

# Measuring the Efficacy of a Quality Improvement Program in Dialysis Adequacy with Changes in Center Effects

JEFFREY C. FINK,<sup>\*†</sup> MIN ZHAN,<sup>†</sup> STEVEN A. BLAHUT,<sup>\*</sup> MICHAEL SOUCIE,<sup>§</sup>  
WILLIAM M. MCCLELLAN<sup>‡§</sup>

<sup>\*</sup>Division of Nephrology, Department of Medicine and <sup>†</sup>Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, Maryland; <sup>‡</sup>Renal Division, Department of Medicine, Emory University, Atlanta, Georgia; <sup>§</sup>Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia.

**Abstract.** End-stage renal disease (ESRD) patients treated with hemodialysis have a high mortality rate, which is largely due to inadequate dialysis. Dialysis adequacy, measured by the urea reduction ratio (URR), tends to be correlated within dialysis facilities with wide variations in average center adequacy. These are characteristics of a center effect, which can have an important impact on dialysis adequacy. This study measured the center effect observed in an ESRD Network before and after a successful quality improvement project (QIP). URR values were recorded on patients sampled from 196 facilities in ESRD Network 6 before (pre-QIP,  $n = 5309$ ) and after (post-QIP,  $n = 5753$ ). These data were used to determine the *within center* correlation ( $\rho$ ) of individual URR values and *between center* variation in aggregate URR values in both samples. The

overall mean URR improved from the pre- to post-QIP sample (mean URR  $64.7 \pm 0.1$  versus  $69.8 \pm 0.1$ , respectively;  $P = 0.001$ ). There was a high degree of *within center* correlation in dialysis adequacy across the facilities, which significantly diminished post-QIP ( $\rho$ , 0.15 [95% CI, 0.12 to 0.18] versus  $\rho$ , 0.06 [95% CI, 0.04 to 0.08]). The *between center* variation in mean URR also declined from the pre-QIP to the post-QIP sample (SD, 3.6 versus 2.8). In conclusion, there is a center effect on dialysis adequacy measurable in an ESRD Network, which diminishes after a successful QIP; therefore, when implementing a QIP to improve dialysis adequacy, changes in the center effect should be considered a potential indicator of the efficacy of the intervention.

The health and well being of patients with end-stage renal disease (ESRD) continues to be dismal (1), and much of this poor outcome can be attributed to inadequate treatment with hemodialysis, which is the renal replacement therapy for a majority of patients with this disease (2–6). Although there are national standards for adequate dialysis (7), there are several barriers to achieving these mandated treatment goals. Efforts to improve dialysis adequacy have emphasized correcting patient-specific factors that are linked to inadequate dialysis (8–11). It is likely that factors at the center level also contribute to the shortcomings in dialysis treatment.

When present, such center effects on dialysis adequacy results in patients from the same dialysis unit receiving similar doses of dialysis compared with patients from different centers. A high degree of intrafacility or *within center* correlation in dialysis adequacy would suggest not only that patients are influenced by a center effect but also that the patients with poor care were not randomly distributed among dialysis facilities. A

related finding to an observed center effect in a population of dialysis patients is that aggregate achieved adequacy in centers is widely variant. The *between center* variation in performance results in some centers having overall adequacy exceeding mandated goals while others underperform, which is also consistent with the possibility that poorly dialyzed patients aggregate in certain treatment centers (12). In this study, we hypothesized that such a center effect in a population of hemodialysis patients exists and is undesirable. If this hypothesis were correct, the center effect and its associated between center variation in adequacy would diminish if a quality improvement initiative is applied to the dialysis population of interest. To test this hypothesis, we observed a dialysis population for which an effective quality improvement project (QIP) was implemented to assess if there was a significant center effect and whether it changed after implementation of the successful QIP.

## Materials and Methods

### Study Sample

This was an observational, retrospective secondary analysis of the dialysis population in ESRD Network 6 (North Carolina, South Carolina, Georgia), which had a survey of dialysis adequacy, as measured by the urea reduction ratio (URR) both before and after a QIP conducted across the Network. The URR measurements were made using pretreatment and posttreatment blood urea nitrogen (BUN) values obtained from study participants in the October before and after the QIP. The URR was calculated using the following equation:

Received August 13, 2001. Accepted June 11, 2002.

Correspondence to Dr. Jeffrey C. Fink, Rm N3W143, University of Maryland Medical System, 22 S. Greene St., Baltimore, MD 21201. Phone: 410-328-5720; Fax 410-328-5685; E-mail: jfink@medicine.umaryland.edu

1046-6673/1309-2338

Journal of the American Society of Nephrology

Copyright © 2002 by the American Society of Nephrology

DOI: 10.1097/01.ASN.0000027978.98194.1F

$$URR = ([BUN_{pre} - BUN_{post}]/BUN_{pre}) \times 100 \quad (1)$$

The details of the study and QIP are described elsewhere but summarized here (13). In October 1994 (pre-QIP), 30 patients from each facility in the Network were randomly selected for inclusion in the study with entry of measured URR from the final quarter of that year into the pre-QIP data sample. The QIP was conducted between November 1994 and October 1995. At the end of the QIP, 30 patients were again randomly selected from all Network facilities with submission of each individual’s measured URR from the final quarter of 1995 in the post-QIP data set. Facilities with a census of fewer than 30 patients had all patients included in the study before and after the QIP. While sampling done after the QIP was independent from that done before the QIP, patients selected for the pre-QIP sample were not excluded from the post-QIP group if sampled again. Along with URR values, several patient-specific covariates were recorded and a facility practice and quality improvement survey was sent to all dialysis facilities before the QIP and self-administered by facility staff.

### Quality Improvement Project

The QIP was conceived, developed, and supervised by the ESRD Network 6 Medical Review Board and administrative staff (MRB). The program involved education on proper measurement of dialysis adequacy parameters, workshops designed to assist facilities to conduct quality improvement activities, setting of improvement goals with monitoring by the MRB, and site visits from the MRB staff if improvement goals were not achieved within the QIP intervention period. The QIP was mandatory for the 10% of facilities with the highest proportion of patients having a URR value less than 50% in the pre-QIP survey. All other facilities were invited to participate in the workshops and QIP activities; however, this was not mandatory for them.

### Statistical Model

The study analysis utilized a mixed effect linear model designed to accommodate clustered data sets, which often arise in observational

studies where there is a natural hierarchical structure. Dialysis populations represent one type of clustered data where a set of ESRD patients is dialyzed in a smaller group of dialysis units. Data from subjects within a smaller set of centers can be correlated, and ordinary linear regression models are often not appropriate for this type of data. Linear regression models typically include variables for factors such as race, which have a finite set of values (fixed effect variables) but they cannot accommodate a variable that might be designated for facilities like dialysis centers where there are a range of values that are a subset of all possible dialysis centers (*random effects* variable) (14,15). Mixed effect linear models that include both fixed effects and random effects variables, are generally better suited for clustered data. The variance related to random effects in mixed effects linear models accounts for the *between center* variation that is not explained by the covariates in the model and can be used to quantify *within center* correlations (16).

The basic mixed effect linear model used to estimate the *within center* correlation in this study was set up with a response variable  $Y_{ij}$ , which represented the URR value for the  $j$ th subject in facility  $i$ . The predictor variables in the model included the patient and facility level characteristics listed in Tables 1 and 2 and were considered fixed-effect covariates. The facility itself was treated as a random effect variable with a *between center* variance  $\sigma_b^2$ . We assumed the within facility error term for URR values followed a normal distribution with a *within center* variance  $\sigma_e^2$ . The total variance,  $\sigma^2$ , of individual URR values from a population of dialysis patients can be broken down into two components:

$$\sigma^2 = \sigma_b^2 + \sigma_e^2 \quad (2)$$

The components of variance can be used to calculate the intraclass correlation parameter  $\rho$ , which estimates the average *within center* correlation of dialysis centers in the sample:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2} \quad (3)$$

Table 1. Characteristics of patients in Network 6 before and after a quality improvement program (QIP) to enhance dialysis adequacy

Patient Characteristics	Pre-QIP Sample (n = 5309)	Post-QIP Sample (n = 5753)
URR (± SE)	64.7 ± 0.1	69.8 ± 0.1 <sup>a</sup>
URR coefficient of variation <sup>b</sup>	12.1%	12.1%
Age (yr)	57.7 ± 0.2	59.2 ± 0.2 <sup>a</sup>
Gender		
male	2472 (46.6%)	2780 (48.3%)
female	2837 (53.4%)	2673 (51.7%)
Race		
African American	3828 (72.1%)	4112 (71.5%) <sup>a</sup>
White	1397 (26.3%)	1427 (24.8%)
other	84 (1.6%)	214 (3.7%)
Cause of renal failure		
diabetes	1709 (32.2%)	1962 (34.1%) <sup>a</sup>
all other causes	3600 (67.8%)	3791 (65.9%)
Time on dialysis (min)	208.8 ± 0.5	217.9 ± 0.5 <sup>a</sup>
Blood flow (ml/min)	374.9 ± 0.9	410.6 ± 0.9 <sup>a</sup>
Serum albumin (mg/dl)	3.8 ± 0.1	3.9 ± 0.1 <sup>a</sup>

<sup>a</sup> P < 0.05.

<sup>b</sup> Coefficient of variation is equal to 100% × (standard deviation of URR)/(mean of URR).

**Table 2.** Facility characteristics of dialysis units included in Network 6 dialysis adequacy improvement program

Free-standing	181 (92%)
Hospital-based	15 (8%)
For-profit	160 (82%)
Nonprofit	36 (18%)
Corporate-owned	166 (85%)
Private	30 (15%)
Attending physician	
permanent	83 (42%)
rotating	113 (58%)
Physician visits	
fewer than 1 per mo	17 (9%)
once or three times a mo	51 (26%)
approximately weekly	99 (51%)
more than once a week	29 (14%)
Peritoneal dialysis availability	
at the facility	93 (47%)
at an affiliated facility	65 (33%)
not available	38 (19%)
Accepts transients	186 (95%)
Does not accept transients	10 (5%)
Proportion of nursing home patients	
10% or less	166 (85%)
11 to 20%	23 (12%)
more than 20%	7 (3%)
Patient care conference	
monthly	88 (45%)
quarterly	21 (11%)
less often than quarterly	87 (44%)
Vital signs monitoring on dialysis	
every 15 minutes	11 (6%)
every 30 minutes	144 (73%)
every hour	41 (21%)

This ratio can be interpreted by considering opposite scenarios for individuals in dialysis centers and their measured URR values. If all dialyzed subjects in a facility are independent without a center effect, then there will be no relationship between their measured URR values, and the associated facility mean URR can be considered a random sample of the entire population of URR values. Therefore, each center mean value will approximate the population mean,  $\sigma_b$  will be small, and the total variance will approach the within center variance,  $\sigma_e^2$ . In this scenario,  $\rho$  will approach 0 when there is no *within center* correlation. However, in a scenario where observations within dialysis centers are highly correlated and have a strong center effect, the mean of each dialysis center will have the potential of varying greatly from the population mean with a large variance  $\sigma_b$  and  $\rho$  will approach one with a strong degree of *within center* correlation (14).

### Analysis

The presence and significance of a center effect on dialysis adequacy was determined using the *within center* correlation,  $\rho$ , in URR values across all facilities and the *between center* variation in facility mean URR values, which is statistically related to  $\rho$  (see statistical

model). These measurements were made on the pre-QIP and then the post-QIP sample. The  $\rho$  estimates of *within center* correlations across the study samples were calculated with a mixed effect linear model, which adjusted for patient and center-specific covariates.

To further illustrate the impact of the QIP on the center effect related to dialysis adequacy as measured by URR in the Network 6 dialysis population we examined the *between center* variation in facility mean URR values by comparing the distribution of facility mean URR values before the QIP to the corresponding distribution after the QIP. To depict differences in the center effect across facilities with varying levels of performance in the sample and demonstrate that changes in the center effect were independent of any regression to the overall mean, the units were stratified into tertiles. Facilities were assigned to tertiles on the basis of the rank of that unit's mean URR value in both the pre-QIP and the post-QIP survey. The parameter  $\rho$  was then calculated for the centers within each tertile. We used a mixed effect model that adjusted for patient-specific covariates and restricted the  $\sigma_b^2$  to the centers within each tertiles while assuming the same  $\sigma_e^2$  for all tertiles. The 95% confidence intervals for  $\rho$  were constructed on the basis of the approximate normality of the estimates of the two variance components,  $\sigma_b^2$  and  $\sigma_e^2$ . The SAS (SAS Institute, Cary NC) MIXED procedure with RANDOM statement and the method of restricted maximum likelihood were employed to calculate  $\rho$  estimates (17).

### Results

There were 196 centers included in both study samples, with 5309 patients in the pre-QIP and 5753 patients in the post-QIP samples. There were 527 (9.9%) and 575 (10.0%) patients in the pre-QIP and post-QIP samples, respectively, from the 10% of centers ( $n = 20$ ) in the pre-QIP distribution with the highest proportion of patients having a URR value less than 50%. There was a significant improvement in mean URR from the pre- to post-QIP samples (URR:  $64.7 \pm 0.1$  versus  $69.8 \pm 0.1$ , respectively;  $P = 0.001$ ); there was an even more substantial improvement in URR among the 20 centers who had the highest proportion of patients with a URR of 50% or less in the pre-QIP sample ( $57.9 \pm 0.4$  versus  $67.6 \pm 0.4$ , pre- versus post-QIP, respectively.) Table 1 illustrates patient characteristics, which were measured as part of the QIP. While there was a substantial improvement in overall mean URR from the pre to post-QIP samples, the two distributions had similar degrees of variability as indicated by the equal coefficients of variation. There were similar racial and gender distributions in the pre and post-QIP samples; however, the patients sampled in the post-QIP data set had a higher mean time on dialysis, average blood flow, and serum albumin.

Table 2 displays the facility characteristics of the centers included in the QIP. The facilities were more likely to be free-standing, for-profit, and corporate-owned. Most centers accepted transients, had rotating physicians that rounded at least weekly and a low percentage of nursing home patients. Vital signs were generally not checked more than once every 30 min, there was a variable frequency of patient care conferences, and availability of peritoneal dialysis.

Table 3 lists the intraclass correlation,  $\rho$ , for both the pre and post-QIP study samples across all facilities and for the centers then grouped into tertiles on the basis of the rank of each facility's mean URR value in the pre-QIP and then in the

Table 3. Mean URR (95% CI) and intraclass correlations ( $\rho$ ) for 196 centers before and after a Network-wide quality improvement program (QIP)

Facilities	Pre-QIP		Post-QIP	
	Mean URR	$\rho$	Mean URR	$\rho$
All centers	64.7 (64.4 to 64.9)	0.15 (0.12 to 0.18)	69.8 (69.6 to 70.0)	0.06 (0.04 to 0.08) <sup>b</sup>
Centers in tertiles				
lowest tertile	60.7 (60.3 to 61.1)	0.28 (0.19 to 0.36)	69.1 (68.7 to 69.5)	0.07 (0.04 to 0.10) <sup>b</sup>
middle tertile	65.1 (64.7 to 65.4)	0	69.9 (69.5 to 70.3)	0.09 (0.05 to 0.12) <sup>b</sup>
highest tertile	68.2 (67.9 to 68.5)	0.17 (0.10 to 0.23)	70.4 (70.0 to 70.8)	0.06 (0.03 to 0.09) <sup>b</sup>

	Pre-QIP Ranking <sup>a</sup>		Post-QIP Ranking <sup>a</sup>	
	Mean URR	$\rho$	Mean URR	$\rho$
	69.1 (68.7 to 69.5)	0.07 (0.04 to 0.10) <sup>b</sup>	66.7 (66.3 to 67.1)	0.18 (0.11 to 0.24) <sup>c</sup>
	69.9 (69.5 to 70.3)	0.09 (0.05 to 0.12) <sup>b</sup>	70.0 (69.7 to 70.4)	0
	70.4 (70.0 to 70.8)	0.06 (0.03 to 0.09) <sup>b</sup>	72.5 (72.2 to 72.9)	0.05 (0.02 to 0.09) <sup>b</sup>

<sup>a</sup> Centers with post-QIP values divided into tertiles based on rank of mean URR before and after the QIP.

<sup>b</sup>  $P \leq 0.01$  when compared with pre-QIP  $\rho$ .

<sup>c</sup>  $P = 0.06$  when compared with pre-QIP  $\rho$ .

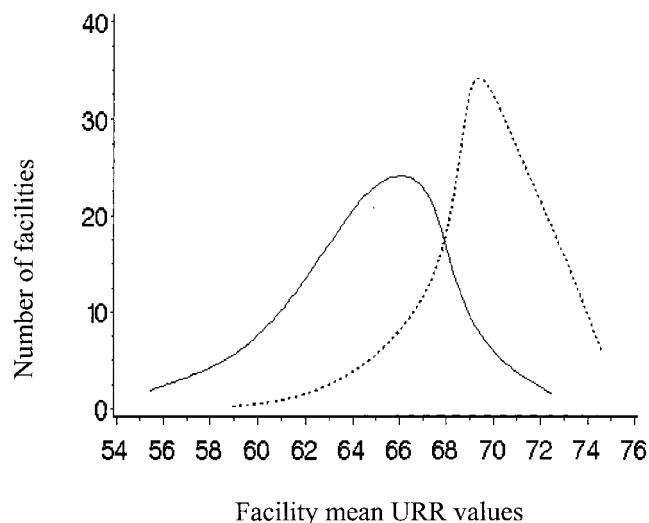
post-QIP sample. There was a substantial *within center* correlation across all facilities in the pre-QIP sample and when the facilities were stratified into tertiles on the basis of the mean URR values, the strongest *within center* correlation was among patients from the tertile of facilities with the lowest mean URR values in the pre-QIP sample. There was a significant but lower *within center* correlation among patients in the highest tertile but no correlation among patients assigned to the middle tertile. When contrasting the  $\rho$  values from the two samples, there was a significant attenuation in the overall degree of *within center* correlation across all facilities in the pre-QIP to post-QIP sample. The most significant decline in  $\rho$  from the pre to post-QIP samples was observed among centers assigned to the tertile with the lowest mean URR values in the pre-QIP sample. The  $\rho$  estimate also diminished in centers assigned to the highest tertile on the basis of pre-QIP URR values; those centers in the mid-tertile of facilities when ranked by pre-QIP URR values became correlated in adequacy post-QIP, but the *within center* correlation was similar to those found in the highest and lowest quartile.

After reclassifying centers and their associated patients into tertiles on the basis of their post-QIP mean URR rank, there was a similar trend in  $\rho$  values when contrasting the post-QIP with the original pre-QIP estimates. The estimate of the *within center* correlation observed among centers in the bottom third of facilities on the basis of post-QIP rank showed a trend of being lower than the  $\rho$  value for centers in the lower tertile of facilities based on pre-QIP URR values and rank. However, the reduction in  $\rho$  was less substantial than that seen for centers with their post-QIP URR values and retained in the lowest tertile based on pre-QIP mean URR rank. The centers reassigned to the highest tertile on the basis of post-QIP mean URR rank also had a diminished estimate for  $\rho$  relative to the centers in the pre-QIP highest tertile, but those centers allocated to the mid-tertile in the post-QIP tertile continued to show no correlation in URR values.

Corresponding to the changes in  $\rho$  was a reduction in *between center* variation in mean URR from the pre-QIP to the post-QIP sample. Figure 1 illustrates the distribution of facility means of URR values for the pre and post-QIP samples and shows the broader between-center variation in the former relative to the latter. There was a leftward skew in the post-QIP facility mean distribution reflecting the general improvement in adequacy across the Network as a result of the QIP. Of the 20 centers, which were in the lowest 10<sup>th</sup> percentile of the pre-QIP sample, 15 moved out of the bottom decile in the post-QIP sample.

### Discussion

In this study we have confirmed that there is a strong center effect on dialysis adequacy as demonstrated by the magnitude of the *within center* correlation and the wide *between center* variation in URR values noted in the Network 6 pre-QIP period. The center effect was greatest for those centers that generally performed poorly followed by those facilities that excelled in overall adequacy suggesting that different factors might be responsible for the center effect experienced by the



*Figure 1.* The frequency distribution of facility means of individual urea reduction ratio (URR) values in the pre-quality improvement project (QIP) and post-QIP samples. The mean value, SD, and interquartile range of the facility means with URR values of the pre-QIP (solid line) versus post-QIP (dashed line) samples were 64.7, 3.6, and 4.3 versus 69.8, 2.8, and 3.2, respectively.

different units. A QIP implemented in the Network led to an overall improvement in individual URR values; however, the variability in the distribution of individual URR values did not change from the pre- to post-QIP samples. In addition to the observed improvement in mean adequacy, there was a diminished *within center* correlation in URR values as measured by the intraclass correlation,  $\rho$ . The QIP also resulted in a reduction in the *between center* variation in mean URR values with a leftward skew in the distribution of facility mean URR reflecting an overall improvement in facility performance.

The reduced center effect was not only observed across the entire set of facilities but also could be demonstrated in various portions of the facility distribution. Centers tracked within their initial tertile classification based on pre-QIP mean URR rank demonstrated an “equalization” of *within center* correlation values across the tertiles, which suggests that the effect of the QIP was to impose a “best” practice policy across all units in the network. However, even after reassigning the set of centers to tertiles on the basis of their new ranking in the post-QIP sample a year later, there was a trend of lower *within center* correlations within the high and low groups.

The assumption that subjects are independent in population-based clinical trials has been challenged frequently, and much has been written about methods to accommodate the correlation that occurs within clinics, classrooms, communities, and multi-center intervention trials (15,18–21). The emphasis has been placed on treating the intraclass correlation inherent to clustered subjects as a “nuisance” parameter, especially in intervention trials, where such correlation results in diminished confidence intervals because of reductions in effective sample size. Several reviews have also discussed how diminished effective samples related to this type of correlation result in

inflation of variance and impact on power calculations (18–21). Less has been written, however, about the intraclass correlation as a parameter of interest with utility for investigators. Murray *et al.* surveyed community-based studies of smoking interventions in classrooms of students and described variations in intraclass correlation within classrooms across the studies (22). Using data from the Health Survey for England, Guillford *et al.* reported several levels of intraclass correlation from the district health authority to the household (23).

An ESRD population is distributed in a way that lends itself to clustering or a center effect. Patients who receive hemodialysis in the United States are organized into ESRD Networks; within each network, individuals aggregate in a smaller set of dialysis facilities. There are several reasons to believe that individuals who receive their renal replacement at a facility are more likely to behave the same than patients from different units. Certainly, facilities may attract patients with common values or habits, rendering them similar, or patients within a facility may interact in ways that make them alike in their behaviors (24). With regard to dialysis adequacy, which is a fixed measurement with small potential for confounding by behavioral factors, it is most likely that characteristics of the facility lead to correlation among individuals within units in dialysis adequacy. Such policies and procedures, which might be either planned or unintentional, are unique to all facilities, and these practice patterns are likely to account for the center effect leading to correlation between subjects or between center variations in dialysis adequacy.

Previous reports have shown that dialysis patients do not adhere to the assumption of independence but do demonstrate a strong degree of *within center* correlation and *between center* variation in outcomes that have been attributed to center effects inherent to dialysis facilities. In a cross-sectional analysis of URR data collected from ESRD patients who received dialysis in ESRD Network 5 there was a significant *within center* correlation estimated by the parameter  $\rho$ , across all the facilities studied. The extent of correlation was not dependent on basic facility characteristics or timing of post-BUN sampling. The substantial *within center* correlation in adequacy corresponded to a wider *between center* variation in mean URR than was expected if the patients had not been influenced by center effects (12). Another analysis conducted on a different sample of dialysis patients from the same ESRD Network confirmed the presence of a significant *within center* correlation among dialysis patients from the same center and wide *between center* variations in achieved adequacy. The impact of the center effect on adequacy was determined to be equivalent to the patient-specific factors that were presumed to affect adequacy in ESRD patients (25). McClellan *et al.* have also confirmed a substantial center effect influencing dialysis outcomes by demonstrating that there is a wide variation in achieved dialysis adequacy across facilities in Network 6 (26). Likewise, they also demonstrated a wide variation in mortality rates found within dialysis facilities, even after controlling for known risk factors for death (26,27).

In this study, we tested the primary hypothesis that a center effect on dialysis adequacy is an indication of overall poor

performance in a population of dialysis patients by measuring the center effect that was present in Network 6 before a QIP. We then tracked changes in the center effect after implementation of a successful QIP in the Network and demonstrated the utility of the *within center* correlation and *between center* variation in adequacy as a gauge for testing efficacy of a QIP. There is a national standard for dialysis adequacy (7), and it is expected that patients will distribute around the standard for adequacy; it is therefore sub-optimal that patients should adhere to distributions that are unique to their facility rather than the universal standard. Indeed some facilities exceed expectations with better mean adequacy than others do, but there should be uniformity in the practice patterns across units to ensure adherence with the universal standard. It is quite possible that the primary effect of a QIP is to standardize the care delivered across a Network.

Retrospective observational analyses have inherent weaknesses that should be taken into consideration when interpreting the results of this study. The patients were not randomly assigned to dialysis facilities before study inception; it is therefore difficult to exclude the possibility that individuals had an affinity for certain units that might contribute to their correlation in adequacy. The analysis attempted to get at the potential for confounding by reporting *within center* correlation estimates that were adjusted for measured patient characteristics. Furthermore, it would be difficult for behavioral factors to have a direct influence in the fixed procedure of blood sampling for URR measurement.

One must also consider whether regression to the mean was relevant to the diminished center effect observed in this study. First, it is important to remember that regression to the mean is the result of random variation in a population (28). Therefore, although a set of subjects in a study sample might have extreme values at one time point, on repeated measurement these subjects regress to the mean while other subjects take their place with extreme values. In Network 6, we showed that the overall center variance diminished with a leftward skew not observed before the QIP, and while the underperformers showed significant improvement in dialysis adequacy they were not replaced by other new outlying centers as might be expected with random variability. Second, the overall variability in adequacy of the study population did not change as was demonstrated by comparing the coefficient of variation of the pre-QIP to post-QIP distributions of individual URR values. Finally, to demonstrate that our findings were not biased by regression to the mean we stratified the centers into tertiles, which focused our *within center* correlation determinations to restricted segments of the distribution. The stratified analysis limited the influence of any changes in the overall sample variance and any potential regression to the mean, enabling us to demonstrate the same reduction in *within center* correlation within segments of the entire sample of centers.

The approach we used to measure the efficacy of the QIP in improving dialysis adequacy has only been applied to network-wide samples of dialysis patients and has not been used in smaller sets of patients in limited numbers of dialysis units. It is unknown how sensitive measurement of *within center* cor-

relations or *between center* variations in adequacy is in detecting important factors responsible for the observed center effect, especially in smaller samples of patients and units. Furthermore, it is not known whether these methods are superior to the standard analytic tools of quality improvement, which are often used to analyze barriers to good care at the facility level. It is quite possible that stepwise addition of measured facility characteristics as covariates to mixed effect linear models while monitoring changes in  $\rho$  will provide insight into the critical factors that account for deleterious center effects in adequacy.

In conclusion, the study findings provide strong evidence that a measured center effect in dialysis adequacy should be treated as a shortcoming in quality for a given dialysis region. One potential way to measure the success of a QIP applied to a large population of dialysis patients to improve adequacy is by tracking changes in center effects as reflected in alterations in the estimated *within center* correlation and *between center* variation in adequacy. Future efforts to evaluate QIP in randomized controlled trials of adequacy will provide an ideal opportunity to confirm the importance of changes in center effects as a surrogate for quality improvement.

## Acknowledgments

The analyses upon which this publication is based were performed under Contract Number 500–97-E023 entitled, “ESRD Contract for Network 6,” sponsored by the Health Care Financing Administration, Department of Health and Human Services. The conclusions and opinions expressed and methods used herein are those of the author. They do not necessarily reflect HCFA policy. The author assumes full responsibility for the accuracy and completeness of the ideas presented. Ideas and contributions to the author concerning experience in engaging with issues presented are welcomed.

## References

1. National Institutes of Health. US Renal Data System: USRDS 2000 Annual Data Report. Bethesda, MD, National Institute of Diabetes and Digestive and Kidney Diseases, 1999
2. Owen WF, Lew NL, Liu Y, Lowrie EG, Lazarus JM: The urea reduction ratio and serum albumin concentration as predictors of mortality in patients undergoing hemodialysis. *N Engl J Med* 329: 1001–1006, 1993
3. Collins AJ, Ma JZ, Umen A, Keshaviah P: Urea index and other predictors of hemodialysis patient survival. *Am J Kidney Dis* 23: 272–282, 1994
4. Sehgal AR, Dor A, Tsai AC: Morbidity and cost implications of inadequate dialysis. *Am J Kidney Dis* 37: 1223–1231, 2001
5. Held PJ, Port FK, Wolfe RA, Stannard DC, Carroll CE, Daugirdas JT, Bloembergen WE, Greer JW, Hakim RM: The dose of hemodialysis and patient mortality. *Kidney Int* 50: 550–556, 1996
6. Held PJ, Blagg C, Liska D, Hakim R, Levin N: The dose of hemodialysis according to dialysis prescription in Europe and the United States. *Am J Kidney Dis* 20: S16–S21, 1992
7. NKF-DOQI Clinical Practice Guidelines for Dialysis Adequacy. New York, National Kidney Foundation, 1997
8. Coyne DW, Delmez J, Spence G, Windus DW: Impaired delivery of hemodialysis prescriptions: An analysis of causes and an approach to evaluation. *J Am Soc Nephrol* 8: 1315–1318, 1997

9. Sehgal AR, Snow RJ, Singer ME, Amini SB, DeOreo PB, Silver MR, Cebul RD: Barriers to adequate delivery of hemodialysis. *Am J Kidney Dis* 31: 593–601, 1998
10. Sargent J: Shortfalls in delivery of dialysis. *Am J Kidney Dis* 15: 500–510, 1990
11. Delmez JA, Windus DW: Hemodialysis prescription and delivery in a metropolitan community. The St. Louis Nephrology Study Group. *Kidney Int* 41: 1023–1028, 1992
12. Fink JC, Gardner JF, Armistead NC, Turner MS, Light PD: Within-center correlation in dialysis adequacy. *J Clin Epidemiol* 53: 79–85, 2000
13. McClellan WM, Soucie JM, Krisher J, Caruana R, Haley W, Farmer C: Improving the care of patients treated with hemodialysis: A report from the Health Care Financing Administration's ESRD Core Indicators Project. *Am J Kidney Dis* 31: 584–592, 1998
14. Donner A, Klar N: Statistical considerations in the design and analysis of community intervention trials. *J Clin Epidemiol*. 49: 435–439, 1996
15. Donner A, Klar N: Cluster randomization trials in epidemiology: Theory and application. *J Stat Plann Infer* 85: 1378–1383, 1994
16. Littell RC, Milliken Ga, Stroup WW, Wolfinger RD: SAS System for Mixed Models. Cary NC, SAS Institute Inc., 1996
17. SAS Institute Inc: SAS/STAT User's Guide. Cary, NC, SAS Institute Inc., 1999
18. Katz S, Zeger SL: Estimation of design effects in cluster surveys. *Ann Epidemiol* 4: 295–301, 1994
19. Kirkwood BR, Cousens SN, Victora CG, de Zoysa I: Issues in the design and interpretation of studies to evaluate the impact of community-based interventions. *Trop Med Int Health* 2: 1022–1029, 1997
20. Hannan PJ, Murray DM, Jacobs DR, Jr., McGovern PG: Parameters to aid in the design and analysis of community trials: Intraclass correlations from the Minnesota Heart Health Program. *Epidemiology* 5: 88–95, 1994
21. Localio AR, Berlin JA, Ten Have TR, Kimmel SE: Adjustments for center in multicenter studies: An overview. *Ann Intern Med* 135: 112–123, 2001
22. Murray DM, Rooney BL, Hannan PJ, Peterson AV, Ary DV, Biglan A, Botvin GJ, Evans RI, Flay BR, Futterman R : Intraclass correlation among common measures of adolescent smoking: Estimates, correlates, and applications in smoking prevention studies. *Am J Epidemiol*. 140: 1038–1050, 1994
23. Gulliford MC, Ukoumunne OC, Chinn S: Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: Data from the Health Survey for England 1994. *Am J Epidemiol* 149: 876–883, 1999
24. Donner A, Brown KS, Brasher P: A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *Int J Epidemiol* 19: 795–800, 1990
25. Fink JC, Blahut SA, Briglia AE, Gardner JF, Light PD: Effect of center- versus patient-specific factors on variations in dialysis adequacy. *J Am Soc Nephrol* 12: 164–169, 2001
26. McClellan WM, Soucie JM, Flanders WD: Mortality in end-stage renal disease is associated with facility-to-facility differences in adequacy. *J Am Soc Nephrol* 9: 1940–1947, 1998
27. McClellan WM, Flanders WD, Gutman RA: Variable mortality rates among dialysis treatment centers. *Ann Intern Med* 117: 332–336, 1992
28. Fisher LD, Van Belle G: *Biostatistics: A Methodology for the Health Sciences*. New York, John Wiley & Sons, 1993, pp 389