Published in final edited form as:

Res Gerontol Nurs. 2011 January ; 4(1): 36–46. doi:10.3928/19404921-20101201-02.

Combining Extant Datasets with Differing Outcome Measures Across Studies of Older Adults after Cancer Surgery

Janet H. Van Cleave, PhD, RN[Clinical Assistant Professor],

New York University College of Nursing, 726 Broadway, 10th Floor, New York, NY 10003

Brian L. Egleston, PhD[Assistant Research Professor],

Biostatistics Facility, Fox Chase Cancer Center, Philadelphia, PA 19111

Meg Bourbonniere, PhD, RN[Vice President for Nursing Research], and

Thomas Jefferson University Hospital, Philadelphia, PA 19107

Ruth McCorkle, PhD, FAAN[Florence Schorske Wald Professor of Nursing]

Yale School of Nursing, New Haven, CT 06536

Abstract

Combining extant datasets with differing outcome measures, an economical method to generate evidence guiding older adults' cancer care, may introduce heterogeneity leading to invalid study results. We recently conducted a study combining extant datasets from five oncology nurse-directed clinical trials (parent studies) using norm-based scoring to standardize the differing outcome measures. The purpose of this article is to describe and analyze our methods in the recently completed study. Despite addressing and controlling for heterogeneity, our analysis found statistically significant heterogeneity ($p < 0.0001$) in temporal trends among the five parent studies. We concluded that assessing heterogeneity in combined extant datasets with differing outcome measures is important to ensure similar magnitude and direction of findings across parent studies. Future research should include investigating reasons for heterogeneity to generate hypotheses about subgroup differences or differing measurement domains that may have an impact on outcomes.

Keywords

Combining data; Older Adults; Cancer

Adults ages 65 and over comprise over 60% of the cancer population, yet only 25% to 38% of the patients enrolled in large national cancer clinical trials (Hutchins, Unger, Crowley, Coltman, & Albain, 1999; Lewis et al., 2003; Talarico, Chen, & Pazdur, 2004). Age-associated decline in functional reserve, increase in comorbid conditions, multiple medications, and delayed diagnosis limit accrual. The under-representation of older adults in major cancer clinical trials limits data analysis and generalizability of cancer treatments in this population (Talarico et al., 2004). Implementation of strategies, such as changes in Medicare policy to cover costs associated with clinical trials, have had minimal effect on incentivizing older adults' participation (Kimmick et al., 2005; Unger et al., 2006).

Although randomized clinical trials are the gold standard, researchers may need to use alternative study designs to increase the sample size of older adults with cancer in order to answer important research questions, detect clinically significant changes, and increase generalizability of results (Kiecolt & Nathan, 1985). One such alternative is combining extant datasets to increase the study sample. Combining extant data sets, however, has the potential to introduce heterogeneity into the study and limit the researchers' choice of

variables. Additionally, researchers may need to combine measures of concepts that are not strictly equivalent. All issues create the potential for imprecise measurement of concepts and study bias.

We recently completed such a study exploring the functional status of older adults after cancer surgery by combining five subsets of nurse-directed studies (Van Cleave, Egleston, & McCorkle, 2011). Although similar in design, data collection, and population, the five studies contained differing functional status measures: the Enforced Social Dependency Scale (ESDS) in four studies and SF-36® Health Survey (SF-36) in the fifth study. Based on the literature, we standardized the differing outcome measures for the analysis using norm-based scoring. The results of the study suggest that factors other than age influence recovery of function.

Therefore, the purpose of this article is to describe and examine our methods for combining extant datasets for the index study. Specifically, we will present a brief literature review and then describe the methods we used to standardize the differing outcomes. We will examine the combined dataset for heterogeneity in patterns of functional status across studies, and conclude the article by analyzing the benefits and limitations of our methods. For clarity, we will refer to the five nurse-directed clinical studies as the *parent* studies and refer to the recently conducted study of functional status in older adults as the *index* study.

Literature Review

Combining extant datasets with differing outcomes using standardized measures, a well-described method in the literature and basis for meta-analysis studies, has benefits and limitations (Higgins & Green, 2009; Kiecolt & Nathan, 1985; Orsi et al., 1999; Smith & Glass, 1977). Critics maintain that combining datasets with differing outcomes may lead to study bias due to the mixing of heterogeneous data. As a result, the researcher may falsely reject the null hypothesis when it represents truth (Type I error) or fail to reject the null hypothesis when the alternative hypothesis is true (Type II error) (Eysenck, 1994; Higgins & Green, 2009; Shadish, Cook, & Campbell, 2002). Proponents counter that mixing heterogeneous data is appropriate within the context of a larger contextual phenomenon. In other words, a study of “apples and oranges” can yield meaningless results when interested only in apples or oranges, but can significantly contribute to a wider question about fruit (Higgins & Green, 2009; Smith & Glass, 1977).

Researchers standardize differing outcome measures in combined extant datasets to facilitate analysis and test research questions. Two methods for standardizing differing outcome measures are mean differences and norm-based scoring. Mean differences consist of using standard deviations (SDs) together with sample sizes to compute the weight given to each study. For example, studies with small SDs are given relatively greater weight while studies with larger SDs are given relatively smaller weights. This is appropriate when differences in SD reflect real differences in the variability of outcomes in the study population (Higgins & Green, 2009). Another method, norm-based scoring, uses a linear transformation with mean=50 and SD=10. This creates the possibility of meaningfully comparing scores across studies (Ware, 2010).

Combination of Datasets

For the index study, we combined the five parent studies, conducted by the same principal investigator, to increase the population of postsurgical cancer patients for study power and generalizability. We first evaluated the five parent studies and judged them to be compatible based on data source, study design, procedures, and study participants. All studies collected

data at baseline, three and six months, and targeted postsurgical cancer population. The outcome measure, functional status, was conceptualized broadly across all studies as representing the total domain of function (capacity, reserve, performance, and capacity utilization) (Leidy, 1994) and defined as the individuals' actual performance of activities and tasks associated with their current life roles (Richmond, Tang, Tulman, Fawcett, & McCorkle, 2004).

The index study was approved by Yale School of Nursing (YSN) Human Subjects Institutional Review Board. Informed consent had been previously obtained from all patients during the parent studies, and study identification numbers were used in place of names or personal identifying data in to protect the rights of the participants.

Data Source

The five parent studies were conducted between 1983 and 2007 at academic cancer centers in the northwestern and northeastern United States (Table 1). Complete details of these studies have been provided elsewhere (McCorkle et al., 1989, 1994, 2000, 2009; McCorkle, Siefert, Dowd, Robinson, & Pickett, 2007). Patients were recruited during hospitalization, and baseline interviews conducted within 1 month after discharge except for study one. In study one, patients were recruited after discharge in the outpatient clinics and baseline interviews were conducted on average 60 days after surgery. The data collection times consisted of baseline (enrollment), and then 3 and 6 months after enrollment.

Demographic data and comorbidities were assessed by patient self-report and recorded at baseline. Cancer site, stage, and treatment were collected from treatment records and medical record audits. Functional status, symptoms, and mental health data were reported by patients and collected by research assistants at baseline, and 3 and 6 months. The types of cancers differed among parent studies. Study one consisted of patients with thoracic cancers. Studies two and three enrolled patients with heterogeneous cancers: breast, colorectal, head and neck, thoracic, prostate, gynecologic, bladder, pancreatic, esophageal, renal, and gastric. Study four targeted men with prostate cancer. The patient population for study five consisted of women who had undergone abdominal surgery for presumed gynecological cancers. Patients were recruited to this study prior to the final pathology, so this subset of records included women with a heterogeneous group of cancers including ovarian, uterine, endometrial, and metastatic breast and pancreatic cancers.

Participant Selection and Data Subsets

The inclusion criteria for the index study consisted of age 65 and older, enrollment in five oncology nurse-directed randomized clinical studies, with no other concurrent malignancy or other medical treatment. A total of 537 patients from the five parent studies met the inclusion criteria. Participant attrition occurred because of substantial patient mortality ($n=76$) during the six months of each study. We therefore limited the index study to patients who had data collected at a minimum of two or more time points. For analysis, we also limited the index study to patients with digestive system, thoracic, gynecologic sites, and genitourinary cancer sites treated primarily with surgery for a homogenous population. The final index study population composed of 316 patients.

Main Outcome Measure

Prior to combining datasets, we mapped the selected variables across the five parent studies, using a spread sheet, to examine and plan item compatibility. The mapping process described showed that studies 1, 2, 3, and 5 used a similar functional status measure, the ESDS (Tang & McCorkle, 2002). Researchers in Study 4 used SF-36, specifically the physical

component summary measures, to measure functional status (McHorney, Ware, Lu, & Sherbourne, 1994; McHorney, Ware, & Raczek, 1993; Ware & Sherbourne, 1992).

The ESDS consists of two components; personal and social competence. Personal competence includes six activities: eating/feeding, dressing, walking, traveling, bathing, and toileting. Dependency in each activity was reported by the patient and rated by the interviewer on a 6-point scale. Scores for personal competence were summed and ranged from 6 to 36. Social competence consisted of home, work, and recreational activities, which were rated on 4-point scales, and a communication category, rated on a 3-point scale. Scores for social competence were summed and ranged from 4 to 15. Scores for personal and social competence were summed to generate a total dependency score ranging from 10 to 51, with higher scores reflecting greater dependency. The ESDS has demonstrated reliability (Cronbach's alpha coefficient = 0.72 to 0.96) and both concurrent and predictive validity (Tang & McCorkle, 2002).

The SF-36 is a 36-item survey of health status used to assess eight health concepts: physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health. The SF-36 has demonstrated item-internal consistency and item-discriminant validity. Reliability coefficients have ranged from a low of 0.65 to a high of 0.94 across scales (median=0.85). The measure has also demonstrated validity in discriminating between serious and minor medical conditions as well as psychiatric illnesses (McHorney et al 1993, 1994; Ware & Sherbourne, 1992).

Symptoms—Number of symptoms was measured by the Symptom Distress Scale (SDS) (McCorkle, Cooley, & Shea, 1998) to quantify the effect of each additional symptom on functional status. The SDS consists of 13 common symptoms of cancer patients: frequency of nausea, severity of nausea, appetite, insomnia, frequency of pain, severity of pain, fatigue, bowel pattern, concentration, appearance, breathing, outlook, and cough. Participants rated their distress on a scale from 1 (*low distress*) to 5 (*high distress*). Absence of a symptom was defined as values 1 and 2, and presence of a symptom was equal to values 3, 4, and 5. This categorization is consistent with previous studies using these same values to distinguish between low and high symptom distress (Cooley, Short, & Moriarty, 2003). Number of symptoms, therefore, represented the sum of present symptoms with highest score equal to 13 and lowest score equal to 0. The SDS has demonstrated content, construct, and criterion validity. The SDS has also demonstrated reliability, with reported Cronbach's alpha coefficient ranging from 0.70 to 0.89 (McCorkle, et al 1994, 1998; Sarna, Lindsey, Dean, Brecht, & McCorkle, 1994).

Data Preparation

We prepared the data for combination by reverse coding the ESDS scores to attain consistent direction with the SF-36. The data were then converted to standardized norm-based scores (mean=50, SD=10) using 0–100 metric and Z scores (Figure 1) (Ware, Kosinski, Turner-Bowker, & Gandek, 2005). SF-36 scores were converted to norm-based scores using QualityMetric Health Outcomes Scoring Software, version 2.0

Methods

T-score Transformations

We used norm-based scoring to transform the outcome variables and combine the datasets. By this, we mean that we standardized the outcome measures to a linear T-score having a mean of 50 and SD of 10 within each study (Figure 1). The T-score is based on Student's T-

distribution, the same distribution used for T-test comparisons of sample means. The T-distribution is essentially equivalent to the normal distribution (Z-scores) as the sample gets larger. Standardization of the outcome measures via T-score transformation eases interpretation across different measures. A 10-unit change in any standardized score is equivalent to a one SD change regardless of outcome measure. In many fields, the use of SD units to define clinically relevant changes in metrics is well accepted (Cohen, 1992). Ware et al. (1998) similarly use linear T-score transformations to make comparisons using SF-36 measures.

To combine the datasets in our study, we normalized our data using the study-specific means and SDs. By using study-specific means, we crudely removed study-specific confounding of demographic and clinical characteristics of interest with the outcomes. In using the study-specific SDs, the interpretation of effects is conditional on the SD within each study.

There are limitations to our approach. Using study-specific means to standardize variables limits the ability to detect associations of interest if the study-level effects are highly correlated with other characteristics, such as age. For example, if age is similar within studies but varies across studies, then standardization by study specific means would hinder the ability to find age related associations. Still, this is not a major limitation, as it would nonetheless be difficult to distinguish age-related effects from study-specific effects if participants' ages are highly correlated with the study. In such a case, reducing the ability to detect effects would protect one from falsely concluding that age is related to an outcome of interest when indeed there were other study-specific factors driving the association.

Another limitation to the approach is that the interpretation of effects in SD units makes clinical relevance of effects more difficult to assess. Further, if the SD of the outcomes is highly variable among studies, standardizing by study-specific means can create artificial heterogeneity in effects among studies. Cummings (2004) criticized meta-analyses using standardized effects on both these measures. Despite such criticism, such standardization does not necessarily lead to substantially different overall decision rules regarding associations of variables. In Cummings' (2004) example, pooled effects were statistically significant both before and after standardization.

Despite the limitations to this approach, the main benefit of such standardization is its ease of implementability among researchers. While there has been work on rigorous methods to combine datasets using disparate quality of life measures (Chang & Cella, 1997; Dobrez, Cella, Pickard, Lai, & Nickolov, 2007; Revicki et al., 2009), such methods generally require that the disparate measures are administered simultaneously in a sample so they can be mapped to each other. When a researcher does not have access to information regarding the appropriate mapping of two variables in a population of interest, simpler methods, such as T-score-based transformations, might be the only option available.

Preliminary Validation of Approach

After performing the T-score transformations, we sought evidence concerning the validity of the approach with respect to the ESDS and SF-36 measures of interest. Although we suspected the ESDS and SF-36 measured different outcomes, we expected that the trajectory of the measures over time on average would be the same after standardization. This does not imply that the same people would have the same trajectory assessed by the two measures. People who had significant changes in one measure might not have similar changes in the other. On average, however, we would expect the population changes to be similar.

To assess whether the standardized measures did move similarly in each sample, we fit multiple linear regression models of the standardized outcomes separately for each study.

We included the 3- and 6- month data collection time points as dummy variables in the model. We also included age and number of symptoms as covariates to crudely adjust for confounding by these variables among the studies. We fit the models by generalized estimating equations assuming an unstructured working correlation matrix to account for the correlation of observations within individuals over time. We then assessed whether the direction and magnitude of the regression coefficients were similar across the studies. If the direction and magnitude were similar, we would have preliminary evidence supporting the combination of datasets with ESDS and SF-36 measures via T-score standardization.

To formally test for heterogeneity in temporal trends in functional status among the studies, we examined a multiple linear regression of function in which we included the wave dummy variables (2 covariates, with the baseline time used as the reference time point), study dummy variables (4 covariates, with the first study used as the reference study), and the 8 interaction terms between the wave and study dummy variables (wave dummy variables multiplied by study dummy variables). We also included the total number of symptoms and age in the model. We again fit the model by generalized estimating equations to account for correlation within individuals. Statistically significant interaction terms would suggest that changes in functional status over time varied among the studies.

In the index study, we used an inverse probability of censoring weighted estimator to account for the missing outcome data over the three data collection times. To implement this method, functional status measures were weighted inversely by the probability of staying in the study (Fitzmaurice & Laird, 2000; Scharfstein, Rotnitzky, & Robins, 1999). We did not implement the weight estimator for this analysis to fully compare the functional status measures across all three data collection times. Analyses were performed using STATA, version 10.1 (StataCorp LP, College Station, Texas).

Results

Study Population

The study sample consisted of 316 patients who underwent surgery for digestive, thoracic, genitourinary, and gynecological cancer (Table 2). The mean age for the total population was 71.8 (SD=5.4 years). The study population consisted primarily of patients ages 65 to 69 (44.9%), White race (76.6%), married (61.7%), who did not work (82.3%), yet had some high school education (90.8%). The study sample was balanced between men and women (49.7% versus 48.7%) and 25.3% of the population earned \$40,000 or greater.

Comparison Across Studies

We conducted a regression analysis of the effect of time of data collection and number of symptoms (Table 3). The total study population showed statistically significant improvement of functional status at 3 months ($\beta=6.85$, 95% confidence interval [CI] = 5.76 to 7.93) and 6 months ($\beta=8.47$, 95% CI = 7.21 to 9.73) compared with baseline functional status. Both number of symptoms and age were significantly associated with decreased functional status: functional status declined with each additional symptom ($\beta=-1.58$, 95% CI = -1.79 to -1.38) and each added year of age ($\beta=-0.18$, 95% CI = -0.29 to -0.08). We next analyzed the pattern of the recovery of functional status across studies (Table 3). Time of data collection was significantly associated with functional status in study 1 (3 months: $\beta=-2.19$, 95% CI = -3.76 to -0.62; 6 months: $\beta=-4.71$, 95% CI = -8.18 to -1.24), study 2 (6 months: $\beta=8.02$, 95% CI = 4.26 to 11.78), study 3 (3 months: $\beta=11.31$, 95% CI = 10.09 to 12.54; 6 months: $\beta=13.18$, 95% CI = 11.84 to 14.51), and study 5 (3 months: $\beta=9.32$, 95% CI = 6.51 to 12.13; 6 months: $\beta=10.91$, 95% CI = 7.98 to 13.84).

A two-way line plot portrays the mean standardized functional status scores across studies to compare and contrast the trajectory of recovery of functional status over time of data collection (wave) (Figure 2). The population in Studies 2, 3, and 5 demonstrated increased mean functional status over time while those in Studies 1 and 4 experienced either a downward or non-specific trajectory.

Table 4 lists the results of our interaction model used to test for heterogeneity in functional status temporal trends among the studies. We found that a Wald test for the joint statistical significance of the 8 interaction terms was statistically significant ($p < 0.0001$), which suggests that there is indeed heterogeneity in temporal trends.

We then examined whether the ESDS and SF-36 contained similar items that could be compared for similarity in trajectory over time. The ESDS contained the items of dressing and bathing; the SF-36 had one item titled bathe or self-dress. We again used two-way line plots to portray the mean standardized functional status across times of data collection. This also demonstrated differing trajectories across time. All three items demonstrated improved scores between baseline and 3 months. The ESDS items, however, showed continued improved mean scores at 6 months compared with decreased mean score by the SF-36 (Figure 3).

Discussion

Combining extant datasets represents a potential economical solution to increase the power and generalizability of studies of older adults with cancer, thus providing evidence to guide clinicians. However, the method may introduce bias into study results due to heterogeneous data, imprecise measurement of concepts, or combining measures that are not strictly equivalent. Literature supports that broadly defining the outcome and then using standardized scores for analysis of differing outcome measures across studies may overcome these potential methods problems. Thus, we used norm-based scoring to standardize the differing functional status outcome measures because of its ease of use and interpretation. In the index study, we found a statistically significant improvement in functional status in the total population. An analysis of individual studies and comparable measurement items, however, demonstrated heterogeneity in trajectory of functional status across parent studies. Hence, the results provide evidence for the benefits and limitations of combining extant datasets with differing measures.

Benefits of Combining Extant Data Sets

Results from our analysis suggest that we gained some benefit from combining extant datasets in studies of older adults with cancer. The method may have led to increased study power to generate evidence that older adults have significant recovery of functional status after cancer surgery. Additionally, we were able to expand our index study population to include older adults with differing types of cancer surgery, and thus increased the generalizability of our results.

Limitations of Combining Extant Datasets

Results from our analysis also exposed some of the potential limitations of combining extant datasets using standardized outcome measures. We may have introduced heterogeneity into the index study, as exemplified by the differing trajectories of functional status after cancer surgery across parent studies and supported by the Wald test for the joint statistical significance of the 8 interaction terms. Functional status improved over time in studies 2,3 and 5. In contrast, functional status in studies 1 and 4 either decreased over time or remained approximately the same over time. One reason for the differences in functional status

temporal trends may stem from the 25 year span of the parent studies. For example, study 1 was conducted between 1983 and 1986 with thoracic cancer patients prior to the adoption of life-extending treatments including combinations of surgery, chemotherapy, and radiation therapy for older thoracic cancer patients (Bunn & Lilenbaum, 2003). Study 4, which demonstrated no significant difference of functional status across time, was conducted from 1997 to 2000 and included men ages 65 and over with early prostate cancer undergoing surgery for potential cure.

An additional reason for differences in functional status temporal trends may be related to differing research questions across the parent studies. For example, study 2 evaluated home care for cancer patients and study 4 explored nursing's impact on men post prostatectomy (Table 1). Variables chosen to represent study concepts ultimately provide the information that answers the study question (Hulley et al., 2001). The two measures used across studies, ESDS and SF-36, provide different information for researchers. Compared with the SF-36, the ESDS provides information on personal competence and potential targets for interventions by assessing patients' activities of daily living (i.e., eating/feeding, dressing, walking, travel, bathing, and toileting) (Tang & McCorkle, 2002). The SF-36, in contrast, addresses limitations in self-care: physical, social and role activities, bodily pain, and frequent tiredness (Ware, 2010).

We took several steps to address and control for heterogeneity based on meta-analysis methods. For the index study, we carefully developed inclusion criteria to generate a homogenous population, examined and judged the five parent studies compatible in study design and data collection methods, reverse coded the ESDS to attain consistent direction with the SF-36, and then transformed the outcome measures to norm-based scoring. Hence, the differing trajectories raise the issue that, despite standardization with norm-based scoring, the ESDS and SF-36 may represent different functional status domains. We were unable to test for their equivalency, though, since none of the parent studies measured functional status with both the SF-36 and ESDS.

Conclusion

Assessing heterogeneity in combined extant datasets with differing outcome measures is important to ensure that the magnitude and direction of findings are similar across studies. Finding a great deal of heterogeneity could call into question the validity and generalizability of the combined results. Some have argued that differences in recruitment populations among studies could be a reason for substantial heterogeneity (Elwood, 1998). Therefore, investigators should examine study-level characteristics, such as general age or gender differences among the studies, that could be creating heterogeneity. Such considerations can be helpful in generating hypotheses about subgroup differences that have an impact on outcomes.

In our study, we found that the interaction terms between study and time period were statistically significant in a regression model, which suggests that there is indeed heterogeneity in temporal trends in functional status among the studies. Future research could investigate reasons for such heterogeneity, which may be related to demographic, disease, or treatment characteristics among the studies and differing domains of functional status among measures.

References

- Bunn PA Jr, Lilenbaum R. Chemotherapy for elderly patients with advanced non-small-cell lung cancer. *Journal of the National Cancer Institute*. 2003; 95:341–343. [PubMed: 12618492]

- Chang CH, Cella D. Equating health-related quality of life instruments in applied oncology settings. *Physical Medicine and Rehabilitation: State of the Art Reviews*. 1997; 11:397–406.
- Cohen J. A power primer. *Psychological Bulletin*. 1992; 112:155–159. [PubMed: 19565683]
- Cooley ME, Short TH, Moriarty HJ. Symptom prevalence, distress, and change over time in adults receiving treatment for lung cancer. *Psychooncology*. 2003; 12:694–708. doi 1002/pon.694. [PubMed: 14502594]
- Cummings P. Meta-analysis based on standardized effects is unreliable. *Archives of Pediatrics and Adolescent Medicine*. 2004; 158:595–597. [PubMed: 15184227]
- Dobrez D, Cella D, Pickard AS, Lai JS, Nickolov A. Estimation of patient preference-based utility weights from the functional assessment of cancer therapy - general. *Value in Health*. 2007; 10:266–272. [PubMed: 17645681]
- Elwood, M. Critical appraisal of epidemiological studies and clinical trials. 2. Oxford, UK: Oxford University Press; 1998.
- Eysenck HJ. Meta-analysis and its problems. *BMJ*. 1994; 309:789–792. [PubMed: 7950571]
- Fitzmaurice GM, Laird NM. Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics*. 2000; 1:141–156. [PubMed: 12933516]
- Higgins, JPT.; Green, S., editors. *Cochrane Handbook for Systematic Reviews of Interventions: Version 5.0.2*. 2009. Retrieved from The Cochrane Collaboration website: <http://www.cochrane-handbook.org>
- Hulley, SB.; Cummings, SR.; Browner, WS.; Grady, D.; Hearst, N.; Newman, TB. *Designing clinical research: An epidemiologic approach*. 2. Philadelphia: Lippincott Williams & Wilkins; 2001.
- Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *New England Journal of Medicine*. 1999; 341:2061–2067. [PubMed: 10615079]
- Kiecolt, KJ.; Nathan, LE. *Secondary analysis of survey data*. Beverly Hills, CA: Sage; 1985.
- Kimmick GG, Peterson BL, Kornblith AB, Mandelblatt J, Johnson JL, Wheeler J, Muss HB. Improving accrual of older persons to cancer treatment trials: a randomized trial comparing an educational intervention with standard information: CALGB 360001. *Journal of Clinical Oncology*. 2005; 23:2201–2207. [PubMed: 15800312]
- Leidy NK. Functional status and the forward progress of merry-go-rounds: toward a coherent analytical framework. *Nursing Research*. 1994; 43:196–202. [PubMed: 8047422]
- Lewis JH, Kilgore ML, Goldman DP, Trimble EL, Kaplan R, Montello MJ, Escarce JJ. Participation of patients 65 years of age or older in cancer clinical trials. *Journal of Clinical Oncology*. 2003; 21:1383–1389. [PubMed: 12663731]
- McCorkle R, Benoliel JQ, Donaldson G, Georgiadou F, Moinpour C, Goodell B. A randomized clinical trial of home nursing care for lung cancer patients. *Cancer*. 1989; 64:1375–1382. [PubMed: 2670188]
- McCorkle, R.; Cooley, ME.; Shea, J. A user's manual for the Symptom Distress Scale. Yale School of Nursing; New Haven, CT: 1998. Unpublished manuscript
- McCorkle R, Dowd M, Ercolano E, Schulman-Green D, Williams AL, Siefert ML, Schwartz P. Effects of a nursing intervention on quality of life outcomes in post-surgical women with gynecological cancers. *Psychooncology*. 2009; 18:62–70. [PubMed: 18570223]
- McCorkle R, Jepson C, Malone D, Lusk E, Braitman L, Buhler-Wilkerson K, Daly J. The impact of posthospital home care on patients with cancer. *Research in Nursing & Health*. 1994; 17:243–251. [PubMed: 8036272]
- McCorkle R, Siefert ML, Dowd MF, Robinson JP, Pickett M. Effects of advanced practice nursing on patient and spouse depressive symptoms, sexual function, and marital interaction after radical prostatectomy. *Urologic Nursing*. 2007; 27:65–77. [PubMed: 17390930]
- McCorkle R, Strumpf NE, Nuamah IF, Adler DC, Cooley ME, Jepson C, Torosian M. A specialized home care intervention improves survival among older post-surgical cancer patients. *Journal of the American Geriatrics Society*. 2000; 48:1707–1713. [PubMed: 11129765]
- McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*. 1994; 32:40–66. [PubMed: 8277801]

- McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*. 1993; 31:247–263. [PubMed: 8450681]
- Orsi AJ, Grey M, Mahon MM, Moriarty HJ, Shepard MP, Carroll RM. Conceptual and technical considerations when combining large data sets. *Western Journal of Nursing Research*. 1999; 21:130–142. [PubMed: 11512172]
- Revicki DA, Kawata AK, Harnam N, Chen WH, Hays RD, Cella D. Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. *Quality of Life Research*. 2009; 18:783–791. [PubMed: 19472072]
- Richmond, T.; Tang, ST.; Tulman, L.; Fawcett, J.; McCorkle, R. Measuring function. In: Frank-Stromborg, M.; Olsen, SJ., editors. *Instruments for clinical health-care research*. 3. Sudbury MA: Jones Bartlett; 2004. p. 83-99.
- Sarna L, Lindsey AM, Dean H, Brecht ML, McCorkle R. Weight change and lung cancer: relationships with symptom distress, functional status, and smoking. *Research in Nursing & Health*. 1994; 17:371–379. [PubMed: 8090948]
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*. 1999; 94:1096–1120.
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin; 2002.
- Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *American Psychologist*. 1977; 32:752–760. [PubMed: 921048]
- Talarico L, Chen G, Pazdur R. Enrollment of elderly patients in clinical trials for cancer drug registration: a 7-year experience by the US Food and Drug Administration. *Journal of Clinical Oncology*. 2004; 22:4626–4631. [PubMed: 15542812]
- Tang, S.; McCorkle, R. A user's manual for the Enforced Social Dependency Scale. Yale School of Nursing; New Haven, CT: 2002. Unpublished manuscript
- Unger JM, Coltman CA Jr, Crowley JJ, Hutchins LF, Martino S, Livingston RB, Albain KS. Impact of the year 2000 Medicare policy change on older patient enrollment to cancer clinical trials. *Journal of Clinical Oncology*. 2006; 24:141–144. [PubMed: 16330670]
- Van Cleave JH, Egleston B, McCorkle R. Factors Affecting Recovery of Functional Status In Older Adults After Cancer Surgery. *Journal of American Geriatrics Society*. 2011; 59:34–43.
- Ware, JE, Jr. SF-36[®] Health Survey Update. 2010. Retrieved July 13, 2010, from: <http://www.sf-36.org/tools/sf36.shtml>
- Ware JE Jr, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, Thunelborg K. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. *Journal of Clinical Epidemiology*. 1998; 51:1167–1170. [PubMed: 9817134]
- Ware, JE., Jr; Kosinski, M.; Turner-Bowker, DM.; Gandek, B. *How to score version 2 of the SF-12 Health Survey (with a supplement documenting Version 1)*. Lincoln, RI and Boston: QualityMetric Incorporated and Health Assessment Lab; 2005.
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*. 1992; 30:473–483. [PubMed: 1593914]

The data were prepared for statistical analysis with the following steps:

1. Reverse coding of ESDS scores: = (maximum score + minimum score) – score
2. Transformation of raw scale scores to 0 – 100:

$$\text{Transformed score} = \frac{(\text{actual raw score} - \text{lowest possible raw score})}{\text{Possible raw score range}} \times 100$$

3. Calculating standard scores

$$\text{Standard Score (T- Score)} = \frac{X - \bar{X}}{\text{Estimated standard deviation}}$$

4. Transforming to norm-based score

$$\text{Norm-based score} = (\text{T-Score} \times 10) + 50$$

Figure 1.

PROCEDURE FOR STANDARDIZATION OF FUNCTIONAL STATUS SCORE

Note: If the standard deviation were known, then the T-score would instead be a Z-score.

ESDS = Enforced Social Dependency Scale.

Figure 2.
Standardized Functional Status by Wave and Study

Figure 3.
Comparison of Similar Measurement Items Among SF-36® Health Survey (SF-36) and Enforced Social Dependency Scale (ESDS).

Parent Studies

Table 1

Study Number	Title of Study	Time Period	Type of Cancer	No. of Participants (No. Participants Age $\geq$ 65 with Two Data Collection Time Points)	Design
1	Evaluation of Cancer Management (McCorkle et al., 1989)	1983 – 1986	Thoracic Cancer	165 (42)	Randomized, controlled trial; three groups; five data collection times
2	Evaluation of Home Care for Cancer Patients (McCorkle et al., 1994)	1987 – 1990	Multiple cancers	233 (39)	Quasi-Experimental; two groups; three data collection times
3	Nursing's Impact on Quality of Life in Elders (McCorkle et al., 2000)	1992– 1997	Multiple Solid Tumors	383 (169)	Randomized, controlled trial; two groups; three data collection times
4	Nursing's Impact on Men Post Prostatectomy (McCorkle, Siefert, Dowd, Robinson, & Prickett, 2007)	1997 – 2000	Prostate Cancer	107 (23)	Randomized controlled trial; two groups; three data collection times
5	Nursing's Impact on Quality of Life Outcomes for Women With Cancer (McCorkle et al., 2009)	2004 – 2007	Ovarian Cancer	145 (43)	Randomized controlled trial; two groups; four data collection times

Table 2

Demographic, Biologic, Psychological, and Treatment Variables of Study Sample (N=316)*

Variable	Mean (SD), Range	
	Mean (SD)	Range
Age (Years)	71.8 (5.4)	65 to 93
	n	%
Age Categories		
65–69	142	44.9
70–74	105	33.2
Ages 75 and over	69	21.8
Gender		
Male	157	49.7
Female	154	48.7
Missing data	5	1.6
Race/Ethnicity		
White	242	76.6
Black/Other Race	55	17.4
Missing data	19	6

Variables	Study Population	
	n	%
Marital Status		
Married/Living With Partner	195	61.7
Not Married	120	38.0
Missing data	1	0.3
Employment Status		
Not Working/Retired/Unemployed	260	82.3
Employed	56	17.7
Education		
Grades 1–8	29	9.2
Grades 9–12	145	45.9
Over 12	142	44.9
Annual Income		
\$0–19,999	97	30.7
\$20,000–39,999	65	20.6
\$40,000+	80	25.3
\$Unknown	74	23.4
Cancer Sites		
Genitourinary Sites	87	27.5
Thoracic	86	27.2
Gynecologic Sites	73	23.1
Digestive	70	22.2

Variables	Study Population	
	n	%
Marital Status		
Treatment		
Surgery	149	47.2
Surgery + Chemotherapy	101	32.0
Surgery + Chemotherapy + Radiation	39	12.3
Surgery + Radiation	27	8.5
APN Intervention		
Yes	171	54.1
No	145	45.9
Number of Symptoms		
0	35	11.1
1	42	13.3
2	39	12.3
3 or more	195	61.7
Missing data	5	1.2

Note. APN = advanced practice nurse.

Percentages may not total 100 due to rounding

Table 3

Regression Analysis By Study

	Beta Estimate (SE)						
Variable	Total Population (n=316)	Study 1 (n=42)	Study 2 (n=39)	Study 3 (n=169)	Study 4 (n=23)	Study 5 (n=43)	
Time of Data Collection							
Baseline (intercept) score	62.90 (3.98)***	71.18 (13.45)***	85.17 (15.44)***	54.57 (4.44)***	39.95 (37.74)	77.51 (9.04)***	
3 months	6.85 (0.55)***	-2.19 (0.80)**	2.01 (1.13)	11.31 (0.63)***	-2.59 (2.50)	9.32 (1.43)***	
6 months	8.47 (0.64)***	-4.71 (1.77)**	8.02 (1.92)***	13.18 (0.68)***	-1.66 (2.66)	10.91 (1.50)***	
Total No of Symptoms	-1.58 (0.10)***	-1.43 (0.21)***	-2.08 (0.27)***	-1.38 (0.12)***	-0.74 (0.48)	-1.39 (0.29)***	
Age	-0.18 (0.05)**	-0.20 (0.18)	-0.40 (0.21)	-0.13 (0.06)*	0.21 (0.54)	-0.42 (0.12)**	

Note. SE = standard error.

*

p <0.05;

**

p <0.01;

p <0.001.

Table 4Interaction model to test temporal trend differences among studies (n=316) ^a

Variable	Beta Estimate	Standard Error	95% Confidence Interval
Time of data collection			
Baseline (intercept) ***	71.26	4.10	63.23, 79.29
3 Months *	-2.36	1.19	-4.69, -0.03
6 Months **	-4.64	1.54	-7.66, -1.61
Total Number of Symptoms ***	-1.46	0.10	-1.65, -1.27
Age ***	-0.20	0.05	-0.30, -0.09
Study			
Study 1 (reference)			
Study 2 *	-3.38	1.52	-6.36, -0.39
Study 3 ***	-11.26	1.19	-13.58, -8.93
Study 4	-0.75	1.82	-4.33, 2.82
Study 5 ***	-9.82	1.49	-12.74, -6.89
Study × time of data collection			
Study 2 × 3 Months **	4.65	1.73	1.25, 8.05
Study 2 × 6 Months ***	11.70	2.17	7.44, 15.95
Study 3 × 3 Months ***	13.50	1.33	10.89, 16.12
Study 3 × 6 Months ***	17.68	1.71	14.32, 21.04
Study 4 × 3 Months	-2.40	2.12	-6.55, 1.74
Study 4 × 6 Months	0.46	2.74	-4.91, 5.84
Study 5 × 3 Months ***	11.57	1.67	8.31, 14.84
Study 5 × 6 Months ***	15.28	2.11	11.15, 19.41

^aJoint Wald test for interaction terms: p < 0.0001.

* p < 0.05;

** p < 0.01;

*** p < 0.001