 54% of the 903 individuals who received questionnaires. Finally, the study focused exclusively on psychological well-being; thus, the results cannot shed light on cross-modality differences in emotional distress.

Evans et al³ also measured subjective well-being and compared this across RRTs, investigating the extent to which differences might be attributable to case-mix confounders. Like Bremer et al,² the investigators observed statistically significant differences in psychological well-being across dialysis and posttransplant patients, and these persisted after controlling for case-mix differences. Contrary to the findings of Bremer et al,² however, comorbid conditions and previous transplant failures did not relate systematically to psychological well-being in this data set. In presenting their results, Evans et al³ cautioned against extrapolating their findings to the larger ESRD population because of the nonrepresentativeness of their sample. This consideration, together with the inconsistencies between the results of these two widely respected investigations, indicate that it would be premature to interpret current findings as definitive or conclusive.

It will be crucial in future research to consider and control for case-mix differences across RRTs. One powerful method to minimize the validity threats introduced by case mix would be to use a prospective repeated-measures experimental design in which the same cohort of patients can be assessed repeatedly and at clinically significant milestones, such as when a patient switches to a different RRT modality. A more practical and less costly method might involve careful identification and documentation of relevant case-mix

variables followed by statistical controls, as illustrated in many existing published studies.

It will also be crucial, of course, to sample quality-of-life outcomes representatively, including measures of both positive and negative states. To date, few studies have included both types of instruments. Equally important is the need to ensure that instruments be psychometrically sound and theoretically relevant. In addition to examining comorbid conditions as case-mix confounders, it would be important to measure severity of illness more generally and to investigate whether this contributes incrementally to quality of life. This might incorporate effects of ESRD across organ systems. The ESRD Severity Index,⁴⁰ for example, assesses the effects of comorbid conditions (eg, systemic lupus erythematosus), nonrenal effects of conditions responsible for renal failure (eg, hypertension), and complications of renal failure (eg, congestive heart failure) and/or treatment side effects (eg, dialysis access events). Routine inclusion of such a measure might help provide a more comprehensive assessment as a context in which to compare quality of life across alternative RRTs. It might also be useful to sample cases strategically, limiting research participants to those patients for whom any form of RRT would be equally suitable.

Progress in meta-analytic synthesis of original findings can be facilitated by establishing a standard for reporting relevant case-mix information across research reports. A common core of indicators described in sufficient detail (eg, means and SDs for each treatment group) would facilitate the investigation of case-mix variables as important explanatory factors in relation to quality of life because this may differ across RRTs. Toward this end, it might be useful for journals to standardize policy concerning the types of descriptive data required for publications based on clinical samples.

This meta-analysis found significant differences among RRTs with respect to two fundamental quality-of-life dimensions: psychological well-being and emotional distress. Although the results corroborated existing clinical impressions, their validity was generally threatened by the potential existence of unpublished studies showing nonsignificant differences across treatment groups. Validity was also threatened by the observation that

important case-mix variables differed across the treatment groups. Because these variables are also importantly related to psychosocial outcomes in ESRD, the existing literature cannot rule out the competing hypothesis that observed quality-of-life differences across RRTs are attributable to preexisting nonrenal and/or nontreatment differences. Research should address the problem of case mix directly and consistently, reporting relevant data for each treatment group in sufficient detail to facilitate future meta-analyses. Although the best available evidence to date indicates that quality of life differs systematically across patients receiving alternative RRTs, it is not clear whether this occurs because of valid differences across treatment modalities, preexisting differences among patients, or a combination of these two alternatives. We conclude, therefore, that further research is required.

ACKNOWLEDGMENT

This research was undertaken in partial fulfillment of the Master of Science degree by Jill I. Cameron under the direction of Dr G.M. Devins (supervisor), Dr C. Whiteside, and Dr J. Katz (supervisory committee) in the Institute of Medical Science, Faculty of Medicine, University of Toronto. The authors thank Dr John Hunsley for valuable information and advice concerning meta-analysis, Kirsten Woodend for assistance in coding published research reports, and the Culture, Community, and Health Studies Manuscript Review Seminar for valuable feedback about the manuscript.

REFERENCES

1. Blagg CR: Objective quantification of rehabilitation in dialysis and transplantation, in Friedman EA (ed): *Strategy in Renal Failure*. New York, NY, Wiley, 1978, pp 415-433
2. Bremer BA, McCauley CR, Wrona RM, Johnson JP: Quality of life in end-stage renal disease: A reexamination. *Am J Kidney Dis* 12:200-209, 1989
3. Evans RW, Manninen DL, Garrison LP, Hart LG, Blagg CR, Gutman RA, Hull AR, Lowrie EG: The quality of life of patients with end-stage renal disease. *N Engl J Med* 312:553-559, 1985
4. Simmons RG, Anderson C, Kamstra L: Comparison of quality of life of patients on continuous ambulatory peritoneal dialysis, hemodialysis, and after transplantation. *Am J Kidney Dis* 4:253-255, 1984
5. Devins GM, Binik YM, Hollomby DJ, Barre PE, Guttman RG: Helplessness and depression in end-stage renal disease. *J Abnorm Psychol* 90:531-545, 1981
6. Devins GM, Binik YM, Hutchinson TA, Hollomby DJ, Barre PE, Guttman RD: The emotional impact of end-stage renal disease: Importance of patients' perceptions of intrusiveness and control. *Int J Psychiatry Med* 13:327-343, 1983
7. Johnson JP, McCauley CR, Copley JB: The quality of

life of hemodialysis and transplant patients. *Kidney Int* 22:286-291, 1982

8. Binik YM, Devins GM, Orme CM: Psychological stress and coping in end-stage renal disease, in Neufeld RWJ (ed): *Advances in the Investigation of Psychological Stress*. New York, NY, Wiley, 1989, pp 305-342

9. Greenfield S, Sullivan L, Ailliman RA, Dukes K, Kaplan SH: Principles and practice of case mix adjustment: Applications to end-stage renal disease. *Am J Kidney Dis* 24:289-307, 1994

10. Binik YM: Coping with chronic life-threatening illness: Psychosocial perspectives on end-stage renal disease. *Can J Behav Sci* 15:373-391, 1983

11. De-Nour AK: The renal replacement therapies: Is there a difference in quality of life? *New Trends Exp Clin Psychiatry* 10:109-113, 1994

12. Smith ML, Glass GV, Miller TI: *The Benefits of Psychotherapy*. Baltimore, MD, Johns Hopkins University Press, 1980

13. Rosenthal R: *Meta-Analytic Procedures for Social Research*. Newbury Park, CA, Sage, 1991

14. Rosenthal R: The "file-drawer problem" and tolerance for null results. *Psychol Bull* 86:638-641, 1979

15. Begg CB: Publication bias, in Cooper H, Hedges LV (eds): *The Handbook of Research Synthesis*. New York, NY, Russell Sage Foundation, 1994, pp 399-409

16. Hutchinson TA, Boyd NF, Feinstein AR, Gonda A, Hollomby D, Rowat B: Scientific problems in clinical scales, as demonstrated in the Karnofsky index of performance status. *J Chronic Dis* 32:661-666, 1979

17. Beiser M: Components and correlates of mental well-being. *J Health Soc Behav* 15:320-327, 1974

18. Bradburn NM: *The Structure of Psychological Well-Being* (ed 1). Chicago, IL, Aldine, 1969

19. Devins GM, Beiser M, Dion R, Pelletier LG, Edward RG: Cross-cultural measurement of psychological well-being: The psychometric equivalence of Cantonese, Vietnamese, and Laotian translations of the Affect Balance Scale. *Am J Public Health* 87:794-799, 1997

20. Diener E: Subjective well-being. *Psychol Bull* 95:542-575, 1984

21. Devins GM, Mandin H, Hons RB, Burgess ED, Klassen J, Taub K, Schorr S, Letourneau PK, Buckle S: Illness intrusiveness and quality of life in end-stage renal disease: Comparison and stability across treatment modalities. *Health Psychol* 9:117-142, 1990

22. Radloff LS: The CES-D scale: A self-report depression scale for research in the general population. *Appl Psychol Meas* 3:385-401, 1977

23. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J: An inventory for measuring depression. *Arch Gen Psychiatry* 4:561-571, 1961

24. McNair DM, Lorr M, Droppelman LF: *EITS Manual*

for the Profile of Mood States. San Diego, CA, Educational and Industrial Testing Service, 1971

25. Campbell A: Subjective measures of well-being. *Am Psychol* 31:117-124, 1976

26. Cantril H: *The Pattern of Human Concerns*. New Brunswick, NJ, Rutgers University Press, 1965

27. Hedges LV, Becker BJ: Statistical methods in the meta-analysis of research on gender differences, in Hyde JS, Linn MC (eds): *The Psychology of Gender: Analysis Through Meta-Analysis*. Baltimore, MD, Johns Hopkins University Press, 1986, pp 14-50

28. Rosenthal R: Parametric measures of effect size, in Cooper H, Hedges LV (eds): *The Handbook of Research Synthesis*. New York, NY, Sage, 1994, pp 231-244

29. Hasselblad V, Hedges LV: Meta-analysis of screening and diagnostic tests. *Psychol Bull* 117:167-178, 1995

30. Ray JW, Shadish WR: How interchangeable are different estimators of effect size? [published erratum in *J Consult Clin Psychol* 66:532, 1998]. *J Consult Clin Psychol* 64:1316-1325, 1996

31. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ: Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol Methods* 1:170-177, 1996

32. Rosenthal R, Rubin DB: Meta-analytic procedures for combining studies with multiple effect sizes. *Psychol Bull* 99:400-406, 1986

33. Greenhouse JB, Iyengar S: Sensitivity analysis and diagnostics, in Cooper H, Hedges LV (eds): *The Handbook of Research Synthesis*. New York, NY, Sage, 1994, pp 383-398

34. Hedges LV: Fixed effects models, in Cooper H, Hedges LV (eds): *The Handbook of Research Synthesis*. New York, NY, Sage, 1994, pp 285-299

35. Devins GM, Hunsley J, Mandin H, Taub K, Paul LC: The marital context of end-stage renal disease: Illness intrusiveness and perceived changes in family environment. *Ann Behav Med* 19:325-332, 1997

36. Holahan CJ, Moos RH, Holahan CK, Brennan PL: Social support, coping, and depressive symptoms in a late-middle-aged sample of patients reporting cardiac illness. *Health Psychol* 14:152-163, 1995

37. Devins GM, Beanlands H, Mandin H, Paul LC: Psychosocial impact of illness intrusiveness moderated by self-concept and age in end-stage renal disease. *Health Psychol* 16:529-538, 1997

38. Kutner NG: Psychosocial issues in end-stage renal disease: Aging. *Adv Ren Replace Ther* 1:210-218, 1994

39. Pugliesi K: Work and well-being: Gender differences in the psychological consequences of employment. *J Health Soc Behav* 36:57-71, 1995

40. Craven J, Littlefield C, Rodin G, Murray M: The End-stage Renal Disease Severity Index (ESRD-SI). *Psychol Med* 21:237-243, 1991