

# Biomarker Qualification Review for Total Kidney Volume

## EXECUTIVE SUMMARY

This is a summary of the reviews and recommendations by the members of the Biomarker Qualification Review Team (BQRT) of a submission by the Polycystic Kidney Disease Outcomes Consortium (PKDOC), herein referenced as Consortium or submitter. This document describes the data supporting the qualification of total kidney volume (TKV) measured at baseline as a prognostic enrichment biomarker to be used in combination with patient age and baseline estimated glomerular filtration rate (eGFR), to help identify those Autosomal Dominant Polycystic Kidney Disease (ADPKD) patients who are at greater risk for a substantial decline in renal function.

### a. Background

ADPKD, the most common hereditary kidney disease, is characterized by progressive enlargement of the kidneys due to cyst growth and formation. In up to half of those diagnosed with the disease, progressive kidney dysfunction develops over decades, with a typical age of onset of end stage renal disease in the mid to late 50s among those who progress to kidney failure. Currently there are no approved therapies in the United States to treat ADPKD.

The PKDOC's perspective is that TKV measured using imaging techniques such as magnetic resonance imaging (MRI), computed tomography (CT) and ultrasound (US), is a promising biomarker for tracking and predicting the natural history of ADPKD. The Consortium proposed the following context of use for TKV for clinical trial enrichment in patients with ADPKD: Baseline TKV can be applied as a prognostic enrichment biomarker that, in combination with patient age and baseline eGFR, can be used to help identify those ADPKD patients who are at the greatest risk for a substantial decline in renal function defined as (1) 30% worsening of eGFR, (2) 57% worsening of eGFR (equivalent to doubling of serum creatinine), or (3) End-Stage Renal Disease (ESRD, defined as dialysis or transplant).

### b. Sources of Data and Major Findings

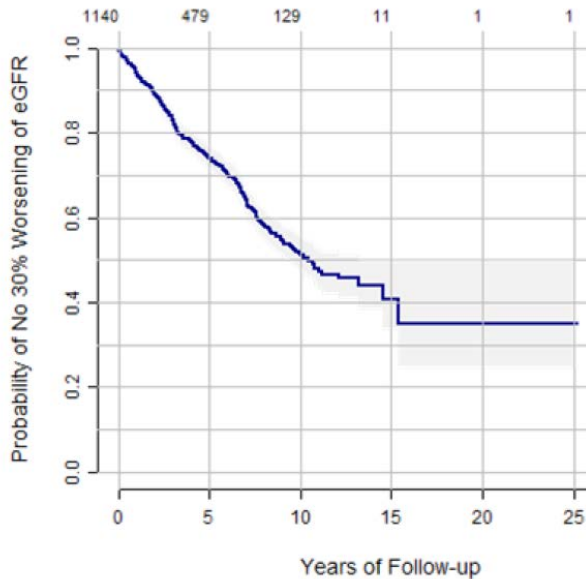
In support of the proposed context of use, the Consortium aggregated data from three patient registries (University of Colorado-Denver, Mayo Clinic and Emory University) and two longitudinal cohort studies (Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease 1 (CRISP1) and Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease 2 (CRISP2)) on the natural history of ADPKD. The common database contained data from a total of 2355 subjects with available TKV imaging. Imaging modalities used included MRI, CT and US.

**1. Submitter approach**

**30% worsening of eGFR**

A total of 2355 patients with at least one TKV measurement (all modalities) in the database were available. GFR was estimated using the MDRD equation. A total of 1215 patients with missing covariates, missing baseline eGFR, or an insufficient number of post-baseline eGFR measurements were excluded. Overall, the analysis dataset included 1140 patients of which 361 (31.7%) patients had a 30% worsening of eGFR (two measurements 30% lower than baseline). A Kaplan-Meier plot was generated for the probability of no 30% worsening of eGFR by years of follow-up, calculated from the time of the first TKV measurement to the time of the first of the qualifying eGFR measurements (see Figure 1).

**Figure 1. Kaplan-Meier plot for the probability of no 30% worsening of eGFR by years of follow-up**

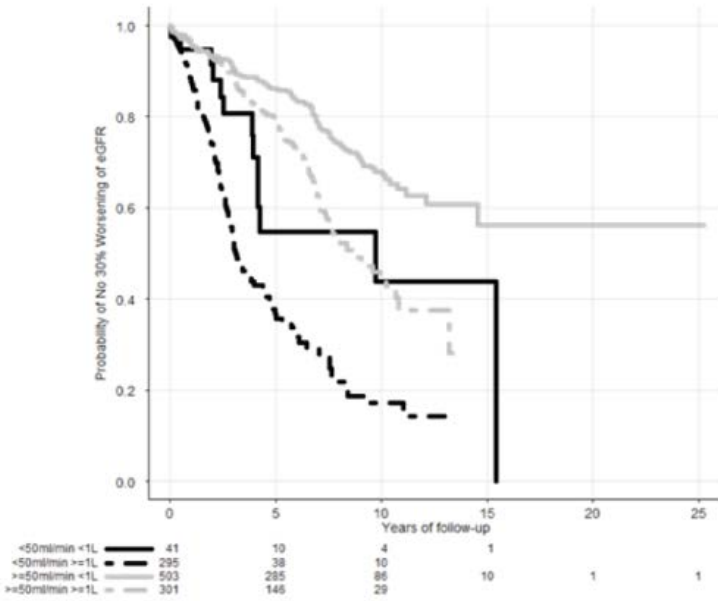


Source: Figure 26 of the final briefing book, page 114

The probability of reaching 30% worsening of eGFR at three years of follow-up was approximately 20% and increased to approximately 25% at five years of follow-up.

A Kaplan-Meier plot for the probability of avoiding a 30% decline of eGFR for different subgroups defined by baseline TKV (< 1 or ≥1 L) and baseline eGFR (< 50 or ≥ 50 mL/min/1.73m<sup>2</sup>) was constructed and is shown in Figure 2.

**Figure 2: Kaplan-Meier plot for the probability of avoiding a 30% decline in eGFR for different subgroups defined by baseline TKV and baseline eGFR**



Source: Figure 27 of the final briefing book, page 115

For patients with eGFR  $\geq 50$  mL/min per  $1.73 \text{ m}^2$ , the risk of a 30% worsening in ADPKD patients with larger TKV ( $\geq 1$  L) was greater than that observed in patients with smaller TKV ( $< 1$  L) (grey dashed vs. grey solid lines). For patients with eGFR  $< 50$  mL/min per  $1.73 \text{ m}^2$ , the risk of a 30% worsening of eGFR in ADPKD patients with larger TKV ( $\geq 1$  L) was greater than that observed in patients with smaller TKV ( $< 1$  L) (black dashed vs. black solid lines).

### Multivariate Cox Analysis

The submitter used multivariate Cox regression to investigate the relationship between the covariates and the time to 30% worsening of eGFR. The submitter showed that the three covariates age, baseline eGFR, and log-transformed baseline TKV were each associated with the time to 30% decline in eGFR. Note that “log” in “log-transformed baseline TKV” refers to taking the natural logarithm of the baseline TKV value. However, these covariates are not completely independent. The submitter presented the multivariate model that resulted in the highest area under the ROC curves at the 1-year and 5-year time points. The area under the curve was 0.75 and 0.70 at years 1 and 5, respectively. This model includes age, baseline eGFR, log{baseline TKV}, and all two-way interactions.

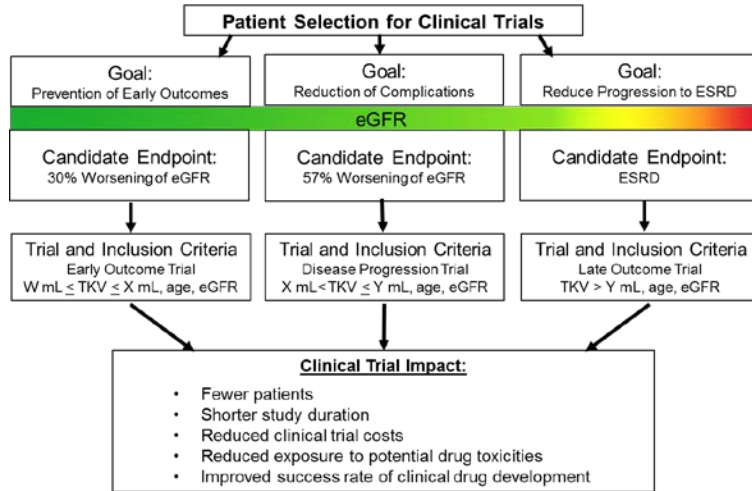
### 57% worsening of eGFR and ESRD

Similar analyses were performed for these two endpoints by the submitter. The submitter suggests that TKV is prognostic for selecting patients most likely to progress to these events.

Decision Tree

A decision tree was developed by the submitter to assist sponsors in the use of TKV as a prognostic enrichment biomarker for patient selection in clinical trials, see Figure 3.

**Figure 3. Decision tree for use of baseline TKV, eGFR and age for prognostic clinical trial enrichment**



Source: Figure 40 from the final briefing book, page 161

The submitter’s table below (Table 1) shows how the model components (baseline TKV, age, and baseline eGFR) interact, based on predicted probabilities of a 30% decline in eGFR according to selected example cut-offs for age (< 40 vs ≥ 40 years), baseline TKV (<1 L vs ≥ 1 L) and baseline eGFR (< 50 vs ≥ 50 mL/min per 1.73 m<sup>2</sup>).

**Table 1: A trial enrichment example**

Follow-Up Times (Years)	Probabilities of Avoiding 30% Worsening of eGFR							
	TKV < 1 L				TKV ≥ 1 L			
	Age: < 40 years		Age: ≥ 40 years		Age: < 40 years		Age: ≥ 40 years	
	eGFR ≥ 50 mL/min	eGFR < 50 mL/min	eGFR ≥ 50 mL/min	eGFR < 50 mL/min	eGFR ≥ 50 mL/min	eGFR < 50 mL/min	eGFR ≥ 50 mL/min	eGFR < 50 mL/min
1	0.991	0.992	0.992	0.991	0.984	0.982	0.985	0.979
2	0.980	0.980	0.981	0.979	0.963	0.959	0.966	0.953
3	0.950	0.949	0.951	0.947	0.907	0.899	0.915	0.884
4	0.917	0.916	0.918	0.913	0.852	0.839	0.863	0.815
5	<b>0.887</b>	0.888	0.889	0.884	<b>0.805</b>	0.789	0.818	<b>0.757</b>

Source: Table 49 from the final briefing book, page 162

For example, if a sponsor wishes to evaluate their therapy in younger patients (< 40) with more preserved renal function (eGFR ≥ 50), enrollment could be limited to patients with a TKV ≥ 1 L who have

an approximately 10% (or 20%) probability of reaching a 30% worsening in eGFR over three years (or five years). Based on these probabilities, statistical power calculations may be performed to determine the sample size needed for a 30% decline in eGFR, considering patient characteristics (age, baseline eGFR, and baseline TKV), the study duration, the probability of reaching the endpoint in the control arm, and the hypothetical effect of the therapeutic intervention on the outcome.

## **2. FDA approach**

### Data considerations

- The submitter's dataset included very young patients and also patients with end-stage disease (i.e., an eGFR < 15 mL/min/1.73m<sup>2</sup>) at baseline/time of entry into the dataset. The review team felt that the analyses would be more applicable to the trial setting if the criteria for including patients in the analysis reflected, to the extent possible, the design of clinical trials. Thus, the FDA analyses were limited to patients with an eGFR ≥25 and at least 12 years of age, which represents the population that the submitter felt was most likely to be enrolled in clinical trials. This resulted in 925 subjects with 300 events, compared to 1140 subjects with 361 events of 30% decline in eGFR in the submitter's analyses.
- Some subjects had imaging performed with more than one modality. Given data indicating that US measurements are less accurate and precise than MRI measurements, the FDA statistical reviewers used an MRI measurement if available. If no MRI measurement was available for a subject, then a CT measurement was used; if no MRI or CT measurement was available, then a US measurement was used. In contrast, the submitter took the average of the measurements obtained using different modalities if more than one was available.

**Note:** Caveats related to the relevance of the source data for predicting event risks in a clinical trial are discussed on pages 6-7 of the primary statistical review and evaluation and are summarized in the Appendix.

### 30% decline in eGFR

#### *Methodology*

A Cox regression modeling approach was used to predict the event risk at different time points in ADPKD subjects based on age, baseline eGFR, with and without baseline TKV. Cross-validation was performed using the submitter's dataset to assess the predictive ability of the modeling process. External validation was also performed using a separate dataset that was not publicly available.

TKV values measured by MRI, CT, and US modalities were used in the prediction modeling. Because the variability associated with the TKV imaging measurement differs among the three modalities, stratification by imaging modality in the prediction modeling was needed. For the outcome measure,

30% decline in eGFR; a subsequent measurement within any timeframe was required to confirm that the original decline was not transient. eGFR was estimated using the CKD-EPI equation.

The FDA reviewers searched for the best fit model without and with TKV (the original measurement or its log transformation) using the covariates age, baseline eGFR (mL/min per 1.73 m<sup>2</sup> or the log transformation) and the two-way interactions. Based on the Akaike information criterion (AIC), a measure of the relative quality of a statistical model for a given set of data, the best fit model without TKV included terms for age, log(eGFR) and the two-way interaction; the best fit model with TKV included log(TKV), age, eGFR and all two-way interactions. Here, log(TKV) refers to the log transformed baseline TKV value.

### Analyses

- Improvement in model fit:  
The FDA used the AIC to compare the best fit models with and without TKV. The improvement in AIC for the model with log (TKV) over the model without TKV was 86; This indicates substantial improvement in model fit.
- Improvement in discrimination:  
A C-statistic was used that measures the concordance between the patients' risk scores based on the survival model and their event times. This C-statistic was appropriate for right-censored survival data. There appeared to be an improvement of a fitted survival model including log(baseline TKV) as compared to not including log(baseline TKV) based on the C-statistics using the submitter's pooled dataset as can be observed from Table 2.

**Table 2. C-statistics for each model at different time points**

Time (years)	C no TKV	C with log(TKV)	p-value
1	0.424	0.617	0.002
2	0.493	0.660	0.0002
3	0.599	0.698	0.0008
5	0.617	0.703	0.0001
Maximum	0.598	0.686	0.002

Source: Table 9 of the primary statistical review and evaluation, page 15

At 1, 2, 3 and 5 years and over the maximum period of follow-up, the C-statistic was greater for the best fit model with baseline TKV than for the best fit model without, indicating that the model with baseline TKV was better able to discriminate patients who were likely to have a 30% decline in renal function at these time points from those who were not.

- Cross validation:

FDA performed cross-validation using the submitter's dataset. The submitter's dataset was randomly split into 5 mutually exclusive subsets of approximately the same size 181(=925/5) subjects per subset. The first four subsets were combined, the best models with and without log (TKV) were found for this pooled set of 724 subjects. Next, this model was used to predict the results in the fifth subset. The C-statistics were found as well as the event rates at years 2, 3, and 5 and then compared with the observed event rate in that subset. The entire process was done 5 times to allow each combination of 4 subsets to be pooled to serve as the training set and each subset not used in the training set to serve as the test set. The results obtained from the internal cross-validation described above shows that including TKV in the model provides a seemingly modest improvement over the best model using age and eGFR alone in terms of predictive performance on event risk.

- External validation using a separate dataset:

A separate dataset was used to evaluate whether the biomarker's predictive performance can be generalized to populations other than the one in which it was developed. FDA's independent validation using a separate dataset that was available internally supported the predictive performance of baseline TKV and hence baseline TKV's qualification as a prognostic enrichment biomarker.

- Potential utility of using TKV for trial enrichment :

Given questions about the underlying assumptions and approach used in the submitter's analysis reported in Table 1, a similar table was constructed relying on FDA's model. Overall, predicted event rates using the FDA model are higher with a baseline TKV  $\geq 1$  L, relative to baseline TKVs  $< 1$  L in these age and eGFR subsets. Predicted probabilities of not having a 30% worsening of eGFR within subsets defined by baseline age, baseline TKV and baseline eGFR cut-offs were calculated and the estimates and standard errors in the 8 subgroups are shown in Table 3. These estimates are found by calculating the estimated probability of an event for each subject (using their TKV, age, and eGFR) in the subset and then averaging across all the subjects in the subset.

**Table 3. Predicted probabilities of not having confirmed 30% decline in eGFR in subgroups defined by baseline age, baseline TKV and baseline eGFR (estimated standard errors in parentheses)**

Followup Time (Years)	Predicted probability of not having a confirmed 30% decline in eGFR							
	TKV < 1 L				TKV ≥ 1 L			
	Age < 40 years		Age ≥ 40 years		Age < 40 years		Age ≥ 40 years	
	eGFR ≥ 50 mL/min per 1.73 m <sup>2</sup>	eGFR < 50 mL/min per 1.73 m <sup>2</sup>	eGFR ≥ 50 mL/min per 1.73 m <sup>2</sup>	eGFR < 50 mL/min per 1.73 m <sup>2</sup>	eGFR ≥ 50 mL/min per 1.73 m <sup>2</sup>	eGFR < 50 mL/min per 1.73 m <sup>2</sup>	eGFR ≥ 50 mL/min per 1.73 m <sup>2</sup>	eGFR < 50 mL/min per 1.73 m <sup>2</sup>
1	0.971 (0.005)	0.982 (0.004)	0.977 (0.005)	0.954 (0.015)	0.960 (0.008)	0.916 (0.020)	0.961 (0.007)	0.914 (0.014)
2	0.944 (0.008)	0.965 (0.007)	0.958 (0.008)	0.919 (0.024)	0.922 (0.012)	0.851 (0.026)	0.928 (0.010)	0.851 (0.021)
3	0.895 (0.010)	0.938 (0.011)	0.916 (0.013)	0.853 (0.038)	0.846 (0.016)	0.735 (0.034)	0.858 (0.016)	0.730 (0.035)
4	0.860 (0.012)	0.920 (0.015)	0.879 (0.016)	0.803 (0.045)	0.787 (0.020)	0.654 (0.039)	0.800 (0.020)	0.640 (0.043)
5	0.828 (0.013)	0.897 (0.017)	0.854 (0.018)	0.765 (0.052)	0.738 (0.022)	0.594 (0.043)	0.758 (0.023)	0.581 (0.048)

Source: Table 13 of the primary statistical review and evaluation, page 20, and, the FDA model in Appendix E of the primary statistical review and evaluation, page 31.

The results shown in Table 3 can be compared to the estimated event rates using the Kaplan-Meier estimates (no model). For example, the predicted probability of not having a confirmed 30% decline in eGFR in subgroups defined by baseline covariates (age, eGFR and TKV) at year 3 using Kaplan-Meier estimates ignoring covariates do not differ among subgroups, see the table below. In contrast, the predicted probabilities of not having a confirmed 30% decline in eGFR based on the FDA model that includes the three baseline covariates are approximately 4% to 6% higher in the sub-category defined by eGFR ≥ 50 mL/min/1.73m<sup>2</sup> (relatively lower risk) in both age < 40 years and age ≥ 40 years compared to K-M estimates for subjects with TKV < 1L. A reverse trend is observed for subjects with TKV ≥ 1L. That is, these predicted probabilities are approximately 12% lower in the sub-category defined by eGFR < 50 mL/min/1.73m<sup>2</sup> (relatively higher risk) in both age < 40 years and age ≥ 40 years compared to K-M estimates. These predicted probabilities suggest that both baseline eGFR and baseline TKV may have better predictive performance than age, see Table 4.

**Table 4. Predicted probability of not having a confirmed 30% decline in eGFR in subgroups defined by baseline covariates (age, eGFR and TKV) at year-3 using Kaplan-Meier estimates ignoring covariates versus FDA model including all three covariates**

Predicted probability	TKV < 1L	TKV < 1L	TKV < 1L	TKV < 1L	TKV ≥ 1L	TKV ≥ 1L	TKV ≥ 1L	TKV ≥ 1L
	Age < 40yrs	Age < 40yrs	Age ≥ 40yrs	Age ≥ 40yrs	Age < 40yrs	Age < 40yrs	Age ≥ 40yrs	Age ≥ 40yrs
Estimation method	eGFR≥50	eGFR<50	eGFR≥50	eGFR<50	eGFR≥50	eGFR<50	eGFR≥50	eGFR<50
From K-M estimates	0.855	0.855	0.855	0.855	0.855	0.855	0.855	0.855
From FDA Model	0.895	0.938	0.916	0.853	0.846	0.735	0.858	0.730

Source: Table 2 of the secondary statistical review memo, page 6.



### 57% decline in eGFR and ESRD events

There were 354 ESRD events and 115 57% eGFR decline events in the submitter's dataset. In the restricted analysis set (i.e., patients with an eGFR  $\geq$  25 and age  $\geq$  12), there were 182 ESRD events and 99 57% eGFR decline events. Of the ESRD events, 47 occurred in the first five years and 12 events occurred in the first 3 years. Of the 57% decline in eGFR events, 34 occurred with the first five years. There were too few events over the timeframe of a feasible clinical trial to perform meaningful analyses.

### **c. Image Acquisition and Analysis Methods**

TKV was qualified based on a collection of data from multiple sites and modalities including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound (US). The University of Colorado-Denver study used US, the Emory study used MRI or US, while the Mayo Clinic study employed MRI or CT. CRISP1 used Gadolinium-enhanced MRI whereas CRISP II employed non-Gadolinium enhanced MRI. Image analysis methods included ellipsoid based methods (University of Colorado) and stereology techniques (Mayo Clinic, Emory, CRISP I, and CRISP II).

### Performance characteristics of the imaging modalities

The performance characteristics required of the TKV measurement, either at baseline or as an endpoint, will depend on the specific enrichment model implemented for the particular clinical trial. The FDA model stratifying on the imaging modality is an example of when multiple imaging modalities of differing precisions are used in a clinical trial.

The imaging modality or modalities selected should provide complete coverage of the kidneys with sufficient spatial resolution and tissue contrast to permit volumetric analysis of the kidneys. When deciding which imaging modality to use, sponsors should consider the accuracy and precision of the TKV imaging modality and measurement method as well as the risks associated with image acquisition. The accuracy and precision of the TKV measurement depend, in part, on the image analysis methodology used to determine TKV. In general, quantitative stereology, boundary tracing, and similar methods reduce uncertainty in TKV measurement compared to ellipsoid volume methods.

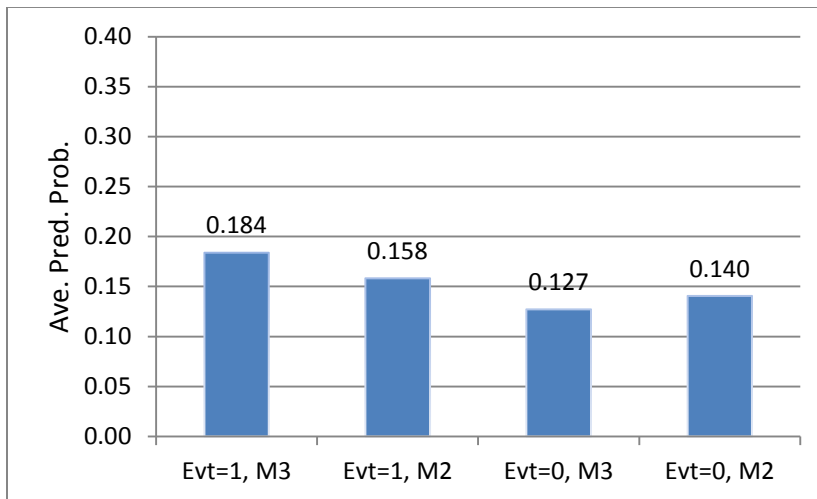
Reducing uncertainty in TKV measurement improves the reliability of the model used for trial enrichment. Sponsors should adopt well-defined, trial-specific standardized image acquisition and analysis protocols to minimize variability of imaging data and TKV measurements within trials. The adoption of standardized imaging acquisition protocols and standardized volume calculation methodology is encouraged to reduce uncertainty in TKV measurement and facilitate future analyses across studies. For details, please refer to, "Clinical Trial Imaging Endpoint Process Standards: Guidance for Industry."

**d. Enrichment considerations**

A statistically significant improvement in the predictive performance was quantified by the C-statistics from the best fit model without baseline TKV (“C without TKV”) to the best fit model with log(baseline TKV) (“C with log (TKV)”). This improvement is observed across the follow-up time in years, see Table 2.

Further analyses showed that compared to the FDA best fit model without baseline TKV (Model-2 or M2) at year 3, the FDA best fit model with baseline TKV (Model-3 or M3) tends to yield a higher predicted probability of not having a confirmed 30% decline in eGFR for ADPKD patients with baseline TKV < 1L and yield a lower predicted probability of not having a confirmed 30% decline in eGFR for ADPKD patients with a baseline TKV ≥ 1L. This finding is confirmed through model discrimination, see Figure 4 . By including baseline TKV in the model, there is a 2.6% absolute increase in the average predicted probability for the subjects experiencing a confirmed 30% decline in eGFR (evt=1 or event) and a 1.3% absolute decrease in the average predicted probability for the subjects without experiencing a confirmed 30% decline in eGFR (evt=0 or non-event). This results in an absolute change of 3.9% (= 2.6% + 1.3%). Interpretation of this absolute change depends on the background event rate observed in the data. A relative measure,  $(0.184 - 0.127) / (0.158 - 0.14) - 1 = 2.2$ , suggests that the difference in predicted probabilities between events and non-events is about 2.2-fold more in the best fit model with baseline TKV compared to best fit model without TKV.

**Figure 4. Average predicted probability by event status for Model-2\* and Model-3\*\* at year 3**



\*Model-2 (or M2): the FDA best fit model without baseline TKV

\*\*Model-3 (or M3): the FDA best fit model with baseline TKV

evt=1 refers to those ADPKD subjects having a confirmed 30% decline in eGFR

evt=0 refers to those ADPKD subjects not having a confirmed 30% decline in eGFR

Source: Figure 5 of the secondary statistical review memo, page 7

In addition, the total potential reclassification based on the best fit model with baseline TKV as compared to the best fit model without is shown to be positive. That is, there is a higher percentage being reclassified to higher risk categories for those having a confirmed 30% decline in eGFR and a higher percentage being reclassified to lower risk categories for those not having a confirmed 30% decline in eGFR with Model-3 than with Model-2.

*Prediction ability by imaging modality*

From the four data sources, about 50% of the subjects had their baseline TKV measurements taken by MRI, 25% by CT and 25% by US. The predicted probability of not having a confirmed 30% decline in eGFR by imaging modality based on the final prediction model that included baseline TKV showed that the interval prediction estimates by MRI is much more precise than those by CT or US. MRI has the smallest standard deviation of the predicted probability and CT has the largest standard deviation. See Figures 6 and 7 (pages 9 and 10) of the secondary statistical review memo.

*Number of ADPKD patients needed to screen versus to enroll in a clinical trial*

Two approaches were taken to determine the impact of using the best fit model with TKV on the number of patients needed to produce one event and the number that would need to be screened. The first approach considered age, eGFR and TKV as separate parameters/entry criteria for enrollment. The second approach used a patient's risk score (as defined by the model) to select patients for enrollment.

Table 5 shows the predicted event rate in the placebo group over 3 years, the number of subjects needed to produce one event and the number of patients that would need to be screened to enroll these subjects using the models with and without TKV and specifying separate entry criteria based on age, eGFR and TKV. Based on the entry criteria described above, 13 patients would need to be screened in the submitter's dataset to enroll 11 patients to get one event using the model without TKV. Using the model with TKV, approximately 25 patients would need to be screened in the dataset to enroll 9 patients to get one event.

**Table 5. Predicted event rate in placebo arm over 3 years, number needed to enroll and number needed to treat to get one event using the best fit models with and without TKV**

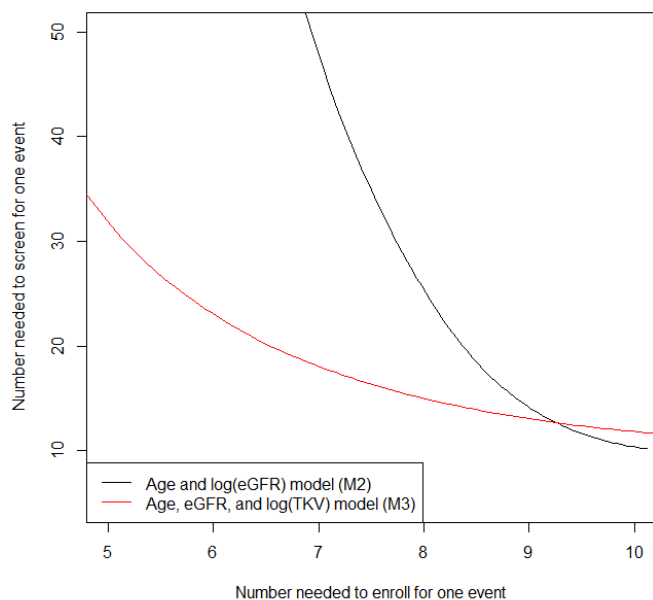
	Model without TKV‡	Model with TKV‡, using added criterion of TKV > 1 L
Predicted event rate in placebo arm over 3years	0.0905	0.11
Number needed to enroll†	11.05	9.09
Number needed to screen	13.01 <sup>λ</sup>	24.5§

‡Assumes entry criteria of eGFR ≥ 50 mL/min per 1.73 m<sup>2</sup> and age between 20 and 50 years; †Calculated as 1/the predicted event rate over 3 years; <sup>λ</sup> Calculated as 11.05 \*676/574; §Calculated as 9.09 \*676/251

Source: Email from the primary statistical reviewer dated May 21, 2015 summarized in Table 7 of the clinical review.

The second statistical analysis was restricted to patients who were between 20 and 50 years of age and the risk scores from the two models were used to define the entry criteria. Figure 5 shows the number needed to enroll for one event vs the number needed to screen as a function of the risk score. According to this analysis, if patients with the top 50% of risk scores are selected using the best fit model without TKV, then the number needed to enroll to get one event would be 8.6 and the number needed to screen would be 17.2. If patients with the top 50% of risk scores are selected using the best fit model with TKV, then the number needed to enroll decreases to 7.8 and the number needed to screen decreases to 15.6. These findings support the clinical utility of using the model with TKV to enrich the trial population. The findings suggest that using a multivariate risk score to enrich the trial population is more efficient than specifying independent entry criteria for the parameters of interest.

**Figure 5. Number needed to enroll for one event vs. number needed to screen using the risk scores from the two models to select patients**



Source: Email from the primary statistical reviewer dated May 29, 2015 summarized in Figure 6 of the clinical review.

An example of a sample size calculation incorporating baseline TKV can be found in pages 21-22 of the primary statistical review and evaluation. In the sample size illustration, the magnitude of new treatment improvement over placebo in a placebo controlled trial was assumed, e.g., an hazard ratio of 0.7. The calculation was based on FDA best fit model including baseline TKV (Model-3 or M3).

There are important factors to consider for predicting event rates to future clinical trials. For example, missing data, the potential for selection bias, as well as other differences in how covariates and endpoint events were defined in the database that were used to derive the model vs. a typical clinical trial setting, may affect the ability to generalize event rate predictions to future clinical trials.

**e. BQRT Conclusions**

Relative to a model that did not include log (TKV), a fitted survival model including log (TKV) improved the predictive performance of event risk (based on a concordance measure for time-to-event data) for a confirmed 30% decline in eGFR. The improvement is observed using either the submitter’s registry data alone for model development and cross validation or using clinical trial data that were available internally to FDA for independent validation. For the endpoints of 57% decline in eGFR and ESRD, there were too few events over the timeframe of a feasible clinical trial to perform meaningful analyses.

**f. BQRT Recommendations**

Based upon consideration of the strengths and limitations of the data, the BQRT recommends that Total Kidney Volume (TKV) determined at baseline, in combination with patient age and baseline eGFR, can be qualified as a prognostic enrichment biomarker for autosomal dominant polycystic kidney disease (ADPKD) subjects at high risk for a progressive decline in renal function, defined as a confirmed 30% decline in eGFR.

Additional considerations are provided below.

**Patient populations:**

For use in clinical trials of patients with ADPKD who are at least 12 years of age.

**TKV measurement:**

Various imaging modalities and post-processing methods are available to determine TKV. The uncertainty associated with the TKV measurement methodology may impact the reliability of the predicted clinical trial population enrichment.

## Appendix: Differences in registry/cohort studies or analysis methodologies versus clinical trials

There were several differences between the registry/cohort studies used for the qualification and the clinical trials for ADPKD. Some of the pertinent differences that can influence TKV and eGFR, specifically, confirmed 30% eGFR decline from baseline determinations are provided below.

- a. The population was not a random sample from the population of people with diagnoses of ADPKD. In addition, the people recruited for a clinical trial are not a random sample from the population.
- b. Generally, baseline eGFR is calculated as an average of two or more values taken within a short time (one or two weeks) before randomization or first dose of study drug in clinical trials, since the average is more reliable than single values. In the datasets provided by the submitter, only one baseline value was used and it could have been measured many months (up to a year) after the defined time 0 (the time of the baseline TKV measurement).
- c. The timing and frequency of serum creatinine measurements can affect the event rate. In a clinical trial, serum creatinine measurements are taken at defined time points such as every 3 months or every 4 months. In these datasets, measurements were not taken at defined time points. In some cases, measurements were taken every day for many days in a row; in other cases, there were gaps of a year or more between measurements.
- d. In a clinical trial, the endpoint of a confirmed 30% change in eGFR from baseline would need to be confirmed by the very next subsequent measurement. If the qualifying and confirmatory measurement were not consecutive, then, an event would not count. However, in these datasets, an event could count even when the qualifying and confirmatory events are years apart with many non-confirmatory measurements in between.
- e. In a clinical trial, an event of confirmed 30% worsening of eGFR would also include more severe endpoints including need for initiation of dialysis or transplant. In this dataset, subjects with these more severe events were censored at the time of the event.