Statistical Methods for Cohort Studies of CKD: Prediction Modeling

Jason Roy,† Haochang Shou,‡ Dawei Xie,§ Jesse Y. Hsu,§ Wei Yang,§ Amanda H. Anderson,§ J. Richard Landis,§ Christopher Jepson,§ Jiang He,§ Kathleen D. Liu,§ Chi-yuan Hsu,§ and Harold I. Feldman§ on behalf of the Chronic Renal Insufficiency Cohort (CRIC) Study Investigators

Abstract
Prediction models are often developed in and applied to CKD populations. These models can be used to inform patients and clinicians about the potential risks of disease development or progression. With increasing availability of large datasets from CKD cohorts, there is opportunity to develop better prediction models that will lead to more informed treatment decisions. It is important that prediction modeling be done using appropriate statistical methods to achieve the highest accuracy, while avoiding overfitting and poor calibration. In this paper, we review prediction modeling methods in general from model building to assessing model performance as well as the application to new patient populations. Throughout, the methods are illustrated using data from the Chronic Renal Insufficiency Cohort Study.

Introduction
Predictive models and risk assessment tools are intended to influence clinical practice and have been a topic of scientific research for decades. A PubMed search of prediction model yields over 40,000 papers. In CKD, research has focused on predicting CKD progression (1,2), cardiovascular events (3), and mortality (4–6) among many other outcomes (7). Interest in developing prediction models will continue to grow with the emerging focus on personalized medicine and the availability of large electronic databases of clinical information. Researchers carrying out prediction modeling studies need to think carefully about design, development, validation, interpretation, and the reporting of results. This methodologic review article will discuss these key aspects of prediction modeling. We illustrate the concepts using an example from the Chronic Renal Insufficiency Cohort (CRIC) Study (8,9) as described below.

Motivating Example: Prediction of CKD Progression
The motivating example focuses on the development of prediction models for CKD progression. In addition to the general goal of finding a good prediction model, we explore whether a novel biomarker improves prediction of CKD progression over established predictors. In this case, urine neutrophil gelatinase–associated lipocalin (NGAL) was identified as a potential risk factor for CKD progression on the basis of a growing literature that showed elevated levels in humans and animals with CKD or kidney injury (2). The question of interest was whether baseline urine NGAL would provide additional predictive information beyond the information captured by established predictors.

The CRIC Study is a multicenter cohort study of adults with moderate to advanced CKD. The design and characteristics of the CRIC Study have been described previously (8,9). In total, 3386 CRIC Study participants had valid urine NGAL test data and were included in the prediction modeling. Details of the procedures for obtaining urine NGAL are provided elsewhere (2,10).

Established predictors include sociodemographic characteristics (age, sex, race/ethnicity, and education), eGFR (in milliliters per minute per 1.73 m²), proteinuria (in grams per day), systolic BP, body mass index, history of cardiovascular disease, diabetes, and use of angiotensin–converting enzyme inhibitors/angiotensin II receptor blockers. In this example, all were measured at baseline.

The outcome was progressive CKD, which was defined as a composite end point of incident ESRD or halving of eGFR from baseline using the Modification of Diet in Renal Disease Study equation (11). ESRD was considered to have occurred when a patient underwent kidney transplantation or began chronic dialysis. For the purposes of this paper, in lieu of a broader examination of NGAL that was part of the original reports (2,10), we will focus on occurrence of progressive CKD within 2 years from baseline (a yes/no variable).

Among the 3386 participants, 10% had progressive CKD within 2 years. The median value of urine NGAL was 17.2 ng/ml, with an interquartile range of 8.1–39.2 ng/ml. Detailed characteristics of the study population are given in the work by Liu et al. (2).

We excluded patients with missing predictors (n=119), leaving a total of 3033 with a valid NGAL measurement, no missing predictors, and an observed outcome. We made this decision, because the percentage of missing
data was low, the characteristics of those with observed and missing data were similar (data not shown), and we wanted to focus on prediction and not missing data issues. In practice, however, multiple imputation is generally recommended for handling missing predictors (12,13).

**Prediction Models**

It is important to distinguish between two major types of modeling that are found in medical research—associative modeling and prediction modeling. In associative modeling, the goal is typically to identify population-level relationships between independent variables (e.g., exposures) and dependent variables (e.g., clinical outcomes). Although associative modeling does not necessarily establish causal relationships, it is often used in an effort to improve our understanding of the mechanisms through which outcomes occur. In prediction modeling, by contrast, the goal is typically to generate the best possible estimate of the value of the outcome variable for each individual. These models are often developed for use in clinical settings to help inform treatment decisions.

Prediction models use data from current patients for whom only predictors are available to learn about the relationship between the predictors and outcomes. The models can then be applied to new patients for whom only predictors are available—to make educated guesses about what their future outcome will be. Prediction modeling as a field involves both the development of the models and the evaluation of their performance.

In the era of big data, there has been an increased interest in prediction modeling. In fact, an entire field, machine learning, is now devoted to developing better algorithms for prediction. As a result of so much research activity focused on prediction, many new algorithms have been developed (14). For continuous outcomes, options include linear regression, generalized additive models (15), Gaussian process regression (16), regression trees (17), and $k$-nearest neighbor (18) among others. For binary outcomes, prediction is also known as classification, because the goal is to classify an individual into one of two categories on the basis of their set of predictors (features in machine learning terminology). Popular options for binary outcome prediction include logistic regression, classification tree (17), support vector machine (19), and $k$-nearest neighbor (20). Different machine learning algorithms have various strengths and weaknesses, discussion of which is beyond the scope of this paper. In the CRIC Study example, we use the standard logistic regression model of the binary outcome (occurrence of progressive CKD within 2 years).

**Variable Selection**

Regardless of which type of prediction model is used, a variable selection strategy will need to be chosen. If we are interested in the incremental improvement in prediction due to a novel biomarker (like urine NGAL), then it is reasonable to start with a set of established predictors and assess what improvement, if any, the biomarker adds to the model. Variable selection is, therefore, knowledge driven. Alternatively, if the goal is simply to use all available data to find the best prediction model, then a data-driven approach can be applied. Data-driven methods are typically automated—the researcher provides a list of a large set of possible predictors, and the method will select from that a shorter list of predictors to include in a final model. Data-driven methods include criterion methods, such as maximizing Bayesian information criterion (21), regularization methods, such as Lasso (22), and dimension reduction methods, such as principal components analysis (23), when there is a large set of predictors. Which type of variable selection approach to use depends on the purpose of the prediction model and its envisioned use in clinical practice. For example, if portability is a goal, then restricting to commonly collected variables might be important.

**Performance Evaluation**

After a model is developed, it is important to quantify how well it performs. In this section, we describe several methods for assessing performance. Our emphasis is on performance metrics for binary outcomes (classification problems); some of the metrics can be used for continuous outcomes as well.

Typically, a prediction model for a binary outcome produces a risk score for each individual, denoting their predicted risk of experiencing the outcome given their observed values on the predictor variables. For example, logistic regression yields a risk score, which is the log odds (logit) of the predicted probability of an individual experiencing the outcome of interest. A good prediction model for a binary outcome should lead to good discrimination (i.e., good separation in risk scores between individuals who will, in fact, develop the outcome and those who will not). Consider the CRIC Study example. We fitted three logistic regression models of progressive CKD (within 2 years from baseline). The first included only age, sex, and race as predictors. The second model also included eGFR and proteinuria. Finally, the third model also included other established predictors: angiotensin–converting enzyme inhibitor/angiotensin II receptor blocker, an indicator for any history of cardiovascular disease, diabetes, educational level, systolic BP, and body mass index. In Figure 1, the risk scores from each of the logistic regression models are plotted against the observed outcome. The plots show that, as more predictors were added to the model, the separation in risk scores between participants who did or did not experience the progressive CKD outcome increased. In model 1, for example, the distribution of the risk score was very similar for both groups. However, in model 3, those not experiencing progressive CKD tended to have much lower risk scores than those who did.

**Sensitivity and Specificity**

On the basis of the risk score, we can classify patients as high or low risk by choosing some threshold, values above which are considered high risk. A good prediction model tends to classify those who will, in fact, develop the outcome as high risk and those who will not as low risk. Thus, we can describe the performance of a test using sensitivity, the probability that a patient who will develop the outcome is classified as high risk, and specificity, the probability that a patient who will not develop the outcome is classified as low risk (24).

In the CRIC Study example, we focus on the model that included all of the established predictors. To obtain
sensitivity and specificity, we need to pick a threshold risk score, above which patients are classified as high risk. Figure 2 illustrates the idea using two risk score thresholds—a low value that leads to high sensitivity (96%) but moderate specificity (50%) and a high value that leads to lower sensitivity (43%) but a high specificity (98%). Thus, from the same model, one could have a classification that is highly sensitive (patients who will go on to CKD progression almost always screen positive) and moderately specific (only one half of patients who will not go on to CKD progression will screen negative) or one that is highly specific but only moderately sensitive.

Receiver Operating Characteristic Curves and c Statistic

By changing the classification threshold, one can choose to increase sensitivity at the cost of decreased specificity or vice versa. For a given prediction model, there is no way to increase both simultaneously. However, both potentially can be increased if the prediction model itself is improved (e.g., by adding an important new variable to the model). Thus, sensitivity and specificity can be useful for comparing models. However, one model might seem better than the other at one classification threshold and worse than the other at a different threshold. We would like to compare prediction models in a way that is not dependent on the choice of risk threshold.

Receiver operating characteristic (ROC) curves display sensitivity and specificity over the entire range of possible classification thresholds. Consider again Figure 2. By using two different thresholds, we had two different pairs of sensitivity and specificity. We could choose dozens or more additional thresholds and record equally many pairs of sensitivity and specificity data points. These data points can be used to construct ROC curves. In
was whether NGAL had prediction value beyond that of the primary research question in the work by Liu et al. (2). Thus, there seems to be incremental improvement in the discrimination ability of the model with NGAL (red curve in Figure 3B) (AUC=0.69) compared to the model without NGAL (blue curve in Figure 3B) (AUC=0.69). A 45° line represents a prediction model equivalent to random guessing.

One way to summarize the information in an ROC curve is with the area under the curve (AUC). This is also known as the c statistic (25). A perfect model would have a c statistic of one, which is the upper bound of the c statistic, whereas the random guessing model would have a c statistic of 0.5. Thus, one way to compare prediction models is with the c statistic—larger values being better. The c statistic also has another interpretation. Given a randomly selected case and a randomly selected control (in the CRIC Study example, a CKD progressor and a nonprogressor), the probability that the risk score is higher for the case than for the control is equal to the value of the c statistic (26).

In Figure 3A, the ROC curve is displayed from a prediction model that includes urine NGAL as the only predictor variable. The c statistic for this model is 0.8. If our goal was simply to determine whether urine NGAL has prognostic value for CKD progression, the answer would be yes. If, however, we were interested in the incremental value of the biomarker beyond established predictors, we would need to take additional steps. In Figure 3B, we compare ROC curves derived from two prediction models—one including only demographics and the other including demographics plus urine NGAL. From Figure 3B, we can see that the ROC curve for the model with NGAL (red curve in Figure 3B) (AUC=0.82) dominates (is above) the one without NGAL (blue curve in Figure 3B) (AUC=0.69). The c statistic for the model with urine NGAL is larger by 0.13. Thus, there seems to be incremental improvement over a model with demographic variables alone. However, the primary research question in the work by Liu et al. (2) was whether NGAL had prediction value beyond that of established predictors, which include additional factors beyond demographics. The blue curve in Figure 3C is the ROC curve for the model that included all of the established predictors. That model had a c statistic of 0.9—a very large value for a prediction model. When NGAL is added to this logistic regression model, the resulting ROC curve is the red curve in Figure 3C. These two curves are nearly indistinguishable and have the same c statistic (to two decimal places). Therefore, on the basis of this metric, urine NGAL does not add prediction value beyond established predictors. It is worth noting that urine NGAL was a statistically significant predictor in the full model (P<0.01), which illustrates the point that statistical significance does not necessarily imply added prediction value.

It is also important to consider uncertainty in the estimate of the ROC curve and c statistic. Confidence bands for the ROC curve and confidence intervals for the c statistic can be obtained from available software. For comparing two models, a confidence interval for the difference in c statistics could be obtained via, for example, nonparametric bootstrap resampling (27). However, this will generally be less powerful than the standard chi-squared test for comparing two models (28).

A limitation of comparing models on the basis of the c statistic is that it tends to be insensitive to improvements in prediction that occur when a new predictor (such as a novel biomarker) is added to a model that already has a high c statistic (29–31). It also should not be used without additionally assessing calibration, which we next briefly describe.

**Calibration**

A well calibrated model is one for which predicted probabilities closely match the observed rates of the outcome over the range of predicted values. A poorly calibrated model might perform well overall on the basis of measures, like the c statistic, but would perform poorly for some subpopulations.

Calibration can be checked in a variety of ways. A standard method is the Hosmer–Lemeshow test, where

---

**Figure 3.** Receiver operating characteristics (ROC) curves for three different prediction models. A is the ROC curve when urine neutrophil gelatinase–associated lipocalin (NGAL) is the only predictor variable included in the model. The c statistic is 0.8. B and C each include two ROC curves. In B, the blue curve (area under the curve [AUC] = 0.69) is from the model that includes only demographic variables. The red curve (AUC=0.82) additionally includes urine NGAL. In C, the blue curve is for the model that includes all established predictors. The red curve includes all established predictors plus urine NGAL. The c statistics in C are both 0.9. This plots show that urine NGAL is a relatively good predictor variable alone and has incremental improvement over the demographics-only model but does not help to improve prediction beyond established predictors.
predicted and observed counts within percentiles of predicted probabilities are compared (32). In the CRIC Study example with the full model (including NGAL), the Hosmer-Lemeshow test on the basis of deciles of the predicted probabilities has a $P$ value of 0.14. Rejection of the test ($P<0.05$) would suggest a poor fit (poor calibration), but that is not the case here. Another useful method is calibration plots. In this method, rather than obtaining observed and expected rates within percentiles, observed and expected rates are estimated using smoothing methods. Figure 4 displays a calibration plot for the CRIC Study example, with established predictors and urine NGAL included in the model. The plot suggests that the model is well calibrated, because the observed and predicted values tend to fall near the 45° line.

**Brier Score**

A measure that takes into account both calibration and discrimination is the Brier score (33,34). We can estimate the Brier score for a model by simply taking the average squared difference between the actual binary outcome and the predicted probability of that outcome for each individual. A low value of this metric indicates a model that performs well across the range of risk scores. The perfect model would have a Brier score of zero. The difference in this score between two models can be used to compare the models. In the CRIC Study example, the Brier scores were 0.087, 0.075, 0.057, and 0.056 for the demographics-only, demographics plus NGAL, established predictors, and established predictors plus NGAL models, respectively. Thus, there was improvement in adding NGAL to the demographics model (Brier score difference of 0.012—a 14% improvement). There was also improvement by moving from the demographics-only model to the established predictors model (Brier score difference of 0.020). However, adding NGAL to the established predictors model only decreased the Brier score by 0.001 (a 2% improvement). Although how big of a change constitutes a meaningful improvement is subjective, an improvement in Brier score of at least 10% would be difficult to dismiss, whereas a 2% improvement does not seem as convincing.

**Net Reclassification Improvement and Integrated Discrimination Improvement**

Pencina et al. (35,36) developed several new methods for assessing the incremental improvement of prediction due to a new biomarker. These include net reclassification improvement (NRI), both categorical and category free, and integrated discrimination improvement. These methods are reviewed in detail in a previous Clinical Journal of the American Society of Nephrology paper (37), and therefore, here we briefly illustrate the main idea of just one of the methods (categorical NRI).

The categorical NRI approach assesses the change in predictive performance that results from adding one or more predictors to a model by comparing how the two models classify participants into risk categories. In this approach, therefore, we begin by defining risk categories as we did when calculating specificity and sensitivity—that is, we choose cutpoints of risk for the outcome variable that define categories of lower and higher risks. NRI is calculated separately for the group experiencing the outcome event (those with progressive CKD within 2 years in the CRIC Study example) and the group not experiencing the event. To calculate NRI, study participants are assigned a score of zero if they were not reclassified (i.e., if both models place the participant in the same risk category), a score of one if they were reclassified in the right direction, and a score of −1 if they were reclassified in the wrong direction. An example of the right direction is a participant who experienced an event who was classified as low risk under the old model but classified as high risk under the new model. Within each of the two groups, NRI is calculated as the sum of scores across all patients in the group divided by the total number of patients in the group. These NRI scores are bounded between −1 and one, with a score of zero implying no difference in classification accuracy between the models, a negative score indicating worse performance by the new model, and a positive score showing improved performance. The larger the score, the better the new model classified participants compared with the old model on the basis of this criterion.

We will now consider a three-category classification model, where participants were classified as low risk if their predicted probability was $<0.05$, medium risk if it was between 0.05 and 0.10, and high risk if it was above 0.10. These cutpoints were chosen a priori on the basis of what the investigators considered to be clinically meaningful differences in the event rate (2). The results for the comparison of the demographics-only model with the demographics and urine NGAL model are given in Figure 5, left panel. The NRIs for events and nonevents were 3.6% (95% confidence interval, −3.3% to 9.2%) and 33% (95% confidence interval, 30.8% to 35.8%), respectively. Thus, there was large improvement in risk prediction for non-events when going from the demographics-only model to the demographics and NGAL model. Next, we compared the model with all established predictors with the same model with urine NGAL as an additional predictor. The reclassification data are given in Figure 5, right panel. Overall, there was little reclassification for both events.

---

**Figure 4.** Calibration plot for the model that includes established predictors and urine neutrophil gelatinase–associated lipocalin. The straight line is where a perfectly calibrated model would fall. The curve is the estimated relationship between the predicted and observed values. The shaded region is ±2 SEMs.
and nonevents, indicating no discernible improvement in prediction when NGAL was added to the established predictors model, which is consistent with the c statistic and Brier score findings described above.

The categorical NRI depends on classification cutpoints. There is a category-free version of NRI that avoids actual classification when assessing the models. Although NRI has practical interpretations, there is concern that it can be biased when the null is true (when a new biomarker is not actually predictive) (38,39). In particular, for poorly fitting models (especially poorly calibrated models), NRI tends to be inflated, making the new biomarker seem to add more predictive value than it actually does. As a result, bias-corrected methods have been proposed (40). In general, however, it is not recommended to quantify incremental improvement by NRI alone.

**Decision Analyses**

The methods discussed above tend to treat sensitivity and specificity as equally important. However, in clinical practice, the relative cost of each will vary. An approach that allows one to compare models after assigning weights to these tradeoffs is decision analysis (41,42). That is, given the relative weight of a false positive versus a false negative, the net benefit of one model over another can be calculated. A decision curve can be used to allow different individuals who have different preferences to make informed decisions.

**Validation**

**Internal**

A prediction model has good internal validity or reproducibility if it performs as expected on the population from which it was derived. In general, we expect that the performance of a prediction model will be overestimated if it is assessed on the same sample from which it was developed. That is, overfitting is a major concern. A method that assesses internal validation or one that avoids overfitting should be used. Split samples (training and test data) can be used to avoid overestimation of performance. Alternatively, methods, such as bootstrapping and cross-validation, can be used at the model development phase to avoid overfitting (47,48).

**External**

External validity or transportability is the ability to translate a model to different populations. We expect the performance of a model to degrade when it is evaluated in

**Continuous or Survival Outcomes**

We have focused our CRIC Study example primarily on a binary outcome thus far (classification problems). However, most of the principles described above apply to continuous or survival data. Indeed, some of the same model assessment tools can also be applied. For example, for censored survival data, methods have been developed to estimate the c statistic (29,43–45). For continuous outcomes, plots of predicted versus observed outcome and measures, such as $R^2$, can be useful. Calibration methods for survival outcomes have been described in detail elsewhere but are generally straightforward to implement (46).

### Figure 5

<table>
<thead>
<tr>
<th>Demographics, with and without NGAL</th>
<th>Traditional risk factors, with and without NGAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events With NGAL</td>
<td>Events With NGAL</td>
</tr>
<tr>
<td>Without NGAL</td>
<td></td>
</tr>
<tr>
<td>&lt;5%</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
</tr>
<tr>
<td>5-10%</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>&gt;10%</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>37</td>
<td>24</td>
</tr>
<tr>
<td>82</td>
<td>32</td>
</tr>
<tr>
<td>10%</td>
<td>&gt;10%</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>181</td>
<td>214</td>
</tr>
<tr>
<td>202</td>
<td>245</td>
</tr>
<tr>
<td>226</td>
<td>247</td>
</tr>
</tbody>
</table>

### Figure 5 | Classification of patients in the event and nonevent groups in models with and without urine neutrophil gelatinase–associated lipocalin (NGAL). The counts in red are from patients who were reclassified in the wrong direction. The left panel shows the reclassification that occurs when urine NGAL is added to a model that includes only demographics. For events, $3+8+37=48$ were reclassified in the right direction, whereas $16+16+5=37$ were reclassified in the wrong direction. The net reclassification improvement (NRI) for events was $(48/306)−(37/306)=3.6%$. The NRI for nonevents was $(1153/2727)−(243/2727)=33%$. Thus, there was large improvement in risk prediction for nonevents when going from the demographics-only model to demographics and NGAL model. In the right panel, urine NGAL is added to a model that includes all established predictors. Overall, there was little reclassification for both events and nonevents. The NRI for events is simply $(6/306)−(8/306)=−0.65%$. The NRI for nonevents is $(94/2727)−(79/2727)=0.55%$. |
new populations (49–51). Poor transportability of a model can occur because of underfitting. This would occur, for example, when important predictors are either unknown or not included in the original model. Even if the associations between the predictors and outcome stay the same, there is still a possibility that the baseline risk may be different in the new populations. It is, therefore, important to check (via performance metrics described above) external validity and possibly recalibrate the model when applying the model to a new population. Suppose, for example, that our prediction model is a logistic regression, and we have adequately captured the relationship between predictors and the outcome. However, the baseline risk in the new population might be higher or lower. Baseline risk is represented by the intercept term. Therefore, a simple recalibration method is to re-estimate the intercept term. Similarly, for survival data, re-estimation of the baseline hazard function can be used for recalibration.

**Dynamic Prediction**

In this paper, we have focused on situations where a prediction model is developed using available data that would potentially be applied unchanged to the assessment of new patients. However, a growing area of research is dynamic prediction models, where models are frequently updated over time as new data become available (52,53). The major points of the article still hold for dynamic prediction, but these models have the potential to be more transportable in that they can rapidly adapt to new populations.

**Reporting**

After the prediction model has been developed, evaluated, and possibly, validated, the next step is to report the findings. A set of recommended guidelines for the reporting of prediction modeling studies, the TRIPOD statement, was recently published and includes a 22-item checklist (54,55). For each item in the checklist, there are detailed examples and explanations. We highly recommend reviewing this document before publishing any results. The document places a strong emphasis on being very specific about the aims and design of the study, the definition of the outcome, the study population, the statistical methods used, and discussion of limitations.

An important part of reporting is a discussion of the clinical implications. Typically, adoption of the prediction model should not be recommended if it has not been validated. After the prediction model has been validated, it could be used or further studied in a variety of ways. For example, a model that accurately predicts CKD progression across a wide range of populations would be helpful to provide clinicians and patients with prognostic information. Such models could inform clinical decision making. For example, high-risk patients might need to be followed more intensely (e.g., clinician visits every 3 months rather than every 12 months) and have evidence-based interventions to slow CKD rigorously applied (e.g., BP<130/80 mmHg for those with proteinuria). Another example is the decision to have surgery for arteriovenous fistula creation if ESRD is imminent. If such models are used for decision making, the relative costs of false positives and false negatives need to be assessed. Along the same lines, whether knowing the risk score improves patient care would need to be studied, possibly in a randomized trial.

**Concluding Remarks**

Prediction modeling is likely to become increasingly important in CKD research and clinical practice, with richer data becoming available at a rapid rate. This paper described many of the key methods in the development, assessment, and application of prediction models. Other papers in this CKD methods series will focus on complex modeling of CKD data, such as longitudinal data, competing risks, time-dependent confounding, and recurrent event analysis.

**Acknowledgments**

Funding for the Chronic Renal Insufficiency Cohort Study was obtained under a cooperative agreement from the National Institute of Diabetes and Digestive and Kidney Diseases (grants U01DK080990, U01DK060698, U01DK061022, U01DK061028, U01DK060988, U01DK060963, and U01DK060902). Additional funding was provided by grants K01DK092353 and U01DK05649 (CKDBiocon).

**Disclosures**

None.

**References**


