

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

204441Orig1s000

STATISTICAL REVIEW(S)



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #: NDA 204-441
Supplement #:
Drug Name: Tolvaptan
Indication(s): Slow progressive kidney disease in adults with autosomal dominant polycystic kidney disease
Applicant: Otsuka
Date(s): 10/24/2017
Review Priority: Priority
Biometrics Division: DBI
Statistical Reviewer: John Lawrence, Ph D
Concurring Reviewers: Jim Hung, Ph D
Medical Division: Cardiorenal
Clinical Team: Melanie Blank, MD, Aliza Thompson MD, Nhi Beasley MD
Project Manager: Anna Park
Keywords:
survival analysis, benefit-risk, mixed models, longitudinal data analysis, surrogate endpoints

Table of Contents

EXECUTIVE SUMMARY	5
INTRODUCTION	6
1.1 OVERVIEW.....	6
1.2 DATA SOURCES	6
STATISTICAL EVALUATION	6
1.3 DATA AND ANALYSIS QUALITY	6
1.4 EVALUATION OF EFFICACY	7
1.4.1 <i>Study Design and Endpoints</i>	7
1.4.2 <i>Statistical Methodologies</i>	7
1.4.3 <i>Patient Disposition, Demographic and Baseline Characteristics</i>	7
1.4.4 <i>Results and Conclusions</i>	10
1.5 EVALUATION OF SAFETY	14
1.6 BENEFIT-RISK ASSESSMENT (OPTIONAL).....	14
FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	14
1.7 GENDER, RACE, AGE, AND GEOGRAPHIC REGION	14
1.8 OTHER SPECIAL/SUBGROUP POPULATIONS	15
SUMMARY AND CONCLUSIONS	15
1.9 STATISTICAL ISSUES	15
1.10 COLLECTIVE EVIDENCE.....	15
1.11 CONCLUSIONS AND RECOMMENDATIONS	15
1.12 LABELING RECOMMENDATIONS (AS APPLICABLE).....	16
APPENDICES.....	17
ESTIMATING eGFR OVER TIME GIVEN STARTING VALUES	17
TKV GROWTH IN TEMPO 3:4 AND TEMPO 4:4	18
CORRELATION BETWEEN CHRONIC TKV SLOPE AND CHRONIC eGFR SLOPE IN TEMPO 3:4.....	20
PROPORTION OF TREATMENT EFFECT ON CHRONIC eGFR SLOPE EXPLAINED BY CHRONIC TKV SLOPE IN TEMPO 3:4	20

LIST OF TABLES

Table 1 Patient disposition (Table 10.1-1 of Study Report)	8
Table 2 Patient demographic characteristics (Table 11.2-1 of Study Report)	9
Table 3 Patient demographic characteristics (Table 11..2-1 of Study Report)	10
Table 4 SAS GLM results for primary endpoint (p. 56 Section 16.1.9 of Study Report and confirmed by the FDA) 10	
Table 5 Linear Model results for modified model in Study 156-13-210	11
Table 6 Linear Model results for modified model in Study 156-04-251	12
Table 7 Linear Model results for longitudinal log(TKV) in TEMPO 3:4 (Study 156-04-251) (FDA analysis).....	18
Table 8 Linear Model results for longitudinal log(TKV) in TEMPO 3:4 (Study 156-04-251) (FDA analysis).....	18
Table 9 Linear Model results for longitudinal log(eGFR) in Study 156-04-251 (FDA analysis).....	20
Table 10 Linear Model results for longitudinal log(eGFR) adjusting for chronic TKV slope in Study 156-04-251 (FDA analysis).....	21

LIST OF FIGURES

Figure 1 Kaplan-Meier plot of time to discontinuation from randomized drug (Figure 10.1-1 of Study Report).	9
Figure 2 Predicted eGFR over time for two hypothetical patients. Black=no tolvaptan treatment, Red=tolvaptan immediately, Blue=tolvaptan 5 years after, Orange=tolvaptan 10 years after (FDA analysis).	13
Figure 3 Sponsor's results of key secondary endpoint within subgroups (Study Report Figure 11.4.1.1-3)	14
Figure 4 Percentage change in TKV from TEMPO 3:4 baseline to TEMPO 4:4 Month 24 visit. Open circles and triangles represent off-treatment time points. Break in treatment from TEMPO 3:4 to TEMPO 4:4 ranged from 13 to 829 days from the Month 36 visit in TEMPO 3:4. (Figure 2 of Torres, Vicente E., et al.)	19

EXECUTIVE SUMMARY

Two Phase 3 studies have been completed to assess the safety and efficacy of tolvaptan for the treatment of polycystic kidney disease (PKD). The first study was completed in 2012 and has been reviewed (June 2013). This review focuses mainly on the second trial. Based on the collective evidence of both studies, tolvaptan appears to slow the rate of decline of kidney function as measured by the estimated glomerular filtration rate (eGFR). [REDACTED] (b) (4)

[REDACTED] From the one study where all subjects had TKV measured annually, there appears to be a substantial acute hemodynamic effect, but little or no long term chronic effect on TKV. Depending on what model is used in the double-blind phase, there is either no chronic effect or a very small chronic effect. If there were a small effect, it may or may not be a clinical benefit. When using the model that suggests there is a small chronic effect on TKV, there is a very small correlation between chronic effect on TKV and chronic effect on eGFR. The proportion of the treatment effect on eGFR explained by TKV is very small in that trial. This suggests chronic effect on TKV is not a surrogate endpoint for effect on eGFR for this study population. In a long term extension phase of that trial, all subjects were treated with tolvaptan to determine whether there was a sustained difference between the two arms. At the end of this extension phase, the subjects who were treated 3 years longer with tolvaptan in the double-blind phase had no significant difference in TKV from the subjects that had been randomized to placebo. Both studies have been relatively short term (1-3 years) and have shown a modest effect on annualized change in eGFR. If approved, tolvaptan will be used chronically. It is natural to assume that the effect seen in one year (~1 mL/min) will compound each year, so that after 10 years there will be approximately 10 mL/min effect. We don't know if that is true, but we do have evidence from the two trials that tolvaptan works in both early and late stage of disease.

INTRODUCTION

1.1 Overview

Table: List of all studies included in analysis

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
<i>Study 156-04-251</i>	<i>Phase 3</i>	<i>36 months</i>	<i>36 months</i>	<i>tolvaptan: 961 placebo:484</i>	<i>subjects with ADPKD as defined by a certain number of cysts, estimated creatinine clearance of at least 60 mL/min and TKV>750 mL.</i>
<i>Study 156-13-210</i>	<i>Phase 3b</i>	<i>12 months</i>	<i>15 months</i>	<i>tolvaptan: 683 placebo:687</i>	<i>subjects with ADPKD and eGFR between 25 and 65 mL/min per 1.73 m².</i>

1.2 Data Sources

Electronic datasets and Study Reports:

<\\cdsesub1\evsprod\NDA204441\204441.enx>

<\\cdsesub1\evsprod\NDA204441\0001\m5\datasets\156-04-251\analysis>

<\\CDSESUB1\evsprod\NDA204441\0055\m5\53-clin-stud-rep\535-rep-effic-safety-stud\adpkd\5351-stud-rep-contr\study-156-13-210>

<\\CDSESUB1\evsprod\NDA204441\0055\m5\datasets\156-13-210\analysis\adam\datasets>

STATISTICAL EVALUATION

1.3 Data and Analysis Quality

Both were excellent in Study 156-13-210. Every step the sponsor could conceive- including all the recommendations the FDA made to the sponsor during the planning of this study- were taken to minimize missing data. A high percentage of subjects stopped taking the randomized study drug- more in the tolvaptan arm than in the placebo arm- but the percent of subjects with month 12 measurements was nearly 96% in both arms. Appropriate analyses were planned to assess the

sensitivity to different assumptions about missing data. Because of the acute hemodynamic effects of tolvaptan on eGFR, the design and analysis had to be planned very carefully, and both were done excellently. There were dietary recommendations for sodium, protein, and fluid. These were all excellent and improved the quality of the trial, particularly the encouragement to reduce meat protein. This trial is not just an exemplar for future trials in polycystic kidney disease, but for all clinical trials.

1.4 Evaluation of Efficacy

1.4.1 Study Design and Endpoints

Study 156-13-210 was a double-blind randomized study. The trial consisted of an 8-week pre-randomization period, a 12-month double-blind randomized treatment period, and up to 3-week final follow-up period. 683 (tolvaptan) and 687 (placebo) subjects were randomized. If a subject stopped taking randomized study drug, they were encouraged to remain in the study and return for the 12 month end-of-study assessment. The primary endpoint was the annualized change from baseline in eGFR. This was found by taking the off-treatment mean value at the end of the double-blind period and subtracting the pre-treatment mean value, then dividing this difference by the time elapsed. The sample size was chosen to have 90% power assuming a treatment effect of 1.07 mL/min per 1.73 m². There was an exploratory endpoint based on a PKD outcomes survey.

1.4.2 Statistical Methodologies

For each subject, the annualized change and weights were calculated based on the precision of the estimate. Then, a weighted linear model was used to estimate the difference in mean annualized rate of change in eGFR. The model also included the continuous variable baseline eGFR and categorical variables for age, TKV size, and eGFR category.

1.4.3 Patient Disposition, Demographic and Baseline Characteristics

The patient disposition is shown in Table 1. Almost 96% of the subjects in both arms completed the month 12 visit. Significantly more subjects discontinued treatment in the tolvaptan arm. Figure 1 shows that the proportion of subjects in the tolvaptan arm who discontinued study drug was larger than the proportion in the placebo arm starting from the second month throughout the rest of the trial.

	Tolvaptan (N=683)	Placebo (N=687)	Total (N=1370)
Subjects screened			2292
Subjects entering the placebo run-in phase			1519
Subjects entering the tolvaptan run-in phase			1496
Subject randomized	683 (100)	687 (100)	1370 (100)
Completed month 12 visit	654 (95.8)	659 (95.9)	1313 (95.8)
On-treatment completers ^a	578 (84.6)	637 (92.7)	1215 (88.7)
Off-treatment completers ^b	76 (11.1)	22 (3.2)	98 (7.2)
Non-completers ^c	29 (4.2)	28 (4.1)	57 (4.2)
Discontinued study medication	105 (15.4)	50 (7.3)	155 (11.3)
Analyzed for primary safety	681 (99.7)	685 (99.7)	1366 (99.7)
Analyzed for secondary safety			1491
Analyzed for primary efficacy	668 (97.8)	663 (96.5)	1331 (97.2)
Analyzed for secondary efficacy	680 (99.6)	682 (99.3)	1362 (99.4)

Note: Data are presented as n (%). Analysis populations are defined in Section 9.7.1.1.

^aSubjects were randomized, took IMP up to Month 12, and completed some or all of their required trial assessments to the end of the trial (including the Month 12 visit and at least 1 follow-up serum creatinine assessment).

^bSubjects were randomized, took IMP, but discontinued prior to Day 358 (or never began treatment), and completed some or all of their required assessments to the end of the trial (including the Month 12 visit and at least 1 follow-up serum creatinine assessment).

^cSubjects were randomized, took IMP (or never began treatment), but did not complete the Month 12 visit or complete Month 12 visit but did not have at least 1 follow-up serum creatinine assessment.

Table 1 Patient disposition (Table 10.1-1 of Study Report)

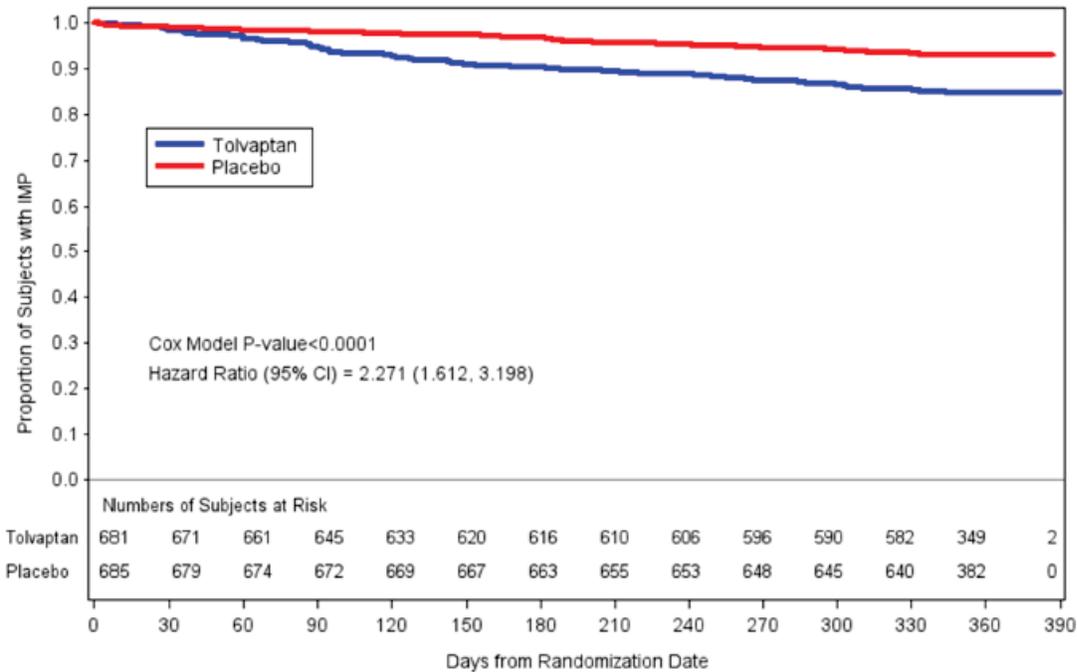


Figure 1 Kaplan-Meier plot of time to discontinuation from randomized drug (Figure 10.1-1 of Study Report).

The patient demographic characteristics are shown in Tables 2 and 3. The demographics were comparable between the two groups.

Variable		Tolvaptan (N=683)	Placebo (N=687)	Total (N=1370)	
Age [years]	Mean (SD)	47.3 (8.2)	47.2 (8.2)	47.3 (8.2)	
	Range	23, 65	21, 65	21, 65	
Gender	Male	n (%)	347 (50.8)	333 (48.5)	680 (49.6)
	Female	n (%)	336 (49.2)	354 (51.5)	690 (50.4)
Height [cm]	Mean (SD)	173.7 (10.4)	172.8 (10.2)	173.3 (10.3)	
	Range	134.0, 206.0	145.0, 205.0	134.0, 206.0	
Weight [kg]	Mean (SD)	84.6 (19.9)	83.2 (19.3)	83.9 (19.6)	
	Range	46.0, 227.0	41.9, 193.6	41.9, 227.0	
Body Mass Index (kg/m ²)	Mean (SD)	28.0 (5.8)	27.7 (5.6)	27.8 (5.7)	
	Range	17.1, 78.5	16.5, 52.0	16.5, 78.5	
Race	White	n (%)	626 (91.7)	632 (92.0)	1258 (91.8)
	Black or African American	n (%)	25 (3.7)	23 (3.3)	48 (3.5)
	American Indian or Alaska Native	n (%)	3 (0.4)	1 (0.1)	4 (0.3)
	Asian	n (%)	22 (3.2)	19 (2.8)	41 (3.0)
	Native Hawaiian or Other Pacific Islander	n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	n (%)	7 (1.0)	12 (1.7)	19 (1.4)

Table 2 Patient demographic characteristics (Table 11.2-1 of Study Report)

CKD Stage				
CKD 2	n (%)	32 (4.7)	39 (5.7)	71 (5.2)
CKD 3a	n (%)	209 (30.6)	202 (29.4)	411 (30.0)
CKD 3b	n (%)	303 (44.4)	315 (45.9)	618 (45.1)
CKD 4	n (%)	139 (20.4)	128 (18.6)	267 (19.5)
UNKNOWN	n (%)	0 (0.0)	3 (0.4)	3 (0.2)
Age				
<= 55 years old	n (%)	585 (85.7)	588 (85.6)	1173 (85.6)
> 55 years old	n (%)	98 (14.3)	99 (14.4)	197 (14.4)
eGFR (CKD-EPI)				
<= 45 mL/min/1.73 m ²	n (%)	442 (64.7)	438 (63.8)	880 (64.2)
> 45 mL/min/1.73 m ²	n (%)	241 (35.3)	249 (36.2)	490 (35.8)
Total kidney volume				
<= 2000 mL	n (%)	76 (11.1)	73 (10.6)	149 (10.9)
> 2000 mL	n (%)	60 (8.8)	60 (8.7)	120 (8.8)
Not known	n (%)	547 (80.1)	554 (80.6)	1101 (80.4)

SD = standard deviation.

Note: CKD stages were defined as follows: Stage 1 = ≥ 90 mL/min/1.73 m²;
Stage 2 = 60-89 mL/min/1.73 m²; Stage 3a = 45-59 mL/min/1.73 m²;
Stage 3b = 30-44 mL/min/1.73 m²; Stage 4 = 15-29 mL/min/1.73 m². The 3 subjects with CKD stage of "unknown" were randomized based on their eGFR level (rate blank method).

Table 3 Patient demographic characteristics (Table 11..2-1 of Study Report)

1.4.4 Results and Conclusions

There was a statistically significant difference in the mean annualized rate of change. The estimated difference was 1.27 mL/min per 1.73 m² per year; p<0.0001 and 95% CI (.858, 1.68). The sponsor's results from SAS are shown in Table 4.

STAT-1
SAS Output of Analysis of Weighted ANCOVA of Annualized Change in eGFR (CKD-EPI) from Pre-treatment Baseline to Post-treatment
Follow-up
Primary Endpoint Efficacy Population

The GLM Procedure
Dependent Variable: AVAL Analysis Value

Parameter	Estimate	Standard Error	t Value	Pr > t
TRT03PN 1	-4.005819600 B	0.60029336	-6.67	<.0001
TRT03PN 2	-5.276987717 B	0.60906165	-8.66	<.0001
BASE	0.011499275	0.01665210	0.69	0.4900
STRATAGE Age is greater than 55 years	1.854211358 B	0.31497217	5.89	<.0001
STRATAGE Age is less than or equal to 55 years	0.000000000 B			
STRATTKV TKV IS GREATER THAN 2,000 ML	-0.225111994 B	0.37241462	-0.60	0.5456
STRATTKV TKV IS LESS THAN OR EQUAL TO 2,000 ML	-0.324123872 B	0.34434849	-0.94	0.3467
STRATTKV UNKNOWN	0.000000000 B			
STRATEGF EGFR IS GREATER THAN 45 ML/MIN/1.73M ²	0.902339745 B	0.38267750	2.36	0.0185
STRATEGF EGFR IS LESS THAN OR EQUAL TO 45 ML/MIN/1.73M ²	0.000000000 B			

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

TRT03PN = 1 (tolvaptan), 2 (placebo) base: baseline eGFR

Table 4 SAS GLM results for primary endpoint (p. 56 Section 16.1.9 of Study Report and confirmed by the FDA)

For the 1331 subjects used in this analysis, the mean baseline eGFR was 41.024, the proportion of subjects greater than 55 years old is 0.14275 (i.e., 14%), the proportion of subjects with TKV >2000 mL is 0.088655 (9%), the proportion of subjects with TKV less than or equal to 2000 mL is 0.10594 (11%), and the proportion of patients with eGFR > 45 mL/min per 1.73 m² is 0.35762 (36%). So, we can calculate the estimated mean annualized change in eGFR from the parameter estimates in Table 4 as follows:

$$\text{Tolvaptan (TRT03PN = 1)} \\ -4.0058 + 0.011499 * 41.024 + 1.8542 * 0.14275 + (-0.22511) * 0.088655 + \\ (-0.32412) * 0.10594 + 0.90234 * 0.35763 = -3.0010$$

$$\text{Placebo (TRT03PN = 2)} \\ -5.2770 + 0.011499 * 41.024 + 1.8542 * 0.14275 + (-0.22511) * 0.088655 + \\ (-0.32412) * 0.10594 + 0.90234 * 0.35763 = -4.2722$$

In this model, the response variable is eGFR slope. This model shows how the eGFR slope varies with different baseline values of eGFR and different ages. In order to try to predict how eGFR would continue to change for a hypothetical patient in the future, both with and without tolvaptan treatment, I changed the model slightly. First, I removed the covariate for TKV stratum because most subjects had unknown TKV and the covariate was not significant in the model anyway. I also made age a continuous variable and I added an interaction term between age and eGFR and a treatment by baseline eGFR interaction. trt is a binary variable (0=placebo, 1=tolvaptan), base is the baseline eGFR. From the results in Table 5, we can see that there is a strong treatment by baseline interaction. The higher the baseline eGFR, the higher the estimated treatment effect. For a subject with baseline eGFR of 41.024 (the mean baseline value), the estimated treatment effect is $-1.082531 + 0.057708 * 41.024 \approx 1.28$, which is very close to the estimated treatment effect in the primary analysis.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-15.413311	2.536601	-6.076	1.60e-09	***
trt	-1.082531	0.802099	-1.350	0.17737	
base	0.156889	0.060629	2.588	0.00977	**
age	0.222030	0.052595	4.221	2.59e-05	***
trt: base	0.057708	0.018887	3.055	0.00229	**
base: age	-0.003020	0.001286	-2.348	0.01902	*

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1

Residual standard error: 0.9606 on 1325 degrees of freedom
 Multiple R-squared: 0.08288, Adjusted R-squared: 0.07942
 F-statistic: 23.95 on 5 and 1325 DF, p-value: < 2.2e-16

Table 5 Linear Model results for modified model in Study 156-13-210

Next, I fit a similar model with the data from the first study (TEMPO 3:4, also called Study 156-04-251). In the first study, subjects had to have estimated creatinine clearance greater of at least 60 mL/min and they had to be less than 50 years old. So, the first study generally had subjects in the early stages of CKD (measured by kidney function) and the second study generally had patients in the later stages of CKD. If we use the model from the first study- Table 6 – and apply it to estimate the treatment effect for a patient with eGFR of 41.024, then we would find $2.176 - 0.007211 * 41.024 = 1.88$. This is quite different from the estimate we found from the model fit to the data from the second study.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.604e+01	1.673e+00	-9.590	< 2e-16	***
trt	2.176e+00	6.022e-01	3.613	0.000313	***
base	1.180e-01	1.742e-02	6.774	1.82e-11	***
age	1.591e-01	4.105e-02	3.876	0.000111	***
trt: base	-7.211e-03	7.109e-03	-1.014	0.310588	
base: age	-1.359e-03	4.431e-04	-3.066	0.002210	**

 Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.803 on 1429 degrees of freedom
 (10 observations deleted due to missingness)

Multiple R-squared: 0.2147, Adjusted R-squared: 0.2119

F-statistic: 78.13 on 5 and 1429 DF, p-value: < 2.2e-16

Table 6 Linear Model results for modified model in Study 156-04-251

I next combined the two models to draw prediction curves for future values of eGFR given several different starting age and eGFR. See the appendix for more details about how this was done. I used the model from Table 5 when the eGFR was less than 50 mL/min per 1.73 m² and the model from Table 6 if eGFR was greater than 50 mL/min per 1.73 m². Figure 2 shows the predicted eGFR over time for two hypothetical patients under 4 different scenarios (no tolvaptan, tolvaptan started immediately, after 5 years, or after 10 years). The black curves represent the predicted eGFR over time for two different hypothetical patients- one with eGFR=100 mL/min per 1.73 m² at age 20 and the second with the same eGFR at age 30. The red curves show the predicted eGFR over time if those patients start taking tolvaptan immediately. The blue curves show the predicted eGFR if they start taking tolvaptan 5 years later, the orange curves, if they start 10 years later. Again, we are using the two separate models from the two studies depending on whether eGFR is greater than or less than 50 mL/min per 1.73 m². That's why there is an abrupt change in the slope at that point. We could combine the data from the two studies together and force everything into one model that is smooth, but I don't think we should do that. The studies had different enrollment criteria and that may account for the difference. Also, I do not think we should take one of the models and use it to extrapolate outside the range of age and eGFR that was observed in the corresponding trial. For a hypothetical 20 year old with eGFR = 100 mL/min per 1.73 m² with similar characteristics to the subjects in the TEMPO 3:4 trial, the models

predict there is a good chance that patient would reach $eGFR = 15 \text{ mL/min per } 1.73 \text{ m}^2$ at around age 35. If this patient were to start tolvaptan immediately at age 20, this could be delayed by about 6 years. If the patient started tolvaptan at age 25, it could be delayed by 3 years. If started at age 30, it could be delayed by about 1 year.

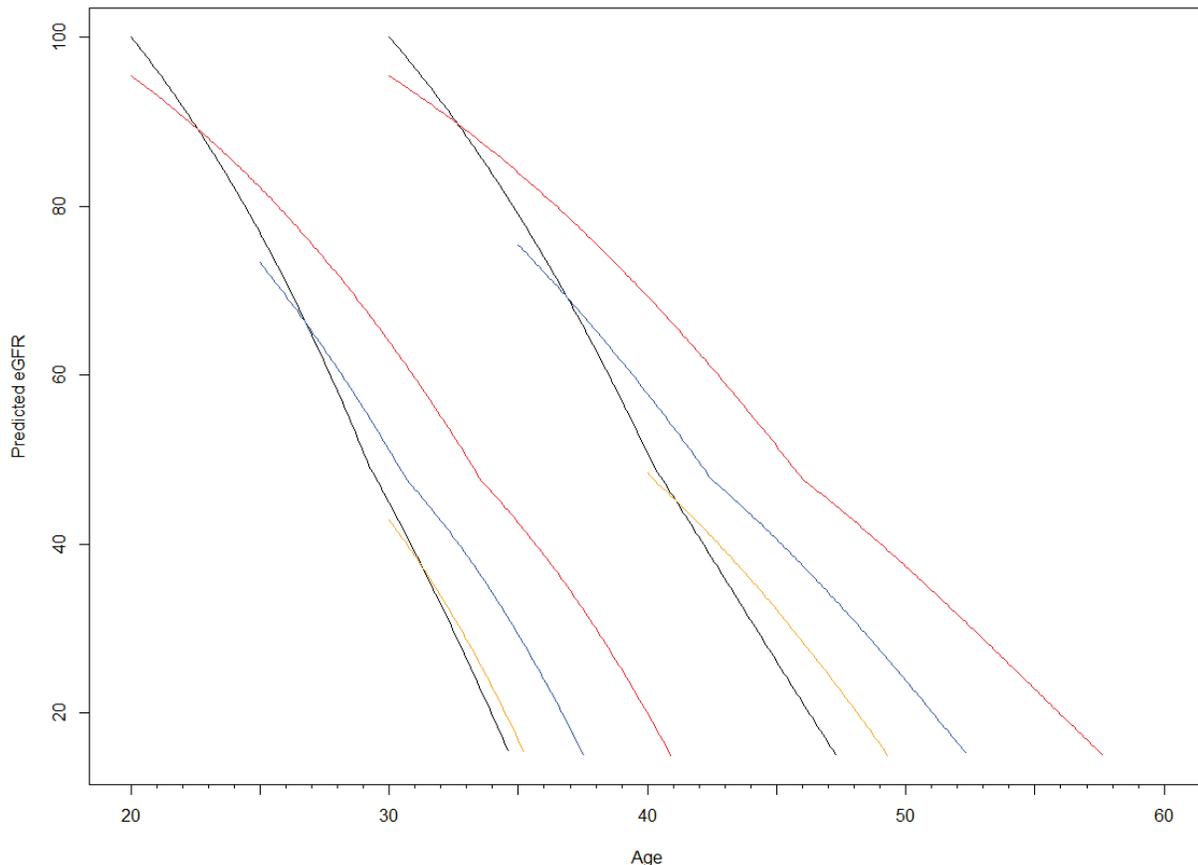


Figure 2 Predicted eGFR over time for two hypothetical patients. Black=no tolvaptan treatment, Red=tolvaptan immediately, Blue=tolvaptan 5 years after, Orange=tolvaptan 10 years after (FDA analysis).

From the TEMPO 3:4 trial, there was some suggestion that tolvaptan slowed the rate of growth of TKV. In the statistical review of that trial, it was argued that there was a large acute hemodynamic effect, but little to no long term effect on TKV. Because of the amount of missing data, it was not possible to conclude whether there was any long term effect on TKV. We now have data from the open-label extension phase of that trial (TEMPO 4:4) that demonstrate clearly that there is little to no long term sustained disease modifying effect on TKV. If we use a simple model for change in $\log(\text{TKV})$ with only an acute effect and a linear growth, there is no treatment effect. If we make the model more complicated to allow a quadratic term, there is a small effect on TKV. Even when using the more generous assumption including the quadratic term, the correlation between chronic TKV slope and chronic eGFR slope is very low ($R^2 \approx 5\%$)

and the proportion of treatment effect on eGFR that is explained by the effect on TKV is very low (~17%). See the appendix for further details. This new data and these new analyses clearly show that the FDA was correct in discouraging the sponsor from using TKV as the primary endpoint in TEMPO 3:4 and the Advisory Committee was correct in voting not to approve tolvaptan based solely on the data from the TEMPO 3:4 trial.

1.5 Evaluation of Safety

See clinical review.

1.6 Benefit-Risk Assessment (Optional)

See clinical review.

FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

1.7 Gender, Race, Age, and Geographic Region

The sponsor's results for the primary endpoint are shown in Figure 3. Tolvaptan appeared to have a larger effect in patients less than or equal to 55 years old compared to the older subjects. This is a statistically significant quantitative interaction in the effect between the two age groups.

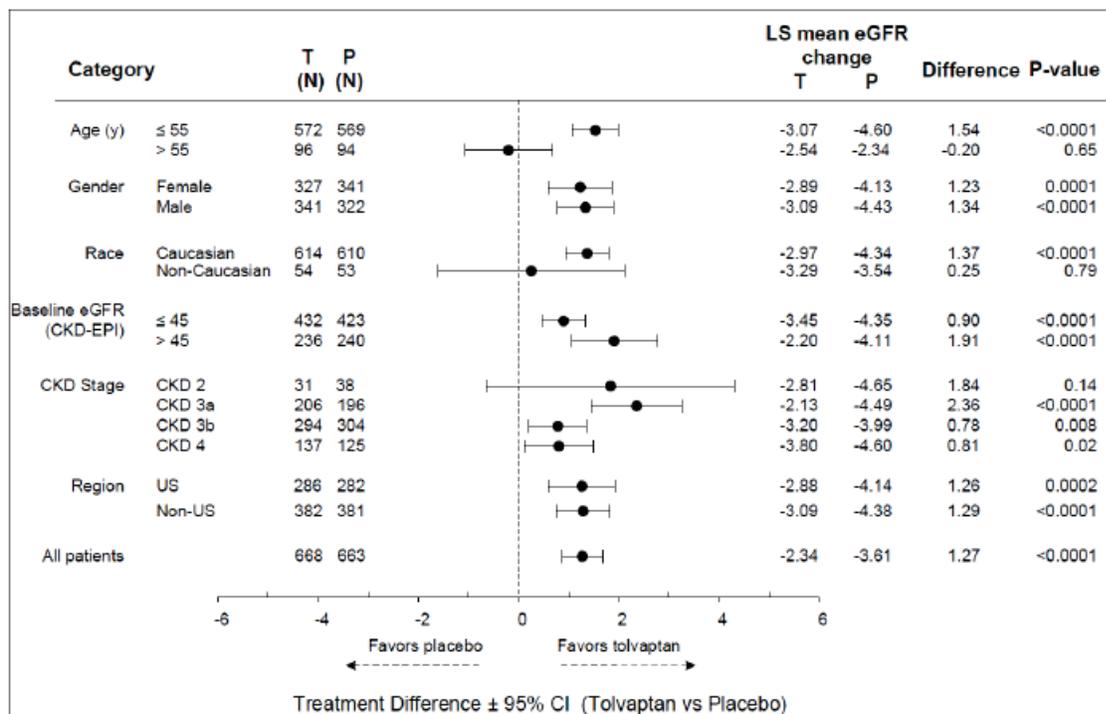


Figure 3 Sponsor's results of key secondary endpoint within subgroups (Study Report Figure 11.4.1.1-3)

1.8 Other Special/Subgroup Populations

NA

SUMMARY AND CONCLUSIONS

1.9 Statistical Issues

Models appropriately included acute and hemodynamic effects. The sponsor did a commendable job through the design of the study at minimizing the amount of missing data. Sensitivity analyses were done to assess impact of missing data. There was a small amount of missing data and the relatively large treatment effect.

1.10 Collective Evidence

Based on the collective evidence of both studies, tolvaptan appears to slow the rate of decline of kidney function as measured by the estimated glomerular filtration rate (eGFR). (b) (4)

In the second trial, TKV was not measured, so there is no evidence from that trial. From the first study where all subjects had TKV measured annually, there appears to be a substantial acute hemodynamic effect, but little or no long term chronic effect on TKV. Depending on what model is used in the double-blind phase, there is either no chronic effect or a very small chronic effect. If there were even a small effect, it may or may not be a cosmetic or clinical benefit. When using the model that suggests there is a small chronic effect on TKV, there is a very small correlation between chronic effect on TKV and chronic effect on eGFR. The proportion of the treatment effect on eGFR explained by TKV is also very small in that trial. This suggests chronic effect on TKV is not a surrogate endpoint for effect on eGFR for this study population. In a long-term extension phase of that trial, all subjects were treated with tolvaptan to determine whether there was a sustained difference between the two arms. At the end of this extension phase, the subjects who were treated 3 years longer with tolvaptan in the double-blind phase had no significant difference from the subjects that had been randomized to placebo.

1.11 Conclusions and Recommendations

Two Phase 3 studies have been completed to assess the safety and efficacy of tolvaptan for the treatment of polycystic kidney disease (PKD). Both studies have been relatively short term (1-3 years) and have shown a modest effect on annualized change in eGFR. If approved, tolvaptan will be used chronically. It is natural to assume that the effect seen in one year (~1 mL/min) will compound each year, so that after 10 years there will be an effect of approximately 10 mL/min.

We don't know if that is true, but we do have evidence from the two trials that tolvaptan works in both early and late stage of disease. It is reasonable to expect some benefit to continue to accrue each year with continued treatment.

1.12 Labeling Recommendations (as applicable)

Labeling should make clear there is no effect demonstrated on chronic change in TKV. Effect seen on chronic change in eGFR may continue to accrue over longer periods of time than measured in the trial, but label should be cautious about using model to extrapolate benefits over long durations of time since we have no data to directly support that.

APPENDICES

Estimating eGFR over time given starting values

Method 1. Numerical method

Suppose we start with a model for how the slope of eGFR depends on the current age and eGFR. Given any starting age and eGFR (denoted Age_0 and $eGFR_0$), we can predict the eGFR at a short time in the future, for example 0.1 years later, by the current eGFR plus 0.1 times the estimated slope. If $eGFR > 50$, then use the slope from the model in Table 6. If $eGFR < 50$, then use the slope from the model in Table 5. Call this predicted eGFR at the next year $eGFR_1$. Also, the age 0.1 years later will be $Age_1 = Age_0 + 0.1$. Continue doing this to find a sequence of values $(Age_i, eGFR_i)$ where each pair in the sequence is found by taking the prior pair $(Age_{i-1}, eGFR_{i-1})$ and finding the predicted slope, then finding $eGFR_i = eGFR_{i-1} + 0.1 * \text{predicted slope}$ and defining $Age_i = Age_{i-1} + 0.1$.

Method 2. Solving a differential equation

Depending on the form of the equation defining the predicted slope of eGFR, the differential equation may be solvable to give us an explicit formula for the predicted eGFR at all future times given a starting age and eGFR (Age_0 and $eGFR_0$). Let $y(t)$ be the predicted eGFR at time t years from the starting time. The age at time t is $eGFR_0 + t$. Thus, the differential equation using our model for the slope is

$$y' = a_0 + a_1 y + a_2(Age_0 + t) + a_3(Age_0 + t)y$$

Where a_0, a_1, a_2 , and a_3 are constants with initial condition $y(0) = eGFR_0$.

The equation can be rewritten in the standard form for a nonhomogeneous first order differential equation as:

$$y' = \{a_1 + a_3(Age_0 + t)\}y + a_2(Age_0 + t) + a_0$$

The solution subject to the initial condition is

$$y(t) = \frac{\left((a_2 + a_3 eGFR_0) e^{a_1 t + a_3 Age_0 t + \frac{a_3 t^2}{2}} - a_2 \right)}{a_3} + \frac{\sqrt{2\pi}(a_1 a_2 - a_0 a_3)}{2a_3^{3/2}} e^{\frac{(a_1 + a_3(Age_0 + t))^2}{2a_3}} \left\{ erf\left(\frac{a_1 + a_3 Age_0}{\sqrt{2a_3}}\right) - erf\left(\frac{a_1 + a_3(Age_0 + t)}{\sqrt{2a_3}}\right) \right\}$$

where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function.

This only describes the solution when there is a single model for y' and the a_0, a_1, a_2 , and a_3 are constants. Since we have two different models depending on the eGFR, if $y(0) > 50$ then we have to solve two different nonhomogeneous first order linear differential equations with 2 different initial conditions.

Note: Method 2 is more elegant, but Method 1 is easier to understand and can always be applied regardless of the form of the equation that models eGFR slope as a function of age and eGFR.

TKV growth in TEMPO 3:4 and TEMPO 4:4

Table 7 shows the estimated effect on chronic TKV slope assuming a model including a quadratic effect of time. This model and the estimated coefficients are the same as on p.14 of the Statistical Review. In this model, there is a rather large acute hemodynamic effect (trt:posttrt interaction = -0.02235) and a relative small, but statistically significant, chronic effect (time:trt:posttrt interaction = -0.004334). The time variable is measured in units of years here.

Fixed effects: log(RERESULT.N, 10) ~ time + time^2 + trt:posttrt + trt:posttrt:time

	Value	Std. Error	DF	t-value	p-value
(Intercept)	3.181218	0.005120845	3541	621.2291	0.000
time	0.015834	0.001458221	3541	10.8583	0.000
time^2	0.002577	0.000379631	3541	6.7878	0.000
trt:posttrt	-0.022350	0.001794210	3541	-12.4569	0.000
time:trt:posttrt	-0.004334	0.001318831	3541	-3.2861	0.001

Table 7 Linear Model results for longitudinal log(TKV) in TEMPO 3:4 (Study 156-04-251) (FDA analysis).

Table 8 shows the estimated effect on chronic TKV slope assuming a model including only the linear effect of time. In this model, there is a rather large acute hemodynamic effect (trt:posttrt interaction = -0.031335) and a very small, and statistically not significant, estimated chronic effect (time:trt:posttrt interaction = -0.001358).

Fixed effects: log(RERESULT.N, 10) ~ time + trt:posttrt + trt:posttrt:time

	Value	Std. Error	DF	t-value	p-value
(Intercept)	3.180461	0.005120114	3542	621.1699	0.0000
time	0.023342	0.000949956	3542	24.5717	0.0000
trt:posttrt	-0.031335	0.001222074	3542	-25.6406	0.0000
time:trt:posttrt	-0.001358	0.001244014	3542	-1.0913	0.2752

Table 8 Linear Model results for longitudinal log(TKV) in TEMPO 3:4 (Study 156-04-251) (FDA analysis).

Recently, the investigators have published the results of the long-term open-label extension phase of the trial (Torres, Vicente E., et al. "Multicenter, open-label, extension trial to evaluate the long-term efficacy and safety of early versus delayed treatment with tolvaptan in autosomal dominant polycystic kidney disease: the TEMPO 4: 4 Trial." *Nephrology Dialysis Transplantation* (2017): gfx043.). More than 90% of the subjects outside Japan that completed the trial enrolled in TEMPO 4:4 (871 enrolled out of 948 completers outside Japan). All subjects were started on tolvaptan and followed for two years to determine whether there was any difference between the arm that had originally been treated with tolvaptan compared to the arm that had originally been treated with placebo (for the 3 years in the initial double-blind study). 763 out of 871 subjects enrolled in TEMPO 4:4 completed 24 months of follow-up. No difference was observed between the two arms at the end of followup.

This is shown in Figure 4. The figure shows that there is a large acute effect at the beginning of TEMPO 3:4. There is little to no chronic effect on TKV after that. During the break in between the two phases, the Early Tolvaptan treated arm tended to get closer to the Delayed Tolvaptan arm because the hemodynamic effect was disappearing. During the first year of TEMPO 4:4, the Delayed Tolvaptan arm nearly matched the Early Tolvaptan arm and there was no difference by the end of the second year.

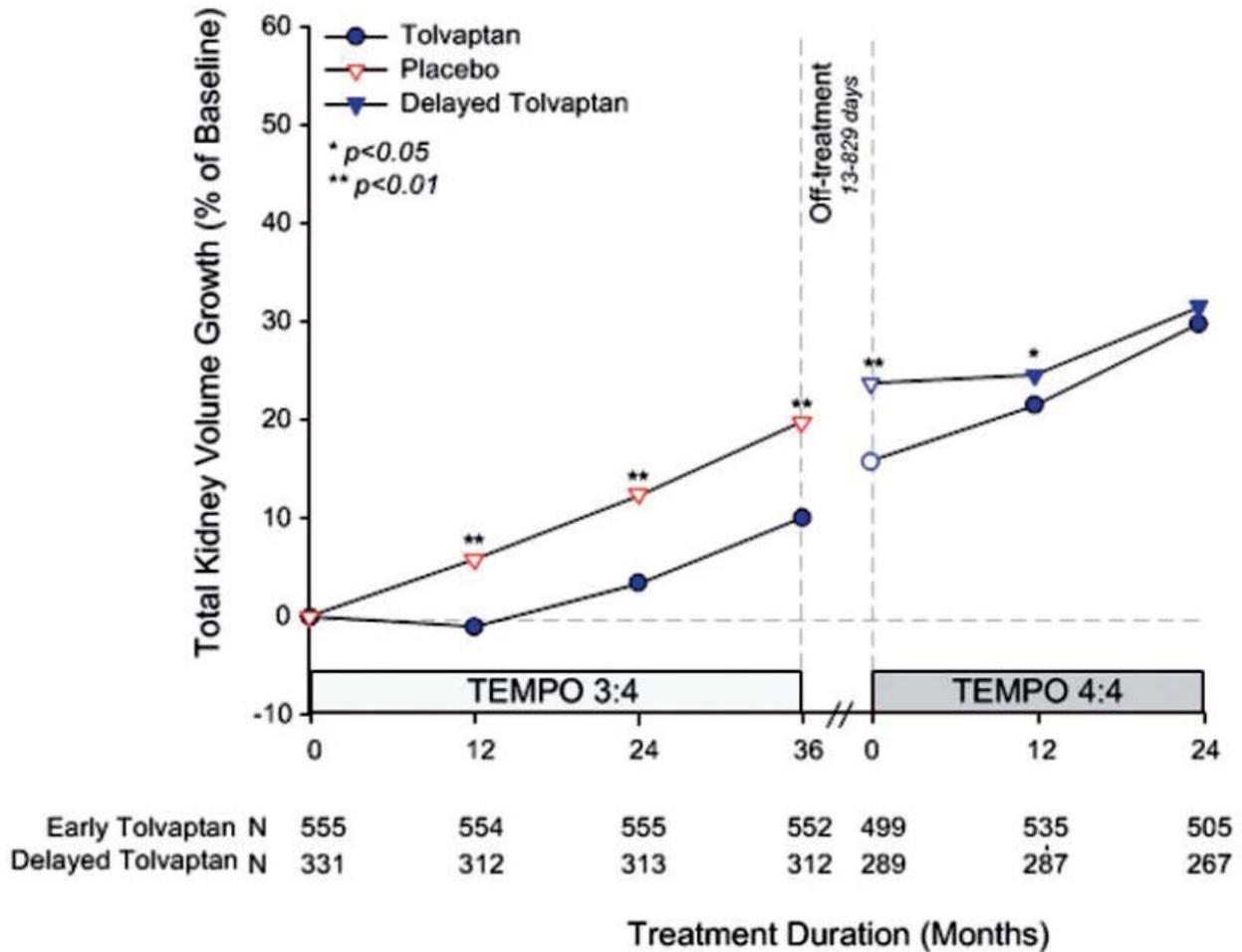


Figure 4 Percentage change in TKV from TEMPO 3:4 baseline to TEMPO 4:4 Month 24 visit. Open circles and triangles represent off-treatment time points. Break in treatment from TEMPO 3:4 to TEMPO 4:4 ranged from 13 to 829 days from the Month 36 visit in TEMPO 3:4. (Figure 2 of Torres, Vicente E., et al.).

TKV was not measured in the new trial (Study 156-13-210). The results for the effect on TKV from the earlier trial were inconclusive before, but in light of the new data from the TEMPO 4:4

extension phase, the results are getting more clear. There is no long term effect on TKV or the effect is very small.

Correlation between chronic TKV slope and chronic eGFR slope in TEMPO 3:4

Table 9 shows the estimated effect on chronic eGFR slope. The covariate beGFR is baseline eGFR and btkv is baseline TKV. This model and the estimated coefficients are the same as on p. 21 of the Statistical Review. For each subject, we have an estimated chronic slope of TKV (from the model in Table 7) and a chronic slope of eGFR from the model in Table 9. The estimated correlation between these two across all subjects is -0.31 with an approximate 95% CI of (-0.35, -0.26). This is a small statistically significant correlation between the two variables. The estimated correlation goes in the direction one would expect based on the hypothesis that slowing the rate of growth of TKV causes a higher rate of change (less negative) in eGFR.

Fixed effects: lgfr ~ time + log(beGFR) + trt: time + postrand: trt + posttrt: trt + AGE + log(btkv) + log(btkv): time + log(beGFR): time

	Value	Std. Error	DF	t-value	p-value
(Intercept)	0.8516663	0.08899565	13317	9.56975	0
time	-0.3578122	0.04613133	13317	-7.75638	0
log(beGFR)	0.8836611	0.01159501	1430	76.21049	0
AGE	-0.0017195	0.00040382	1430	-4.25811	0
log(btkv)	-0.0382131	0.00664043	1430	-5.75461	0
time: trt	0.0204063	0.00320151	13317	6.37395	0
trt: postrand	-0.0457872	0.00567248	1430	-8.07180	0
trt: posttrt	0.0415260	0.00279014	13317	14.88314	0
time: log(btkv)	-0.0205037	0.00378917	13317	-5.41114	0
time: log(beGFR)	0.1016358	0.00608802	13317	16.69440	0

Table 9 Linear Model results for longitudinal log(eGFR) in Study 156-04-251 (FDA analysis).

Proportion of treatment effect on chronic eGFR slope explained by chronic TKV slope in TEMPO 3:4

Table 10 shows the estimated effect on chronic eGFR slope after adjusting for the effect on chronic TKV slope. The estimated proportion of the treatment effect on eGFR explained by the effect on TKV is $(0.0204 - 0.0168) / 0.0204 \approx 18\%$. If we used the model from Table 8, it would not make any sense to even consider the proportion of treatment effect explained by the effect on TKV because there was not any effect shown using that model. So, in the best case scenario, there is a small chronic effect on TKV, there is a weak correlation between the effect on TKV and the chronic effect on eGFR ($r \approx -0.31$), and the proportion of treatment effect explained by TKV is 18%.

Fixed effects: lgfr ~ time + log(beGFR) + trt:time + postrand:trt + posttrt:trt + AGE + log(btkv) + log(btkv):time + log(beGFR):time + time:estTKVsl ope

	Value	Std. Error	DF	t-value	p-value
(Intercept)	0.8675077	0.08894230	13316	9.75360	0e+00
time	-0.3908233	0.04490199	13316	-8.70392	0e+00
log(beGFR)	0.8816709	0.01158690	1430	76.09206	0e+00
AGE	-0.0018973	0.00040500	1430	-4.68466	0e+00
log(btkv)	-0.0382486	0.00663162	1430	-5.76761	0e+00
time:trt	0.0168140	0.00313442	13316	5.36431	0e+00
trt:poststrand	-0.0456990	0.00566529	1430	-8.06648	0e+00
trt:posttrt	0.0414439	0.00278980	13316	14.85549	0e+00
time:log(btkv)	-0.0136379	0.00375979	13316	-3.62732	3e-04
time:beGFR	0.1009707	0.00590511	13316	17.09887	0e+00
time:estTKVsl ope	-0.7805466	0.09059011	13316	-8.61624	0e+00

Table 10 Linear Model results for longitudinal log(eGFR) adjusting for chronic TKV slope in Study 156-04-251 (FDA analysis).

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

JOHN P LAWRENCE
02/27/2018

HSIEN MING J HUNG
02/27/2018



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #: NDA 204-441
Supplement #:
Drug Name: Tolvaptan
Indication(s): Slow progressive kidney disease in adults with autosomal dominant polycystic kidney disease
Applicant: Otsuka
Date(s): 11/15/2012
Review Priority: Priority
Biometrics Division: DBI
Statistical Reviewer: John Lawrence, Ph D
Concurring Reviewers: Jim Hung
Medical Division: Cardiorenal.
Clinical Team: Aliza Thompson MD, Nhi Beasley MD, Steven Grant, MD
Project Manager: Anna Park
Keywords:
survival analysis, benefit-risk, mixed models, longitudinal data analysis

Table of Contents

EXECUTIVE SUMMARY	5
INTRODUCTION	6
1.1 OVERVIEW.....	6
1.2 DATA SOURCES	7
STATISTICAL EVALUATION	7
1.3 DATA AND ANALYSIS QUALITY	7
1.4 EVALUATION OF EFFICACY	8
1.4.1 <i>Study Design and Endpoints</i>	8
1.4.2 <i>Statistical Methodologies</i>	8
1.4.3 <i>Patient Disposition, Demographic and Baseline Characteristics</i>	9
1.4.4 <i>Results and Conclusions</i>	13
1.5 EVALUATION OF SAFETY	23
1.6 BENEFIT-RISK ASSESSMENT (OPTIONAL).....	23
FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	23
1.7 GENDER, RACE, AGE, AND GEOGRAPHIC REGION	23
1.8 OTHER SPECIAL/SUBGROUP POPULATIONS	24
SUMMARY AND CONCLUSIONS	24
1.9 STATISTICAL ISSUES	24
1.10 COLLECTIVE EVIDENCE.....	24
1.11 CONCLUSIONS AND RECOMMENDATIONS	25
1.12 LABELING RECOMMENDATIONS (AS APPLICABLE).....	25
APPENDICES.....	26
DISTRIBUTION OF RESIDUALS FROM FDA MODEL OF EGFR	26
DISTRIBUTION OF THE RANDOM EFFECTS	27
MODEL FOR EGFR BETWEEN 30 AND 45 mL/MIN/1.73 M ²	27
EXAMPLES	28
USING THE MODEL TO PREDICT INTO THE FUTURE	30

LIST OF TABLES

Table 1 Patient disposition (Table 8.1-1 of Study Report)	9
Table 2 Distribution of number of eGFR measurements per subject used in sponsor's analysis.....	11
Table 3 Distribution of number of eGFR measurements per subject actually measured.....	12
Table 4 Patient demographic characteristics (Table 8.2-1 of Study Report)	12
Table 5 Patient demographic characteristics (Table and 8.2-3 of Study Report).....	13

LIST OF FIGURES

Figure 1 Kaplan-Meier plot of time to discontinuation for all reasons (Figure 8.1-1 of Study Report).	10
Figure 2 Kaplan-Meier plot of time to last eGFR measurement used in sponsor's analysis (Source: FDA)	11
Figure 3 Scatterplot of residuals versus predicted log(TKV) from sponsor's model.	14
Figure 4 Mean observed and predicted TKV over time.....	15
Figure 5 Estimated density of change in eGFR during initial 3 week titration period.....	18
Figure 6 Residuals versus predicted scatterplot from sponsor's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 14 residuals with magnitude larger than 30 not shown.	20
Figure 7 Normal probability plot for residuals from sponsor's eGFR model.....	20
Figure 8 Residuals versus predicted scatterplot from FDA's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 19 residuals with magnitude greater than 0.5 not shown.	22
Figure 9 Normal probability plot for residuals from FDA's log-eGFR model.....	22
Figure 10 Sponsor's results of key secondary endpoint within subgroups (Study Report Figure 9.4.2-1)	23

EXECUTIVE SUMMARY

This review replaces the statistical review dated June 25, 2013.

This submission contains one Phase 3 study to support the indication. Accordingly, the level of evidence from that trial must be equivalent to two trials with a type I error rate of 0.05 each. According to the medical division, the primary endpoint of total kidney volume is not acceptable for approval. This review focuses on the sponsor's key secondary endpoint, a composite endpoint consisting of events defined by hypertension, renal function, renal pain, and albuminuria. In addition, this review focuses on exploratory analyses of longitudinal changes in estimated kidney function (glomerular filtration rate estimated by the CKD-EPI formula).

There were several statistical issues with the analyses. There was possibly non-ignorable missing data and substantially more missing data in the tolvaptan arm compared to the placebo arm. In some analyses, the ITT population could not be used because there were no valid observations. In addition, assumptions used in the models were clearly violated (assumptions about linear responses over time and assumptions about homogeneous variance of residual errors). Tolvaptan has substantial acute effects on estimated GFR and on total kidney volume that are different than chronic effects. Therefore, simple models do not adequately fit the data.

INTRODUCTION

1.1 Overview

Table: List of all studies included in analysis

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
<i>Study 156-04-251</i>	<i>Phase 3</i>	<i>36 months</i>	<i>36 months</i>	<i>tolvaptan: 961 placebo:484</i>	<i>subjects with ADPKD as defined by a certain number of cysts, estimated creatinine clearance of at least 60 mL/min and TKV>750 mL.</i>

There was one Phase 3 trial conducted to support this indication. A Special Protocol Assessment was done, but the FDA did not agree with the Protocol. Since there was only one study, a type 1 error rate of 0.01 was to be used for approval decisions. This was communicated to the sponsor. The primary endpoint of TKV was never acceptable to the FDA, but the key secondary endpoint was an acceptable endpoint.

The meeting minutes from a face to face meeting between the FDA and the sponsor on June 10, 2009 state:

"2) We propose that a significance level of 0.0491 (two-sided) will be used to declare statistical significance at the final analysis for the primary endpoint. In addition, we propose that a significance level of 0.05 (two-sided) will be used to declare statistical significance at the final analysis for the key secondary composite endpoint. In a Type A meeting with the Division on 15 Nov 2005 (minutes provided as [Attachment 2](#)), Otsuka proposed, "if the primary endpoint and composite key secondary endpoint are both statistically significant, and if the other specified endpoints are supportive, the data from this single phase 3 trial will be sufficient to support a New Drug Application (NDA) approval for the proposed indication." The Division agreed to Otsuka's proposal. **Does the FDA agree that the significance levels specified in the draft SAP are acceptable for approval based on a single pivotal trial?**

Preliminary FDA Response: A p-value < 0.05 from a single trial is acceptable for your primary efficacy endpoint because we do not consider this endpoint a surrogate of benefit. In order to provide convincing evidence of treatment benefit, the composite key secondary endpoint will need a p-value < 0.01.

Additional discussion during the meeting: The sponsor has decided to continue their study as proposed and is aware the Division will likely review the results in a more stringent fashion. Dr. Stockbridge reiterated that the Division was less interested in the primary endpoint as compared to the secondary endpoints. The Division acknowledged the sponsor's decision."

However, the FDA defines a primary endpoint in its guidance document as "Endpoint(s) necessary and/or sufficient to establish efficacy" (not published as of this date, but that definition appears in the slide presentation here:

http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2013/03/WC50

0140627.pdf). Since TKV was not necessary or sufficient to establish efficacy, then by the FDA guidance document's definition, it was not a primary endpoint. Even if you don't rely on the definition from the guidance document (which is fair since it is not even published now), it is clear from the minutes above that the FDA told the sponsor that TKV could not be the primary endpoint of the trial. Despite multiple attempts to explain to the sponsor that TKV was not a primary endpoint, the company insisted on calling it the primary endpoint and the FDA was powerless to stop them. In these same meeting minutes, they discuss a plan to stop the trial early at an interim analysis if a benefit was shown on TKV (this adjustment for the interim analysis is the reason for the significance level of 0.0491). This illustrates the difference between how much importance the FDA put on TKV compared to the how much the company put; the company intended to stop the trial early and claim victory if a benefit was shown on TKV while the FDA was telling them they had no interest in TKV.

1.2 Data Sources

Electronic datasets and Study Reports:

<\\cdsesub1\evsprod\NDA204441\204441.enx>

<\\cdsesub1\evsprod\NDA204441\0001\m5\datasets\156-04-251\analysis>

STATISTICAL EVALUATION

1.3 Data and Analysis Quality

The data quality and analysis quality were both poor.

Many (several thousand) serum creatinine measurements were not included in the sponsor's analysis. There were many subjects that were not included in the sponsor's analysis at all. Other subjects had partial data. Subjects with some missing data is common in clinical trials, but the amount of missing or unreliable data in this trial is uncommon (compared to other trials of cardiovascular or renal disease). In many cases, subjects were not followed at all, or only for a short time if they stopped treatment early. A true intent-to-treat analysis should follow all subjects for all outcomes for the entire planned period (36 months). This was not done here. Baseline for changes in serum creatinine or eGFR was defined as the measurement after titration. This caused many subjects to be excluded from the analysis completely if they could not tolerate the drug during the titration phase. It is very uncommon to define a baseline value so long after randomization (approximately 3 weeks). If all the subjects are still in the trial at that time, there is less of a concern, but that was not the case here.

The sponsor's analysis used assumptions that in some cases can be demonstrated to be false and in other cases could not be verified. The mixed effects models include an assumption that the

residual error variance is homogeneous and that those errors are normally distributed. For the TKV endpoint (after log transformation) and for the eGFR endpoint, both of these assumptions can be shown false using the data. In addition, the sponsor's analyses used simple linear response models. For both endpoints, those models were not adequate and that can be shown with the data. Furthermore, these models use other assumptions about the distribution of random effects and the nature of missing data (missing at random) that cannot be verified. Lastly, the analysis of recurrent events uses assumptions in the estimate of the variance that may exaggerate the significance of the p-value for that analysis (see Section 1.4.4).

1.4 Evaluation of Efficacy

1.4.1 Study Design and Endpoints

Study 156-04-251 was a multinational, multicenter trial. 1445 subjects were randomized 2:1 to tolvaptan or placebo. The primary endpoint was change in TKV (total kidney volume) over time. TKV was measured at baseline and every 12 months up to month 36 by MRI. The key secondary endpoint was a composite of clinically relevant outcomes. The composite consisted of four types of events: hypertensive progression (change in category or addition of hypertension medication); renal pain; worsening of albuminuria; worsening of renal function (confirmed rise of 33% in serum creatinine). The composite endpoint was counted with recurrence possible, i.e. not just the first event for each subject, but rather multiple events for each subject were possible and all were counted. Change in renal function (inverse of serum creatinine and other estimates of creatinine clearance or GFR) were also secondary or exploratory endpoints.

1.4.2 Statistical Methodologies

The primary endpoint, TKV, was analyzed using a mixed effects model. First, the TKV was transformed using the base 10 logarithm. Time was measured in years from the time of the first (baseline) TKV (number of days divided by 365.25) and was included as a continuous variable in the model.

The following linear mixed-effect model was fitted to the log-transformed TKV repeated-measures data:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 \text{Group}_i + \beta_4 t_{ij} \times \text{Group}_i + b_{1i} + b_{2i} t_{ij} + e_{ij},$$

In this model, Y_{ij} is the \log_{10} (TKV) of subject i at visit j ($j = 0, 1, 2, 3$), where $\text{Group}_i = 0$ for a subject in the placebo group and $\text{Group}_i = 1$ for a subject in the tolvaptan group. β_1 , β_2 , β_3 , and β_4 are fixed effects (β_1 is the intercept of placebo, $\beta_1 + \beta_3$ is the intercept of tolvaptan, β_2 is the slope of placebo, and $\beta_2 + \beta_4$ is the slope of tolvaptan), while b_{1i} and b_{2i} are random effects assumed to be normally distributed with mean 0 and unknown

variance covariance structure. The error terms in the model, e_{ij} , are assumed mutually independent and normally distributed as $N(0, \sigma^2)$, and they are also assumed to be independent of the random effects. The primary null hypothesis is $H_0: \beta_4 = 0$ versus the alternative hypothesis $H_1: \beta_4 \neq 0$.

The key secondary endpoint was analyzed using the Anderson-Gill recurrent events model. No covariates were included other than treatment group. Subjects were censored at the last censoring time for all components and were considered to have no events of the type without follow-up at those times where unknown.

1.4.3 Patient Disposition, Demographic and Baseline Characteristics

The patient disposition are shown in Table 1. Significantly more subjects discontinued in the tolvaptan arm. The bottom row shows some of the subjects who discontinued were followed for some PKD outcomes, but that means a phone call in many cases and is not the same as complete follow-up on all outcomes. Figure 1 shows that the proportion of subjects in the tolvaptan arm who discontinued was larger than the proportion in the placebo arm uniformly throughout the trial.

Number of Subjects	Tolvaptan (N = 961) n (%)	Placebo (N = 484) n (%)	Total (N = 1445) n (%)
Screened	-	-	2122
Randomized	961 (100.0)	484 (100.0)	1445 (100.0)
Treated	961 (100.0)	483 (99.8)	1444 (99.9)
Completed	740 (77.0) ^a	417 (86.2) ^b	1157 (80.1)
Discontinued IMP	221 (23.0)	67 (13.8)	288 (19.9)
Lost to follow-up	15 (1.6)	8 (1.7)	23 (1.6)
AE	148 (15.4)	24 (5.0)	172 (11.9)
Subject met withdrawal criteria	4 (0.4) ^c	0 (0.0)	4 (0.3)
Investigator withdrew subject	3 (0.3)	4 (0.8)	7 (0.5)
Subject withdrew consent	50 (5.2)	30 (6.2)	80 (5.5)
Protocol deviation	1 (0.1) ^d	1 (0.2) ^d	2 (0.1)
Discontinued and followed for PKD Outcomes	102 (10.6)	27 (5.6)	129 (8.9)

Table 1 Patient disposition (Table 8.1-1 of Study Report)

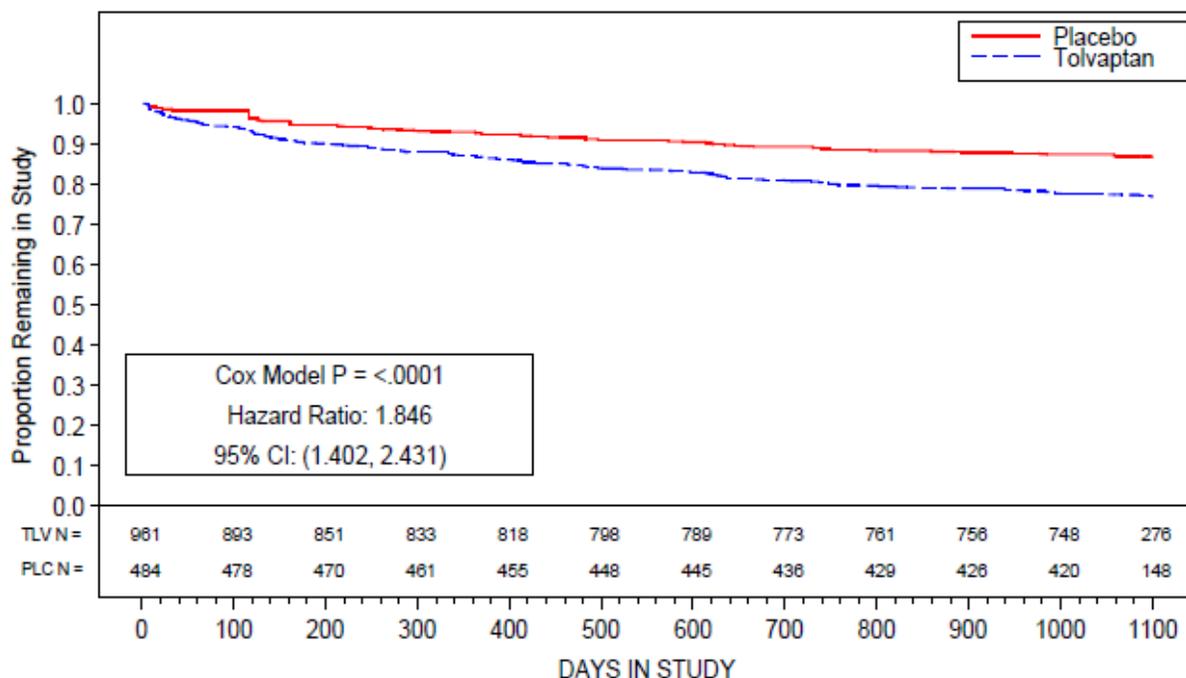


Figure 1 Kaplan-Meier plot of time to discontinuation for all reasons (Figure 8.1-1 of Study Report).

The distribution of the number of reliable eGFR measurements per subject used in the sponsor's analysis by treatment arm is shown in Table 2. In this table, only subjects and measurements used in the sponsor's longitudinal eGFR analysis (Table 9.5.1.1-1 in the Study Report) are included. This includes measurements from end of titration through month 36 for subjects with at least 4 months of follow-up and at least 2 measurements and only counting measurements labeled as reliable. There are only 10 possible visits: End of titration/week 3, Months 4, 8, 12, 16, 20, 24, 28, 32, 36. However, a few subjects had two measurements that fell within a single visit window and both measurements were included. One subject had 11 measurements included in this analysis because they had two measurements in the Month 24 window and one measurement at every other possible visit. Figure 2 shows the cumulative distribution plot of time to last eGFR used in the sponsor's analysis. It can be seen that a relatively high proportion of subjects in the tolvaptan arm were not used in the analysis at all. More than 10% of the subjects in the tolvaptan arm were not included at all and more than 20% had no measurements beyond 1 year from randomization. Table 3 shows the distribution of number of eGFR measurements. The difference between this and the previous table is that it includes measurements labeled unreliable, subjects with less than 4 months follow-up, off-treatment measurements, and measurements from subjects with only one valid measurement. Of note, the sponsor's analysis used 11,785 measurements from 1306 subjects while there were 16,197 measurements from 1445 subjects in the full dataset. If every patient randomized had 13 measurements, there would have been 18,785 measurements.

Number of observations	Tolvaptan n (%)	Placebo n (%)	Total
2	33 (4)	9 (2)	42 (3)
3	23 (3)	7 (2)	30 (2)
4	19 (2)	7 (2)	26 (2)
5	19 (2)	11 (2)	30 (2)
6	15 (2)	12 (3)	27 (2)
7	16 (2)	4 (1)	20 (2)
8	27 (3)	20 (4)	47 (4)
9	105 (12)	57 (12)	162 (12)
10	584 (69)	337 (73)	921 (71)
11	1 (0)	0	1 (0)

Table 2 Distribution of number of eGFR measurements per subject used in sponsor's analysis.

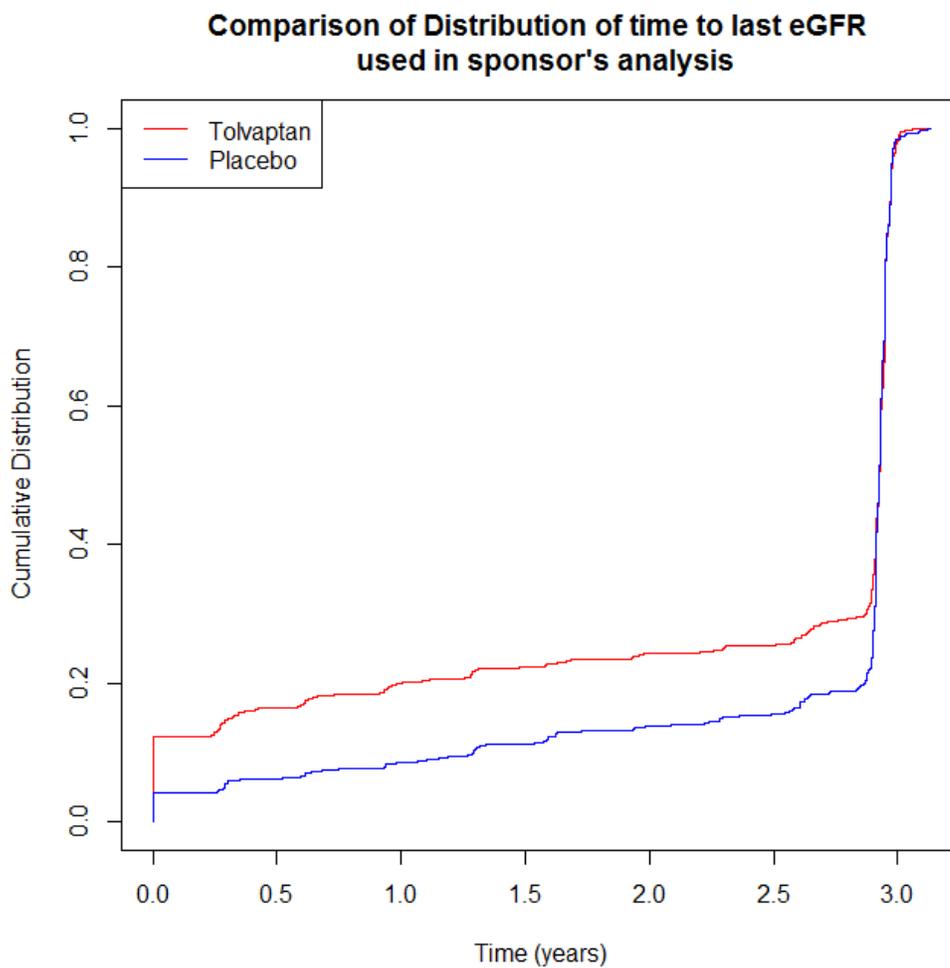


Figure 2 Kaplan-Meier plot of time to last eGFR measurement used in sponsor's analysis (Source: FDA)

Number of observations	Tolvaptan n (%)	Placebo n (%)	Total
1	7 (1)	3 (1)	10 (1)
2	45 (5)	5 (1)	50 (3)
3	56 (6)	5 (1)	61 (4)
4	23 (2)	10 (2)	33 (2)
5	19 (2)	8 (2)	27 (2)
6	22 (2)	9 (2)	31 (2)
7	13 (1)	9 (2)	22 (2)
8	13 (1)	9 (2)	22 (2)
9	7 (1)	4 (1)	11 (1)
10	13 (1)	4 (1)	17 (1)
11	23 (2)	19 (4)	42 (3)
12	110 (11)	49 (10)	159 (11)
13	609 (63)	349 (72)	958 (66)
14	1 (0)	1 (0)	2 (0)

Table 3 Distribution of number of eGFR measurements per subject actually measured.

The patient demographic characteristics are shown in Tables 2 and 3. The demographics were comparable between the two groups.

Demographic Characteristic	Tolvaptan			Placebo			Total		
	Male (N = 495)	Female (N = 466)	Total (N = 961)	Male (N = 251)	Female (N = 233)	Total (N = 484)	Male (N = 746)	Female (N = 699)	Total (N = 1445)
Age (years)									
Number of subjects	495	466	961	251	233	484	746	699	1445
Mean	38.2	38.9	38.6	38.3	39.4	38.8	38.3	39.1	38.7
SD	7.1	7.1	7.1	7.3	7.0	7.1	7.1	7.1	7.1
Median	39.0	40.0	39.0	39.0	40.0	39.0	39.0	40.0	39.0
Minimum	18	19	18	18	18	18	18	18	18
Maximum	51	50	51	50	50	50	51	50	51
Height (cm)									
Number of subjects	495	465	960	251	232	483	746	697	1443
Mean	180.4	166.2	173.5	180.0	166.6	173.6	180.3	166.4	173.6
SD	7.9	7.3	10.4	7.4	6.5	9.7	7.8	7.0	10.2
Median	180.0	166.0	173.0	180.0	167.0	173.0	180.0	167.0	173.0
Minimum	150	143	143	159	150	150	150	143	143
Maximum	210	192	210	201	188	201	210	192	210
Weight (kg)									
Number of subjects	495	466	961	251	233	484	746	699	1445
Mean	87.68	70.74	79.47	86.13	70.30	78.51	87.16	70.59	79.15
SD	15.80	16.59	18.27	16.96	16.04	18.31	16.21	16.4	18.28
Median	86.00	66.50	78.20	82.70	66.90	75.50	85.00	66.70	77.50
Minimum	50.6	40.6	40.6	54.7	46.0	46.0	50.6	40.6	40.6
Maximum	160.6	133.6	160.6	151.8	135.2	151.8	160.6	135.2	160.6
Race, ^a n (%)									
Caucasian	418 (84.4)	392 (84.1)	810 (84.3)	204 (81.3)	204 (87.6)	408 (84.3)	622 (83.4)	596 (85.3)	1218 (84.3)
Black	7 (1.4)	9 (1.9)	16 (1.7)	3 (1.2)	0	3 (0.6)	10 (1.3)	9 (1.3)	19 (1.3)
Hispanic	8 (1.6)	5 (1.1)	13 (1.4)	6 (2.4)	3 (1.3)	9 (1.9)	14 (1.9)	8 (1.1)	22 (1.5)
Asian	61 (12.3)	60 (12.9)	121 (12.6)	37 (14.7)	25 (10.7)	62 (12.8)	98 (13.1)	85 (12.2)	183 (12.7)
Other	1 (0.2)	0	1 (0.1)	1 (0.4)	1 (0.4)	2 (0.4)	2 (0.3)	1 (0.1)	3 (0.2)

Table 4 Patient demographic characteristics (Table 8.2-1 of Study Report)

Parameter	Tolvaptan (N = 961)	Placebo (N = 484)	Total (N = 1445)
1/serum creatinine ($[\text{mg/mL}]^{-1}$)			
Number of subjects	958	482	1440
Mean	102.27	104.30	102.95
SD	27.21	33.87	29.61
Median	100.00	100.00	100.00
Minimum	43.7	35.5	35.5
Maximum	263.2	500.0	500.0
eGFR_{CKD-EPI} ($\text{mL}/\text{min}/1.73 \text{ m}^2$)			
Number of subjects	958	482	1440
Mean	81.35	82.14	81.61
SD	21.02	22.73	21.60
Median	80.76	80.40	80.74
Minimum	32.3	26.4	26.4
Maximum	132.8	186.7	186.7
TKV (mL)			
Number of subjects	961	483	1444
Mean	1704.8	1667.5	1692.3
SD	921.27	873.11	905.31
Median	1456.7	1468.5	1458.8
Minimum	750.0	751.1	750.0
Maximum	7555.4	6751.1	7555.4

Table 5 Patient demographic characteristics (Table and 8.2-3 of Study Report)

1.4.4 Results and Conclusions

The drug had an effect on the primary endpoint, TKV. The effect is not linear over time, but rather there is a large initial drop in TKV in the tolvaptan arm and that difference is maintained for up to 3 years.

The sponsor used the log-transformation in their words, "to reduce heterogeneity in variance and achieve linearity over time" (Study Report). The residual variance was approximately homogeneous (see Figure 3). They were not normally distributed (skewness 2.22, excess kurtosis 6.3). In addition, $\log_{10}(\text{TKV})$ was not linear over time. One simple way to see this is to include a second degree term for time in the model (two extra fixed effects, one for each treatment group). When I did that, the log-likelihood improved by about 300 (note that an improvement of 3 in the log-likelihood with two extra parameters would be a significant improvement) and the AIC improved by almost 600.

As in the sponsor's eGFR analysis, their analysis of TKV did not include all the subjects. All 1445 subjects randomized had a baseline TKV measurement, but only 1277 were included in the sponsor's analysis.

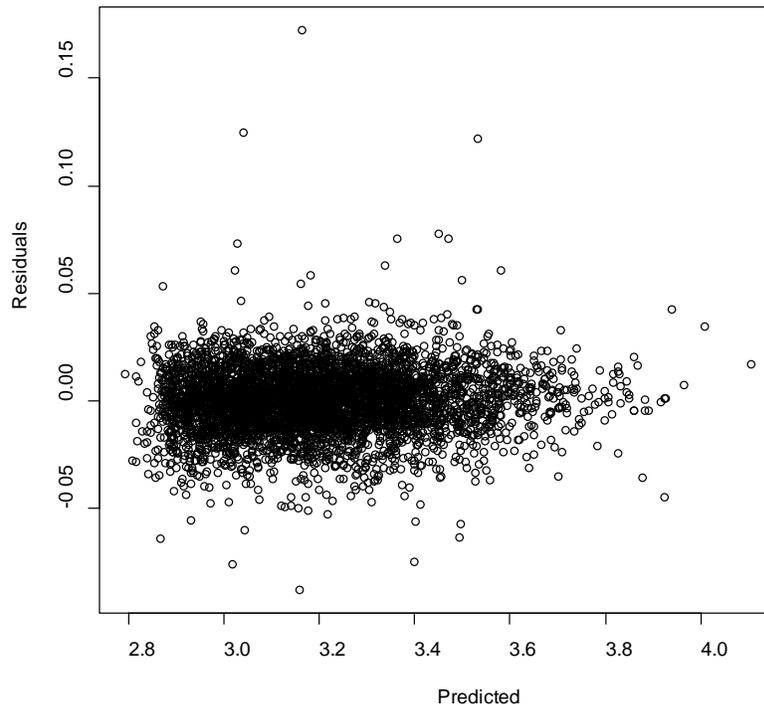


Figure 3 Scatterplot of residuals versus predicted log(TKV) from sponsor's model.

Like I did with the longitudinal change in eGFR, I tried fitting a model with an acute effect and a chronic effect on $\log_{10}(\text{TKV})$ and I included a quadratic function of time. I used all TKV measurements in the dataset. My estimated parameters were:

Intercept	3.18
time	0.0158
time ²	0.00258
acute treatment effect	-0.0223
chronic treatment*time interaction	-0.00433

Treatment*time² was not included because it did not improve the fit. The estimated standard deviations were: random intercept = 0.194, random slope = 0.0183, residual error = 0.0180. The estimated correlation between the random effects was 0.23.

According to the Study Report, the study was designed based on an assumption of a 7% annual increase in the placebo arm (this model actually estimates $10^{(0.0158+0.00258)}=1.043$, or a 4.3% rate of increase in the placebo group during the first year. The sponsor's analysis estimates a mean increase of 5.6% in the placebo group. Either way, the rate of growth was slower than expected. Also, the study design assumed a standard deviation of 0.017 for the residual error and

a standard deviation of 0.0184 for the random effect of slope (on the \log_{10} scale). Those values turned out to be almost exactly what the estimates are for those parameters in the FDA model.

Figure 4 shows the mean of observed $\log_{10}(\text{TKV})$ at each visit, as well as the predictions from the sponsor's model and the FDA model. I transformed everything back to the original scale of TKV by using the exponential function. There is one point for each treatment group at each of four visits. The x-coordinate is at the mean of the times when the observations happened and the y-coordinate is the geometric mean of the observations. The acute effect in the blue curve. It looks in the figure like the solid blue curve doesn't fit the points as well as it could, but that's OK because what the mixed effects model is doing is more complicated than just trying to come close to these group means.

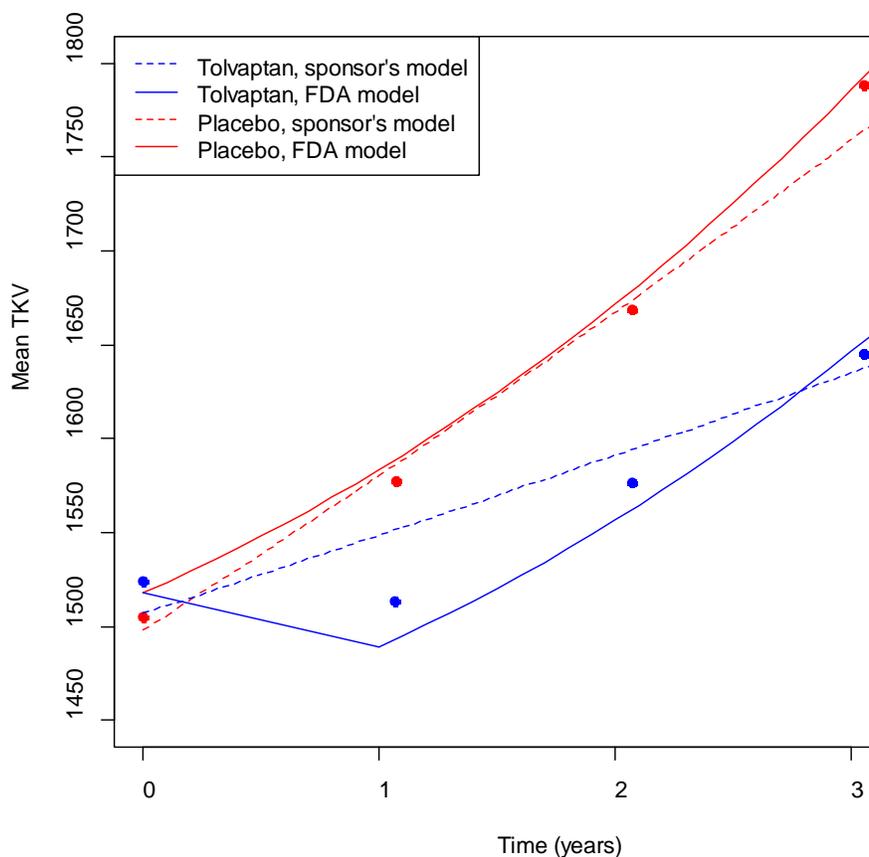


Figure 4 Mean observed and predicted TKV over time.

The trial did not confirm the drug has an effect on the key composite endpoint at a significance level of 0.01. The sponsor's analysis used the Anderson-Gill method for recurrent events. The sponsor's results were a hazard ratio estimate of 0.865 and p-value of 0.0095 using the

investigator reported events. I found the same total number of events (1049 vs. 665) in each group when I tried to repeat the sponsor's analysis and I found the same total number of years of follow-up in both groups (2387 vs. 1329). However, my estimates and p-values were slightly different (hazard ratio of 0.860 and p-value of 0.010).

The statistical issues with this analysis are three-fold. Missing data and ITT analysis, post-randomization baseline used for creatinine component, standard error is estimated under the alternative.

There were more subjects with missing values in the tolvaptan group, as discussed for other endpoints. The way that censoring was done in the sponsor's analysis used the last censoring time for all events, i.e. if there was follow-up on any of the four events, then the subject was not censored for the composite. Handling censoring for a composite endpoint with different censoring times for the components is not straightforward. The sponsor's sensitivity analyses are taken from the Study Report p. 208:

"

In response to a regulatory request to examine the effect of partial missing data on the result of the SAP-defined primary analysis of the key secondary composite endpoint, subjects could only contribute to the treatment group denominator at the last visit where an event occurred or where all 4 components were evaluated. This analysis also favored tolvaptan (HR 0.878, 95% CI 0.787 to 0.979, $p = 0.0194$) (CT-5.2.16.2).

...

Less restrictive ITT analyses were used to include data collected off treatment up to Month 36. Time to multiple composite ADPKD events analyses used a nonrestricted ITT approach (regardless of treatment period) using either predose baseline (CT-5.2.6.1; HR 0.874, 95% CI 0.784 to 0.974, $p = 0.0147$) or Week 3/EOT as baseline (CT-5.2.6.2; HR 0.889, 95% CI 0.797 to 0.992, $p = 0.0354$). Both of these analyses maintained statistical significance.

"

The use of post-randomization baseline for the definition of the creatinine event component and the subjects who dropped out in the first 4 months (and a much higher percentage in the tolvaptan group) complicate the interpretation of this analysis. There is no good way to handle this. It would be better to continue to collect data from subjects after they discontinue study drug. As long as I continue to see studies with a large amount of missing data, I think the best way to handle it is to put some kind of penalty in the analysis whereby subject from the placebo group with missing data are imputed with some kind of neutral or good value, but subjects from the treatment group are given a worse value. Because of the amount of missing data here, that kind of imputation will undoubtedly raise the p-value above 0.05.

Finally, the Anderson-Gill analysis uses a Wald-type estimate of the variance of the treatment effect estimate. That means, the variance is estimated under the alternative hypothesis. For a

clinical trial, when testing the null hypothesis, it is best to calculate the variance using the design-based method. That means, in part, that the variance should be estimated under the null hypothesis. I permuted the treatment assignments 10,000 times and found the variance of those 10,000 estimates. This does not fix the problems with missing data or anything else, it's only an attempt to find the correct distribution of the estimate under the null hypothesis. In comparison to the model-based p-value of 0.0095, the p-value using the permutation distribution is approximately 0.012.

The remainder of this section discusses changes in eGFR using the CKD-EPI equation.

The longitudinal analysis of eGFR is complicated because of acute and chronic effects. Many interventions that have effects on creatinine have different acute and chronic effects. This was anticipated and was the reason that the study was designed to have follow-up visits off treatment. The sponsor's analysis attempted to look only at the chronic effect by eliminating the measurements before titration and the measurements off treatment as well as the measurements that were labeled unreliable. However, besides throwing away a large amount of data, the sponsor's analysis had some other drawbacks. Their model assumes that eGFR changes in a linear way over time. Also, their model assumes the residual errors are independent, normally distributed, with a homogeneous variance. The data actually show that all these assumptions are false.

In the tolvaptan arm: the mean change in the 3 week titration phase was $-3.9 \text{ mL/min/1.73 m}^2$ and 71% of the subjects had a drop in eGFR. The mean change in the placebo arm was $-0.1 \text{ mL/min/1.73 m}^2$ and 47% had a drop in eGFR. These means and percentages are using the observed cases and the data from baseline and end of week 3 only (not based on any model). The estimated densities of the change in eGFR during the titration phase for both groups are shown in Figure 5.

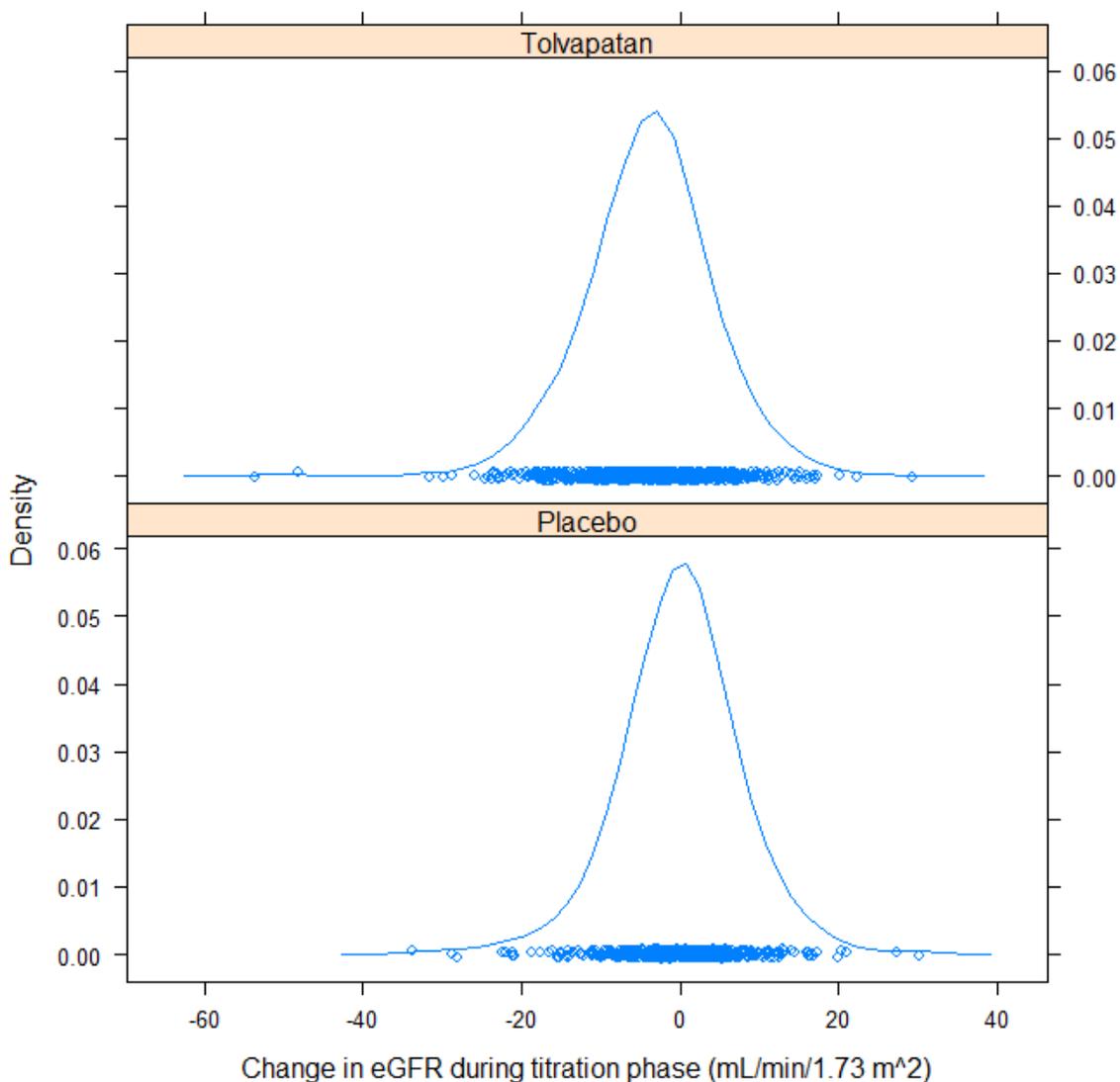


Figure 5 Estimated density of change in eGFR during initial 3 week titration period.

The sponsor's model for longitudinal changes of eGFR over time includes an intercept and terms for baseline, time, treatment group, treatment by time interaction. There are random effects within subject of intercept and time with unstructured covariance matrix. The estimates from the sponsor's model using the sponsor's data are:

Intercept	3.096
treatment	0.749
time	-3.700
baseline	0.954
treatment*time interaction	0.977

The estimated standard deviation of the random intercept is 3.227 and the estimated standard deviation of the random slope is 2.479 and their correlation is 0.663. The residual error standard deviation is 5.560.

The within subject residuals are shown in Figure 6. The red curves show the upper and lower 2.5 percentiles of the distribution as a function of the predicted value. These percentiles are estimated by quantile regression (I used the algorithm from <http://www.e-publications.org/ims/submission/index.php/AOAS/user/submissionFile/4295?confirm=37ca4b7>) and give some sense of whether the variance is homogeneous. In addition, one can divide the graph into 5 parts from left to right with equal number of points in each part and then calculate the sample variance of the residuals in each of the 5 sections. Doing that, I found variances (from left to right) of 11.5, 19.6, 39.9, 32.8, and 32.3. The three on the right are all significantly larger than the two on the left using the F-test for the ratio of the variances. Therefore, the variance is not homogeneous. Figure 7 shows the normal probability plot for the residuals which confirms they are not normal.

To investigate the linearity assumption, one way is to fit a more complicated model and compare the AIC and/or the likelihood ratio if the models are nested. For example, I tried a slightly more complicated model that includes a quadratic term for time and the interaction with treatment (same random effects as before). This more complicated model (with 2 additional parameters) fits the data better than the linear time model; the AIC improves by 40, minus $2 \times \log$ -likelihood ratio is 44, which has a p-value of close to 10^{-10} . Also, the model using $\log(\text{eGFR})$ as the response and replaces the covariate baseline by $\log(\text{baseline})$ fits the data better. It is more complicated to compare these two models and it cannot be done by comparing AIC or likelihood ratios. Instead, to account for the transformation, we have to add the sum of the $\log(\text{eGFR})$ to the likelihood in the first model to compare it with the likelihood of the second model. After accounting for the transformation of the response variable, the log-likelihood of the second model is larger by almost 71. The models have the same number of parameters and clearly the second model (using $\log(\text{eGFR})$) fits much better and so is the preferred model between the two.

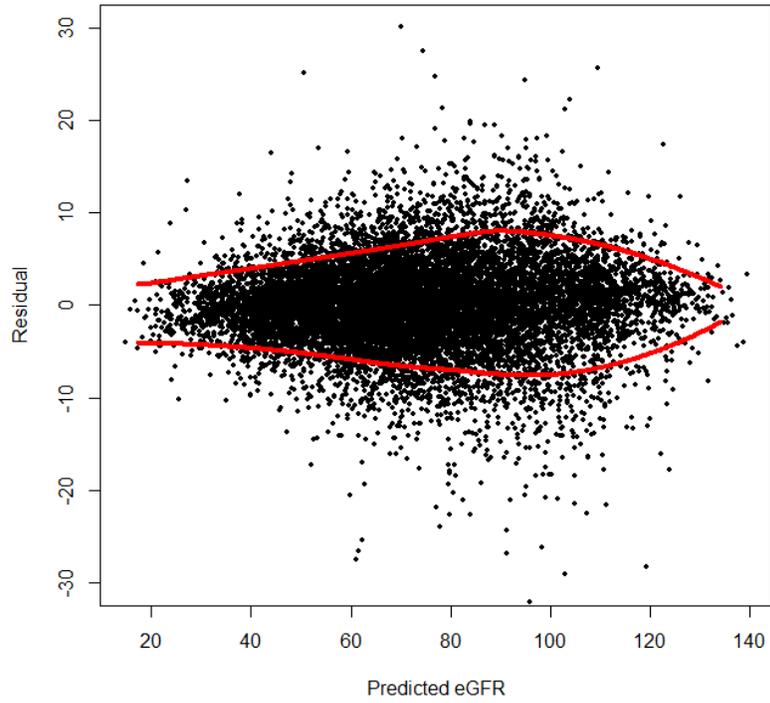


Figure 6 Residuals versus predicted scatterplot from sponsor's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 14 residuals with magnitude larger than 30 not shown.

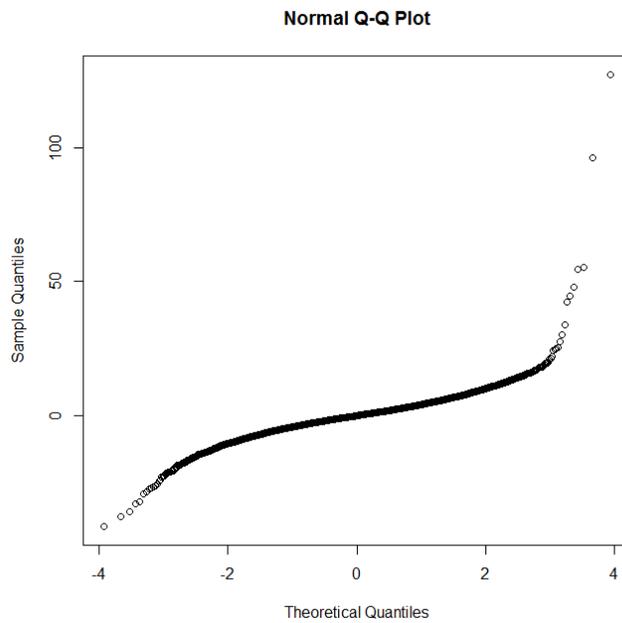


Figure 7 Normal probability plot for residuals from sponsor's eGFR model.

I tried to build a model I thought was reasonable for eGFR that: a) uses all of the measurements and b) accounts for possible acute and chronic effects. Using the log-transformation makes more sense over a long period of time if for no other reason because using a straight line without the transformation will eventually cross into the region where y is negative, but negative values of eGFR are not possible. Since using $\log(\text{eGFR})$ fit the data used in the sponsor's analysis better than eGFR confirms this intuition, I used that as a starting point for a model. Next, I included terms for acute drop of eGFR at the start of treatment and for acute rise of eGFR after stopping treatment. Finally, I considered other covariates, but I found that only baseline eGFR (at randomization), baseline $\log(\text{TKV})$, and age improved the fit significantly among the covariates I tried. Five people had missing baseline eGFR. Since I used baseline $\log(\text{eGFR})$ as a covariate in the model I needed to impute values for those 5 subjects. I tried values that were the same as the subjects' observed data at a nearby timepoint and I also tried other values that were biased against any treatment effect (adding 10 to the reasonable baseline for the two placebo subjects and subtracting 10 to the reasonable baseline from the 3 tolvaptan subjects to make it appear tolvaptan was not effective). However, the estimates in the model were essentially identical in both imputations.

The estimated fixed effects coefficients are:

Intercept	0.0852
$\log(\text{baseline TKV})$	-0.0382
$\log(\text{baseline eGFR})$	0.884
age	-0.00172
time	-0.358
treatment*time interaction	0.0204
$\log(\text{baseline TKV}) \cdot \text{time}$	-0.0205
$\log(\text{baseline eGFR}) \cdot \text{time}$	0.102
acute treatment effect at start	-0.0458
acute effect of withdrawal	0.0415

The estimated standard deviation of the random intercept is 0.0882 and the estimated standard deviation of the random slope is 0.0479 and their correlation is -0.052. The residual error standard deviation is 0.0804.

The residuals from this model are shown in Figure 8. The variance looks homogeneous up to the predicted $\log\text{-eGFR}$ of about 4.5 (eGFR of about 90). The normal probability plot shown in Figure 9 demonstrates that the residuals are not normally distributed. See the appendix for more details about this model including the distribution of the residual errors and the random effects. Also, see the appendix for examples of predictions of GFR for individual subjects based on this model and future predictions for the population based on this model.

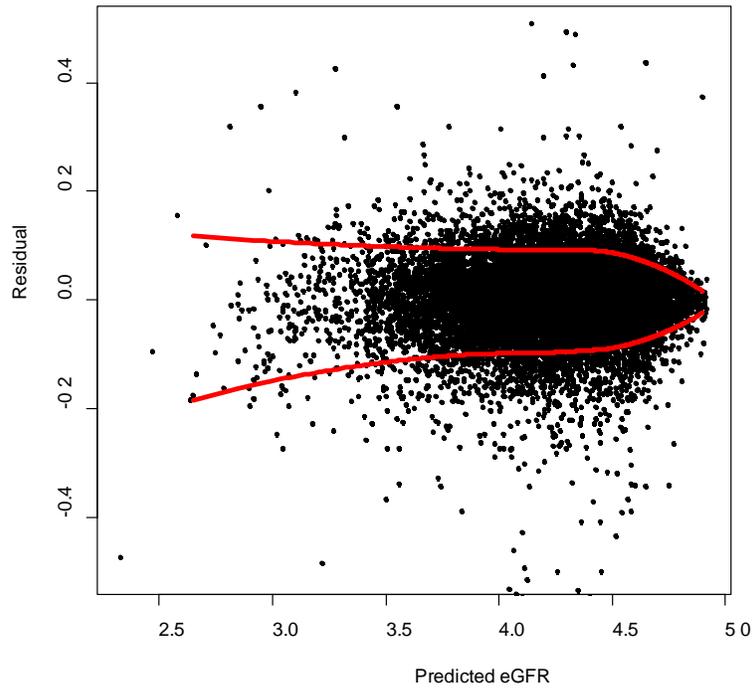


Figure 8 Residuals versus predicted scatterplot from FDA's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 19 residuals with magnitude greater than 0.5 not shown.

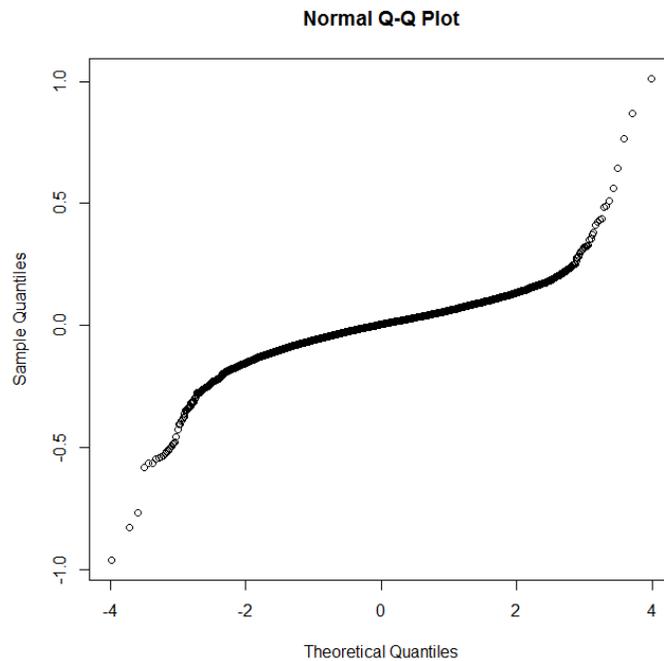


Figure 9 Normal probability plot for residuals from FDA's log-eGFR model.

1.5 Evaluation of Safety

See clinical review.

1.6 Benefit-Risk Assessment (Optional)

See clinical review.

FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

1.7 Gender, Race, Age, and Geographic Region

The sponsor's results for the key secondary endpoint are shown in Figure 10.

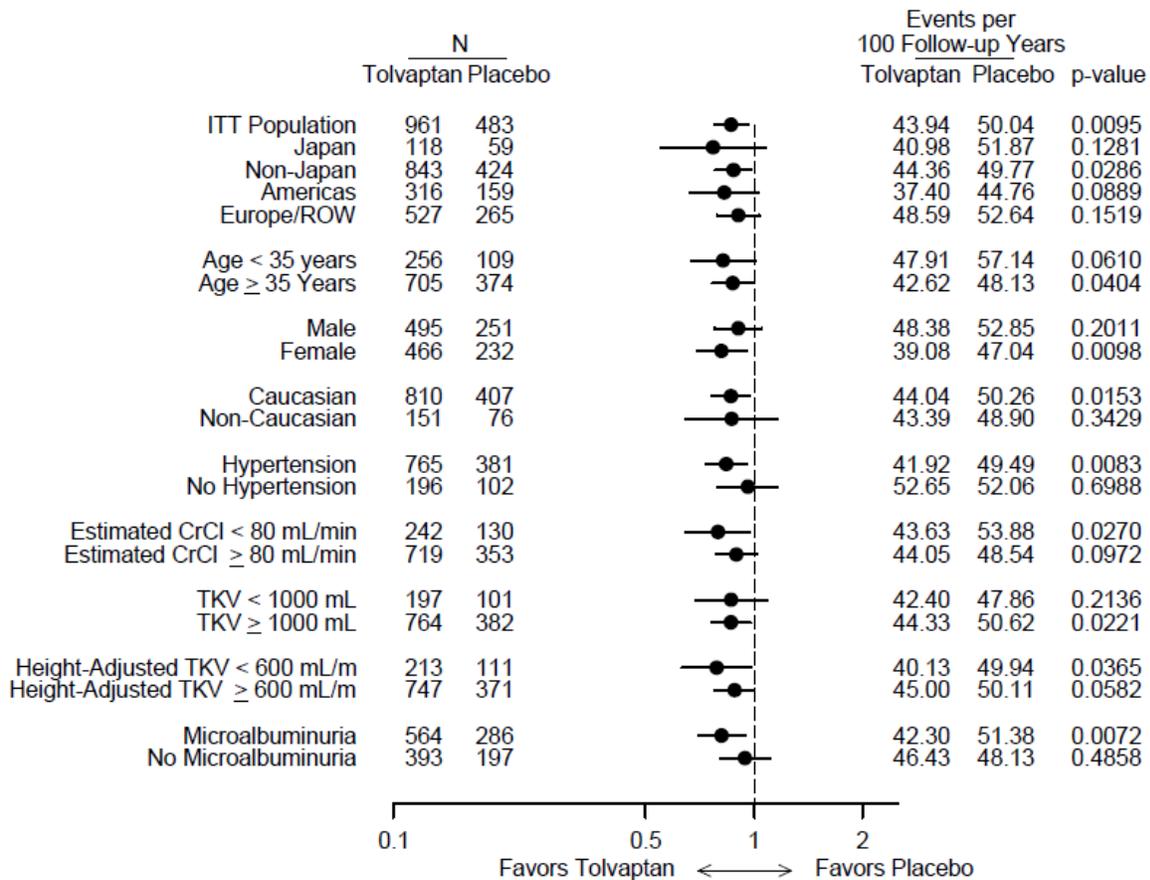


Figure 10 Sponsor's results of key secondary endpoint within subgroups (Study Report Figure 9.4.2-1)

1.8 Other Special/Subgroup Populations

About 4/5 of the subjects were taking an ACE inhibitor or ARB at randomization. The subjects taking those drugs had lower starting eGFR and higher TKV on average (76.4 vs. 89.8 mL/min/1.73 m² and 1598 vs. 1200 mL).

In the subgroup not taking ACEi/ARB, the average number of years of follow-up per subject were 2.12 years (tolvaptan) and 2.65 years (placebo). There were 41.0 events per 100 follow-up years (tolvaptan) and 46.6 events/100 follow-up years (placebo). The estimated hazard ratio for the key secondary endpoint was 0.82 in this subgroup.

In the subgroup taking ACEi/ARB, the average number of years of follow-up per subject were 2.57 years (tolvaptan) and 2.77 years (placebo). There were 44.5 events per 100 follow-up years (tolvaptan) and 50.7 events/100 follow-up years (placebo). The estimated hazard ratio for the key secondary endpoint was 0.86 in this subgroup.

SUMMARY AND CONCLUSIONS

1.9 Statistical Issues

The trial should have been planned with a type 1 error rate of 0.01 (two-sided) for a clinically meaningful endpoint, but was not.

Although the trial was technically blinded, the treatment assignment could have been guessed from effects on dehydration and water intake.

There were a high percentage of dropouts, particularly in the tolvaptan arm. Missing values were not imputed, but many subjects were not included at all in the sponsor's analyses. Other subjects were included with missing values but that is always raises problems, even without imputation.

Endpoints that used change in eGFR defined the baseline using a post-randomization value (post-titration) and a high percentage of subjects (particularly from the tolvaptan arm) had no post-titration value.

The analyses used assumptions that in some cases could be shown false with the data.

1.10 Collective Evidence

There was only one phase 3 trial in the submission.

1.11 Conclusions and Recommendations

The results on the clinical composite endpoint from the phase 3 trial, based on the sponsor's analysis, are just below the level they were told would be needed for approval ($p=0.0095$ when they were told they needed $p<0.01$ for approval). There is a large amount of missing data and use of a post-randomization baseline for change in eGFR. The Anderson-Gill method for recurrent events analysis estimates the variance under the alternative hypothesis. If we do nothing about the missing data or the post-randomization baseline, but just find the p-value from the exact distribution under the null hypothesis, the p-value from the recurrent events analysis is 0.012. Other analyses by the sponsor of the eGFR and TKV endpoints have the same problems related to missing data, but also use unverified model assumptions and in some cases use assumptions that can be demonstrated to be false.

1.12 Labeling Recommendations (as applicable)

NA.

APPENDICES

Distribution of residuals from FDA model of eGFR

The normal probability plot and any test of normality (Anderson-Darling, etc.) show the residuals are not normally distributed. The skewness is 2.88 and the excess kurtosis is 13.9. The empirical cumulative distribution function is shown in Figure A1. Also, the figure shows best fitting normal and Laplace distribution with parameters estimated by maximum likelihood. Neither fits very well, but I believe the Laplace distribution fits a little better. It is not easy to fit mixed effects models outside of the common assumptions of normally distributed errors. However, I think it may still be useful as far as modeling the mean true GFR, at least within the range of the time frame of 3 years from baseline. It may or may not be a reasonable model for extrapolation beyond 3 years.

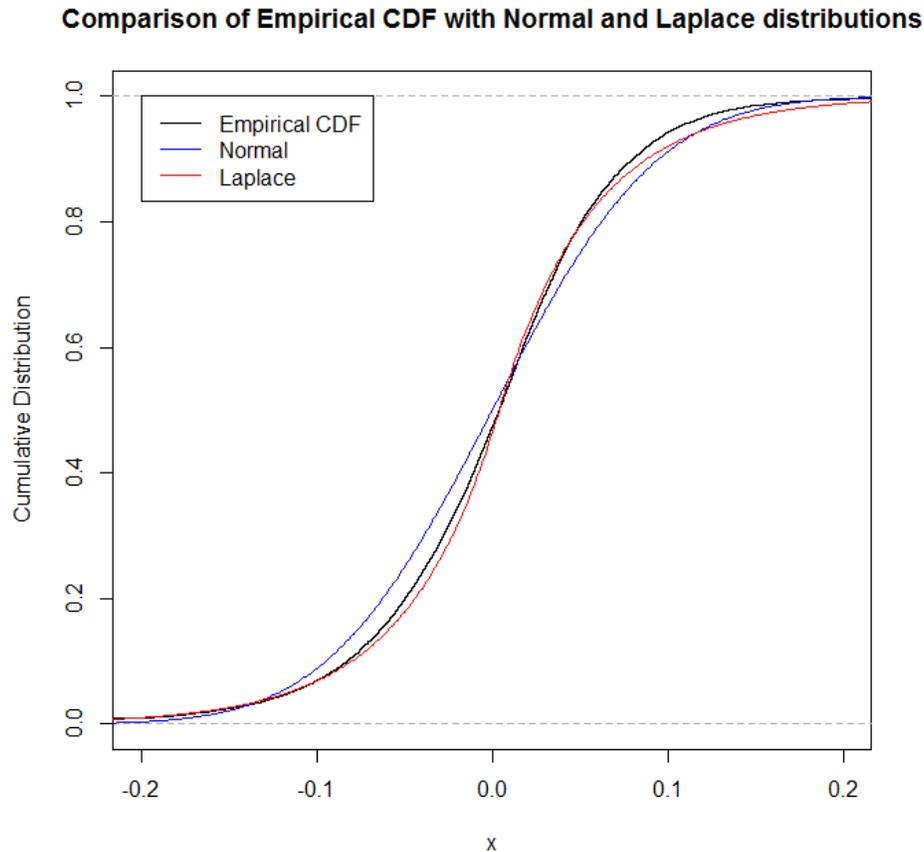


Figure A1 Comparison of Empirical CDF of within subject residuals from model described in the appendix with Normal distribution and Laplace distribution distributions with maximum likelihood estimates of parameters.

Distribution of the random effects

The estimated random effects are also not normally distributed. The scatterplot of the bivariate random effects is shown in Figure A2. Also, some of the estimated slopes are positive. The estimated slopes depend on the random effect for slope, but also on baseline TKV and eGFR. In more than 100 subjects, the estimated slope is still positive after subtracting the estimated chronic effect of treatment on the slope (the estimated slope would be positive for those subjects with no treatment). This doesn't seem to be biologically possible.

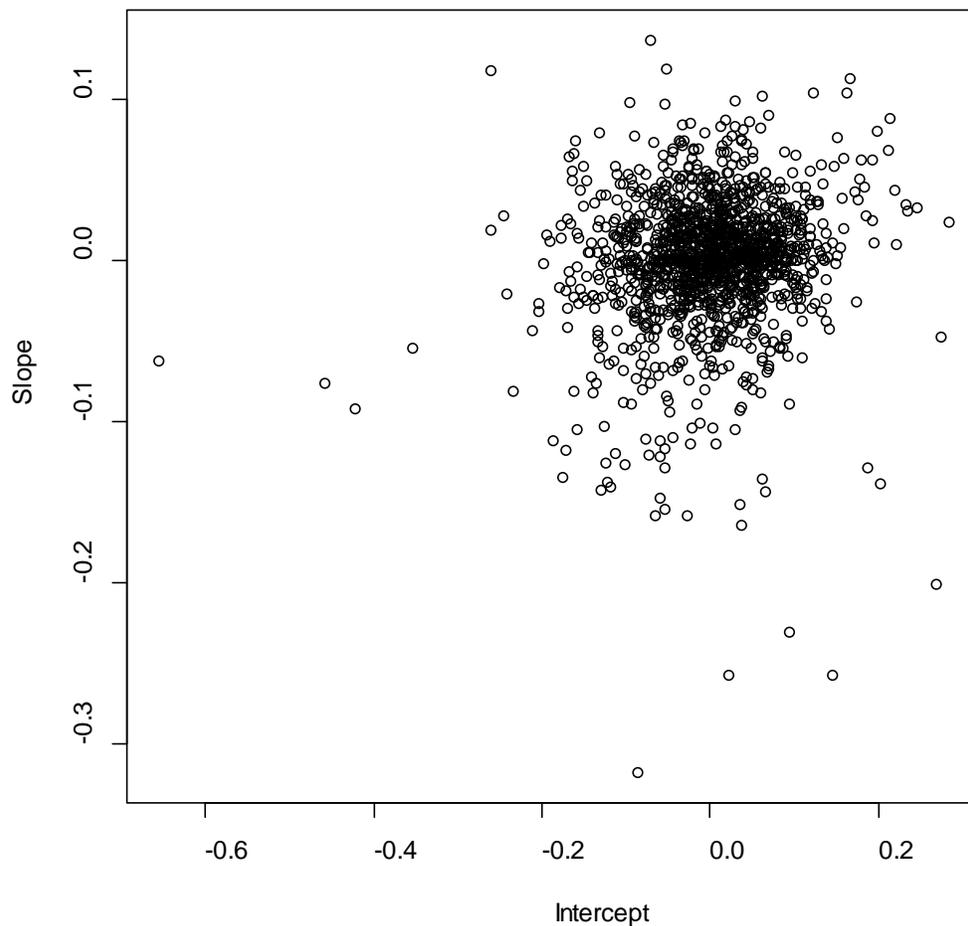


Figure A2 Scatterplot of estimated random intercept and random slope effects.

Model for eGFR between 30 and 45 mL/min/1.73 m²

The function $f(x)$ is the log-eGFR at time x , $y_1 = \log(45)$, x_1 is the time when the log-eGFR is y_1 , d_1 is the slope before reaching a eGFR of 45, $y_2 = \log(30)$, a is the acute effect of the drug withdrawal, d_2 is the slope after reaching the eGFR of 30. We want the acute treatment effect

and the chronic treatment effect to both disappear in a uniform way during the time interval between the eGFR of 30 and 45.

Problem. Suppose that $x_1, y_1, d_1, y_2, a,$ and d_2 are given and that $f(x_1) = y_1, f'(x) = d_1$ for x in a neighborhood to the left of x_1 . Can we define a continuous extension of f onto the interval $(x_1, x_2]$ for some $x_2 > x_1$ such that the following two conditions hold: i) $f(x_2) = y_2$, and ii) $f'(x) = d_1 + (x - x_1) \frac{d_2 - d_1}{x_2 - x_1} + \frac{a}{x_2 - x_1}$ for all $x \in (x_1, x_2)$?

Solution. By taking the anti-derivative of both sides of the equation in the second condition, we find

$$f(x) = c_0 + \left(\frac{a - d_2 x_1 + d_1 x_2}{x_2 - x_1} \right) x + \left(\frac{d_2 - d_1}{x_2 - x_1} \right) \frac{1}{2} x^2$$

Now, use the conditions $f(x_1) = y_1$ and $f(x_2) = y_2$ and solve those two equations simultaneously for the two unknowns c_0 and x_2 to find

$$x_2 = x_1 - \frac{2(a + y_1 - y_2)}{d_1 + d_2}$$

and

$$c_0 = \frac{(d_1 + d_2)x_1(2a + (d_1 - d_2)x_1)}{4(a + y_1 - y_2)} - d_1 x_1 + y_1$$

Examples

Start with one example from the dataset, the first subject in the dataset. This subject was 46 years old with a baseline TKV of 2343.9, baseline eGFR of 70.0 mL/min/1.73 m² and was randomized to tolvaptan. He completed the trial and had 13 total eGFR measured including both follow-up visits. Those two follow-up visits are included in the Figure A3 below using filled circles. There are three scenarios shown, one (in red) assumes he never took the drug, the second (in blue) is where tolvaptan is assumed to always have the same effect. In those first two scenarios, log-eGFR after baseline is a straight line with a constant slope, but the slope is different in the two scenarios. The actual slopes (for log-eGFR) in those two scenarios are estimated from the mixed effects model. The third scenario is shown in brown. This follows the blue curve exactly until GFR hits 45, then uses the solution to the equation shown above for times between GFR of 45 and 30, then has a constant slope identical to the slope of the red curve (on the log scale). It can be seen that during this time period of losing drug effects, the recapture of the acute effect makes the brown curve rise above the blue curve, but later, the blue curve is on top again.

The predicted eGFR shown on the y-axis is a prediction in this sense. For log-eGFR, the prediction is the expected value of an observation at that time point assuming the model with the estimated parameters and the empirical Bayes estimate of the random effects for this subject. It is the mean and median of an observation at that time with those assumptions. I transformed this prediction to the original scale of eGFR by evaluating the exponential function at that prediction. This is no longer the expected value of an observation on the original scale, but it is the median of the distribution of those values. Other ways of handling the transformation in the prediction

may be better. As already noted, the residuals on the log-scale are not normally distributed or even symmetric, so methods based on that assumption might not be adequate.

46 yr old, baseline TKV = 2343.9, 1

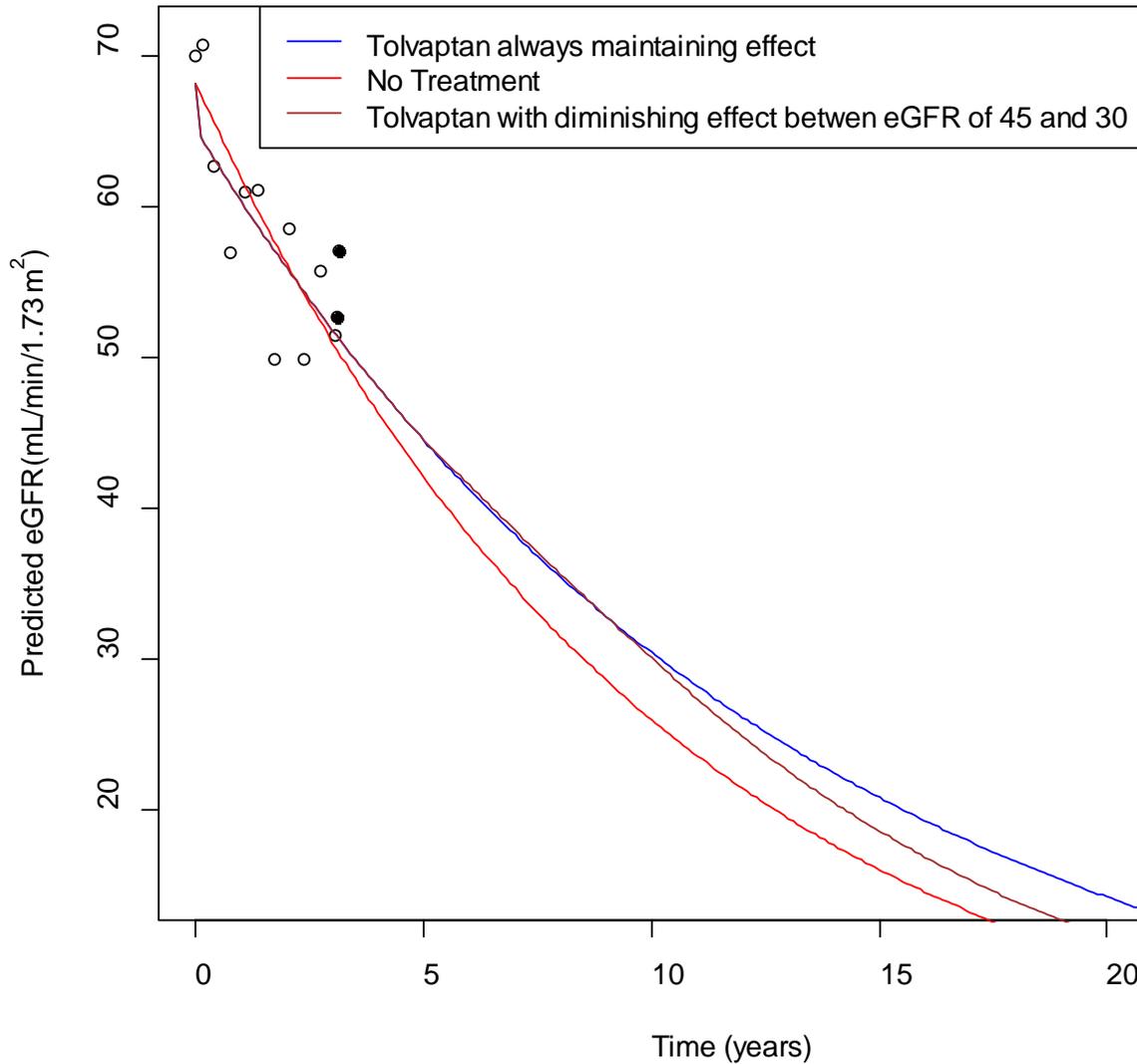


Figure A3 Predicted eGFR for one subject.

A second example from the dataset, a 20 year old subject with baseline eGFR of 88.3 and TKV of 5546.1 is shown in Figure A4. For almost the entire time shown in the figure, the brown curve is on top of the blue curve (the blue curve is there, but hidden under the brown curve). This figure shows that the model predicts an enormous benefit in terms of delaying ESRD. In fact, even reaching a GFR of 60 (CKD stage 3) is delayed by the treatment for a good 15 years in this figure. However, the figure also shows that we only have data for a short period of time. It is left

to the reader to decide how much faith to put into extrapolations of treatment effects 40 years into the future even if they are totally convinced that there is a beneficial effect on eGFR over 3 years.

20 yr old, baseline TKV = 5546.1, 1

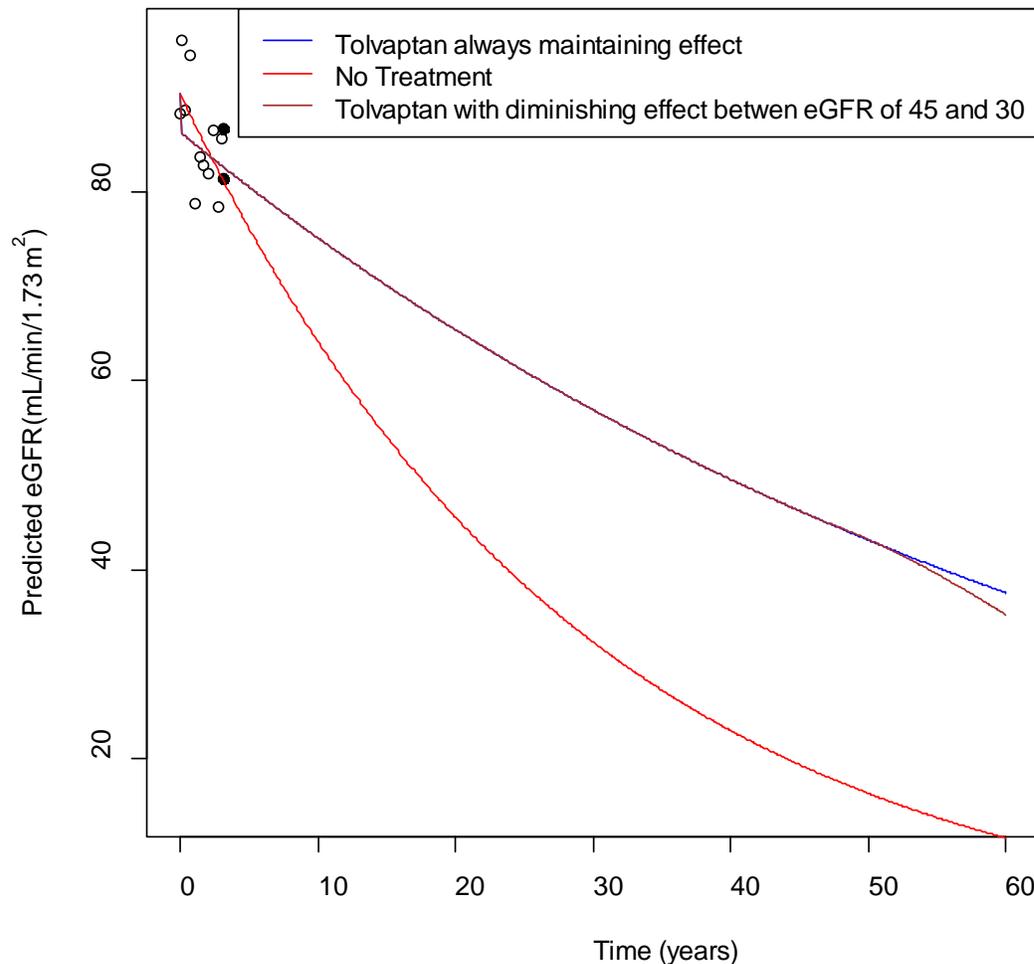


Figure A4 Predicted eGFR for one subject.

Using the model to predict into the future

Taking the subjects in the trial and using their age, baseline TKV, baseline eGFR and estimated random effects, the model allows us to extrapolate into the future and also allows us to estimate what would happen for each subject if they took tolvaptan or did not. We can then estimate for each time point in the future what proportion of subjects would have $GFR < 15 \text{ mL/min/1.73 m}^2$ and compare those proportions assuming all the subjects took tolvaptan and had that estimated

treatment effect versus not taking tolvaptan. These are big assumptions about what will happen in the future and could be very far-fetched. I would not even know how to put confidence intervals around them that would reflect all the uncertainties and assumptions. Figure A5 shows that the treatment effect could be somewhere around a 4 year delay in the time to GFR<15 mL/min/1.73 m².

To be more specific, start with the estimated coefficients in the model:

Intercept	0.0852
log(baseline TKV)	-0.0382
log(baseline eGFR)	0.884
age	-0.00172
time	-0.358
treatment*time interaction	0.0204
log(baseline TKV)*time	-0.0205
log(baseline eGFR)*time	0.102
acute treatment effect at start	-0.0458
acute effect of withdrawal	0.0415

For subject i , let α_i and β_i be their estimated random effects. If they took no drug, the predicted log-GFR at time t years from randomization is:

$$\log(\text{GFR}(t)) = 0.0852 + \alpha_i - 0.0382 \log(\text{baseline TKV}_i) + 0.884 \log(\text{baseline eGFR}_i) - 0.00172 \text{ age}_i + \{-0.358 + \beta_i - 0.0204 \log(\text{baseline TKV}_i) + 0.102 \log(\text{baseline eGFR}_i)\} * t$$

and their estimated time when their GFR is 15 is:

$$\tau_i = \{\log(15) - (0.0852 + \alpha_i - 0.0382 \log(\text{baseline TKV}_i) + 0.884 \log(\text{baseline eGFR}_i) - 0.00172 \text{ age}_i) / \{-0.358 + \beta_i - 0.0204 \log(\text{baseline TKV}_i) + 0.102 \log(\text{baseline eGFR}_i)\}\}$$

The estimated proportion of subjects with GFR<15 at time t is then

$$\frac{1}{n} \sum_{i=1}^n I(\tau_i < t)$$

The red curve in the figure is a graph of this for t between 0 and 40.

The blue curve is more complicated because there is no fixed τ_i for each subject assuming they take the drug. The time to reach GFR<15 depends now on how long they take the drug, which is a random variable. I assumed the time to withdrawal had a constant hazard in the first 4 months and another different constant hazard beyond 4 months. The hazards were defined to make it so they had a 10% chance of withdrawal during the first 4 months and, if they passed that point, a 5% chance of withdrawal each year thereafter.

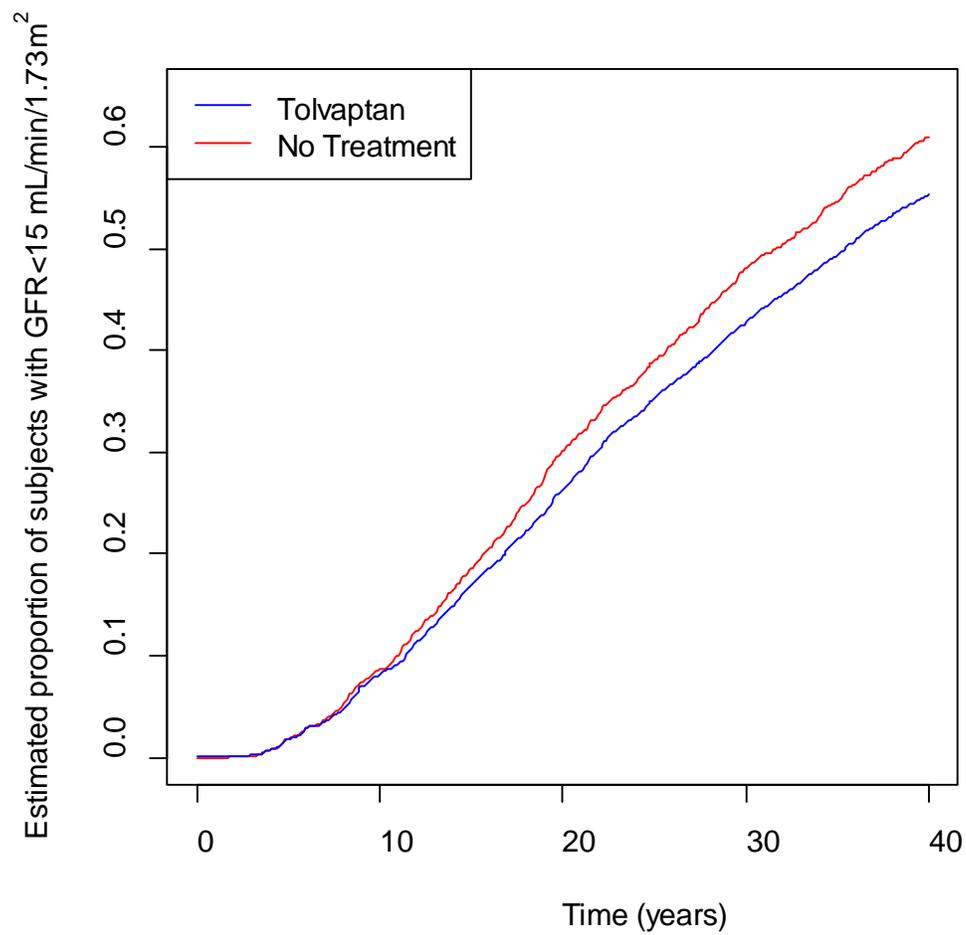


Figure A5 Estimated proportion of subjects with $\text{GFR} < 15 \text{ mL/min/1.73 m}^2$ using FDA's model and extrapolating from 3 years of data into 40 years into the future.

The next figure is based on a similar model, but assumes that after 3 years there is no treatment effect.

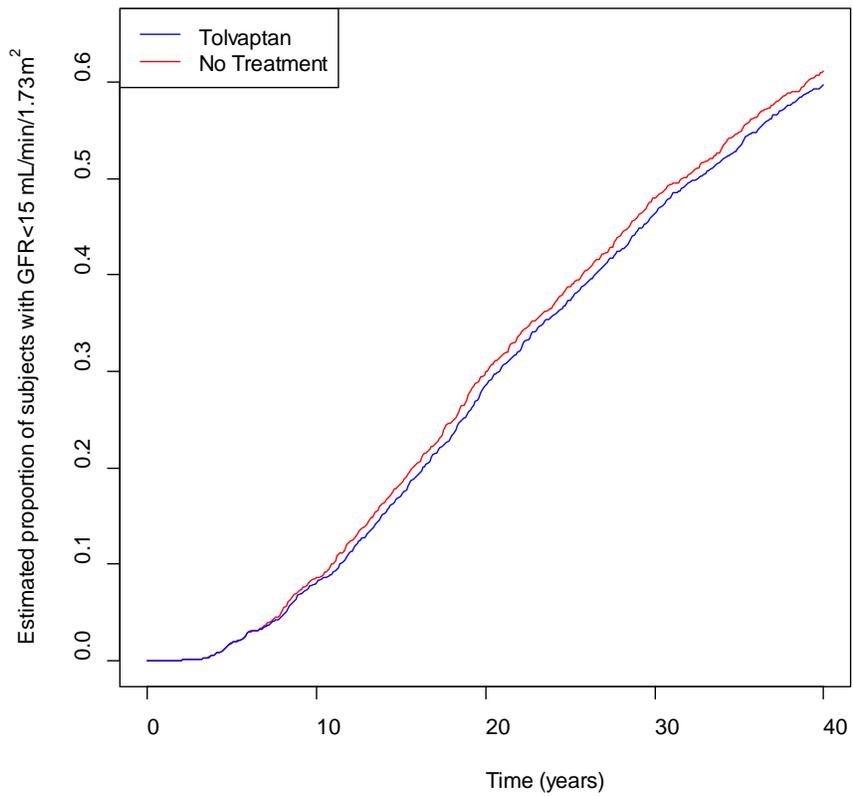


Figure A6 Estimated proportion of subjects with $\text{GFR} < 15 \text{ mL/min/1.73 m}^2$ using FDA's model, assuming no treatment effect beyond 3 years.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

JOHN P LAWRENCE

08/22/2013

corrected some errors in references to numbered tables or figures and other minor corrections. this supersedes review entered in DARRTS June 25, 2013.

HSIEN MING J HUNG

08/22/2013



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #: NDA 204-441
Supplement #:
Drug Name: Tolvaptan
Indication(s): Slow progressive kidney disease in adults with
autosomal dominant polycystic kidney disease
Applicant: Otsuka
Date(s): 11/15/2012
Review Priority: Priority
Biometrics Division: DBI
Statistical Reviewer: John Lawrence, Ph D
Concurring Reviewers: Jim Hung
Medical Division: Cardiorenal.
Clinical Team: Aliza Thompson MD, Nhi Beasley MD, Steven Grant, MD
Project Manager: Anna Park
Keywords:
survival analysis, benefit-risk, mixed models, longitudinal data analysis

Table of Contents

EXECUTIVE SUMMARY	5
INTRODUCTION	6
1.1 OVERVIEW.....	6
1.2 DATA SOURCES	7
STATISTICAL EVALUATION	7
1.3 DATA AND ANALYSIS QUALITY	7
1.4 EVALUATION OF EFFICACY	8
1.4.1 <i>Study Design and Endpoints</i>	8
1.4.2 <i>Statistical Methodologies</i>	8
1.4.3 <i>Patient Disposition, Demographic and Baseline Characteristics</i>	9
1.4.4 <i>Results and Conclusions</i>	13
1.5 EVALUATION OF SAFETY	23
1.6 BENEFIT-RISK ASSESSMENT (OPTIONAL).....	23
FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	23
1.7 GENDER, RACE, AGE, AND GEOGRAPHIC REGION	23
1.8 OTHER SPECIAL/SUBGROUP POPULATIONS	24
SUMMARY AND CONCLUSIONS	24
1.9 STATISTICAL ISSUES	24
1.10 COLLECTIVE EVIDENCE.....	24
1.11 CONCLUSIONS AND RECOMMENDATIONS	25
1.12 LABELING RECOMMENDATIONS (AS APPLICABLE).....	25
APPENDICES.....	26
DISTRIBUTION OF RESIDUALS FROM FDA MODEL OF EGFR	26
DISTRIBUTION OF THE RANDOM EFFECTS	27
MODEL FOR EGFR BETWEEN 30 AND 45 mL/MIN/1.73 M ²	27
EXAMPLES	28
USING THE MODEL TO PREDICT INTO THE FUTURE	30

LIST OF TABLES

Table 1 Patient disposition (Table 8.1-1 of Study Report)	9
Table 2 Distribution of number of eGFR measurements per subject used in sponsor's analysis.....	11
Table 3 Distribution of number of eGFR measurements per subject actually measured.....	12
Table 4 Patient demographic characteristics (Table 8.2-1 of Study Report)	12
Table 5 Patient demographic characteristics (Table and 8.2-3 of Study Report).....	13

LIST OF FIGURES

Figure 1 Kaplan-Meier plot of time to discontinuation for all reasons (Figure 8.1-1 of Study Report).	10
Figure 2 Kaplan-Meier plot of time to last eGFR measurement used in sponsor's analysis (Source: FDA)	11
Figure 3 Scatterplot of residuals versus predicted log(TKV) from sponsor's model.	14
Figure 4 Mean observed and predicted TKV over time.....	15
Figure 5 Estimated density of change in eGFR during initial 3 week titration period.....	18
Figure 6 Residuals versus predicted scatterplot from sponsor's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 14 residuals with magnitude larger than 30 not shown.	20
Figure 7 Normal probability plot for residuals from sponsor's eGFR model.....	20
Figure 8 Residuals versus predicted scatterplot from FDA's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 19 residuals with magnitude greater than 0.5 not shown.	22
Figure 9 Normal probability plot for residuals from FDA's log-eGFR model.....	22
Figure 10 Sponsor's results of key secondary endpoint within subgroups (Study Report Figure 9.4.2-1)	23

EXECUTIVE SUMMARY

This submission contains one Phase 3 study to support the indication. Accordingly, the level of evidence from that trial must be equivalent to two trials with a type I error rate of 0.05 each. According to the medical division, the primary endpoint of total kidney volume is not acceptable for approval. This review focuses on the sponsor's key secondary endpoint, a composite endpoint consisting of events defined by hypertension, renal function, renal pain, and albuminuria. In addition, this review focuses on exploratory analyses of longitudinal changes in estimated kidney function (glomerular filtration rate estimated by the CKD-EPI formula).

There were several statistical issues with the analyses. There was possibly non-ignorable missing data and substantially more missing data in the tolvaptan arm compared to the placebo arm. In some analyses, the ITT population could not be used because there were no valid observations. In addition, assumptions used in the models were clearly violated (assumptions about linear responses over time and assumptions about homogeneous variance of residual errors). Tolvaptan has substantial acute effects on estimated GFR and on total kidney volume that are different than chronic effects. Therefore, simple models do not adequately fit the data.

INTRODUCTION

1.1 Overview

Table: List of all studies included in analysis

	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
<i>Study 156-04-251</i>	<i>Phase 3</i>	<i>36 months</i>	<i>36 months</i>	<i>tolvaptan: 961 placebo:484</i>	<i>subjects with ADPKD as defined by a certain number of cysts, estimated creatinine clearance of at least 60 mL/min and TKV>750 mL.</i>

There was one Phase 3 trial conducted to support this indication. A Special Protocol Assessment was done, but the FDA did not agree with the Protocol. Since there was only one study, a type 1 error rate of 0.01 was to be used for approval decisions. This was communicated to the sponsor. The primary endpoint of TKV was never acceptable to the FDA, but the key secondary endpoint was an acceptable endpoint.

The meeting minutes from a face to face meeting between the FDA and the sponsor on June 10, 2009 state:

"2) We propose that a significance level of 0.0491 (two-sided) will be used to declare statistical significance at the final analysis for the primary endpoint. In addition, we propose that a significance level of 0.05 (two-sided) will be used to declare statistical significance at the final analysis for the key secondary composite endpoint. In a Type A meeting with the Division on 15 Nov 2005 (minutes provided as [Attachment 2](#)), Otsuka proposed, "if the primary endpoint and composite key secondary endpoint are both statistically significant, and if the other specified endpoints are supportive, the data from this single phase 3 trial will be sufficient to support a New Drug Application (NDA) approval for the proposed indication." The Division agreed to Otsuka's proposal. **Does the FDA agree that the significance levels specified in the draft SAP are acceptable for approval based on a single pivotal trial?**

Preliminary FDA Response: A p-value < 0.05 from a single trial is acceptable for your primary efficacy endpoint because we do not consider this endpoint a surrogate of benefit. In order to provide convincing evidence of treatment benefit, the composite key secondary endpoint will need a p-value < 0.01.

Additional discussion during the meeting: The sponsor has decided to continue their study as proposed and is aware the Division will likely review the results in a more stringent fashion. Dr. Stockbridge reiterated that the Division was less interested in the primary endpoint as compared to the secondary endpoints. The Division acknowledged the sponsor's decision."

However, the FDA defines a primary endpoint in its guidance document as "Endpoint(s) necessary and/or sufficient to establish efficacy" (not published as of this date, but that definition appears in the slide presentation here:

http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2013/03/WC50

0140627.pdf). Since TKV was not necessary or sufficient to establish efficacy, then by the FDA guidance document's definition, it was not a primary endpoint. Even if you don't rely on the definition from the guidance document (which is fair since it is not even published now), it is clear from the minutes above that the FDA told the sponsor that TKV could not be the primary endpoint of the trial. Despite multiple attempts to explain to the sponsor that TKV was not a primary endpoint, the company insisted on calling it the primary endpoint and the FDA was powerless to stop them. In these same meeting minutes, they discuss a plan to stop the trial early at an interim analysis if a benefit was shown on TKV (this adjustment for the interim analysis is the reason for the significance level of 0.0491). This illustrates the difference between how much importance the FDA put on TKV compared to the how much the company put; the company intended to stop the trial early and claim victory if a benefit was shown on TKV while the FDA was telling them they had no interest in TKV.

1.2 Data Sources

Electronic datasets and Study Reports:

<\\cdsesub1\evsprod\NDA204441\204441.enx>

<\\cdsesub1\evsprod\NDA204441\0001\m5\datasets\156-04-251\analysis>

STATISTICAL EVALUATION

1.3 Data and Analysis Quality

The data quality and analysis quality were both poor.

Many (several thousand) serum creatinine measurements were not included in the sponsor's analysis. There were many subjects that were not included in the sponsor's analysis at all. Other subjects had partial data. Subjects with some missing data is common in clinical trials, but the amount of missing or unreliable data in this trial is uncommon (compared to other trials of cardiovascular or renal disease). In many cases, subjects were not followed at all, or only for a short time if they stopped treatment early. A true intent-to-treat analysis should follow all subjects for all outcomes for the entire planned period (36 months). This was not done here. Baseline for changes in serum creatinine or eGFR was defined as the measurement after titration. This caused many subjects to be excluded from the analysis completely if they could not tolerate the drug during the titration phase. It is very uncommon to define a baseline value so long after randomization (approximately 3 weeks). If all the subjects are still in the trial at that time, there is less of a concern, but that was not the case here.

The sponsor's analysis used assumptions that in some cases can be demonstrated to be false and in other cases could not be verified. The mixed effects models include an assumption that the

residual error variance is homogeneous and that those errors are normally distributed. For the TKV endpoint (after log transformation) and for the eGFR endpoint, both of these assumptions can be shown false using the data. In addition, the sponsor's analyses used simple linear response models. For both endpoints, those models were not adequate and that can be shown with the data. Furthermore, these models use other assumptions about the distribution of random effects and the nature of missing data (missing at random) that cannot be verified. Lastly, the analysis of recurrent events uses assumptions in the estimate of the variance that may exaggerate the significance of the p-value for that analysis (see Section 1.4.4).

1.4 Evaluation of Efficacy

1.4.1 Study Design and Endpoints

Study 156-04-251 was a multinational, multicenter trial. 1445 subjects were randomized 2:1 to tolvaptan or placebo. The primary endpoint was change in TKV (total kidney volume) over time. TKV was measured at baseline and every 12 months up to month 36 by MRI. The key secondary endpoint was a composite of clinically relevant outcomes. The composite consisted of four types of events: hypertensive progression (change in category or addition of hypertension medication); renal pain; worsening of albuminuria; worsening of renal function (confirmed rise of 33% in serum creatinine). The composite endpoint was counted with recurrence possible, i.e. not just the first event for each subject, but rather multiple events for each subject were possible and all were counted. Change in renal function (inverse of serum creatinine and other estimates of creatinine clearance or GFR) were also secondary or exploratory endpoints.

1.4.2 Statistical Methodologies

The primary endpoint, TKV, was analyzed using a mixed effects model. First, the TKV was transformed using the base 10 logarithm. Time was measured in years from the time of the first (baseline) TKV (number of days divided by 365.25) and was included as a continuous variable in the model.

The following linear mixed-effect model was fitted to the log-transformed TKV repeated-measures data:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 \text{Group}_i + \beta_4 t_{ij} \times \text{Group}_i + b_{1i} + b_{2i} t_{ij} + e_{ij},$$

In this model, Y_{ij} is the \log_{10} (TKV) of subject i at visit j ($j = 0, 1, 2, 3$), where $\text{Group}_i = 0$ for a subject in the placebo group and $\text{Group}_i = 1$ for a subject in the tolvaptan group. β_1 , β_2 , β_3 , and β_4 are fixed effects (β_1 is the intercept of placebo, $\beta_1 + \beta_3$ is the intercept of tolvaptan, β_2 is the slope of placebo, and $\beta_2 + \beta_4$ is the slope of tolvaptan), while b_{1i} and b_{2i} are random effects assumed to be normally distributed with mean 0 and unknown

variance covariance structure. The error terms in the model, e_{ij} , are assumed mutually independent and normally distributed as $N(0, \sigma^2)$, and they are also assumed to be independent of the random effects. The primary null hypothesis is $H_0: \beta_4 = 0$ versus the alternative hypothesis $H_1: \beta_4 \neq 0$.

The key secondary endpoint was analyzed using the Anderson-Gill recurrent events model. No covariates were included other than treatment group. Subjects were censored at the last censoring time for all components and were considered to have no events of the type without follow-up at those times where unknown.

1.4.3 Patient Disposition, Demographic and Baseline Characteristics

The patient disposition are shown in Table 1. Significantly more subjects discontinued in the tolvaptan arm. the bottom row shows some of the subjects who discontinued were followed for some PKD outcomes, but that means a phone call in many cases and is not the same as complete follow-up on all outcomes. Figure 1 shows that the proportion of subjects in the tolvaptan arm who discontinued was larger than the proportion in the placebo arm uniformly throughout the trial.

Number of Subjects	Tolvaptan (N = 961) n (%)	Placebo (N = 484) n (%)	Total (N = 1445) n (%)
Screened	-	-	2122
Randomized	961 (100.0)	484 (100.0)	1445 (100.0)
Treated	961 (100.0)	483 (99.8)	1444 (99.9)
Completed	740 (77.0) ^a	417 (86.2) ^b	1157 (80.1)
Discontinued IMP	221 (23.0)	67 (13.8)	288 (19.9)
Lost to follow-up	15 (1.6)	8 (1.7)	23 (1.6)
AE	148 (15.4)	24 (5.0)	172 (11.9)
Subject met withdrawal criteria	4 (0.4) ^c	0 (0.0)	4 (0.3)
Investigator withdrew subject	3 (0.3)	4 (0.8)	7 (0.5)
Subject withdrew consent	50 (5.2)	30 (6.2)	80 (5.5)
Protocol deviation	1 (0.1) ^d	1 (0.2) ^d	2 (0.1)
Discontinued and followed for PKD Outcomes	102 (10.6)	27 (5.6)	129 (8.9)

Table 1 Patient disposition (Table 8.1-1 of Study Report)

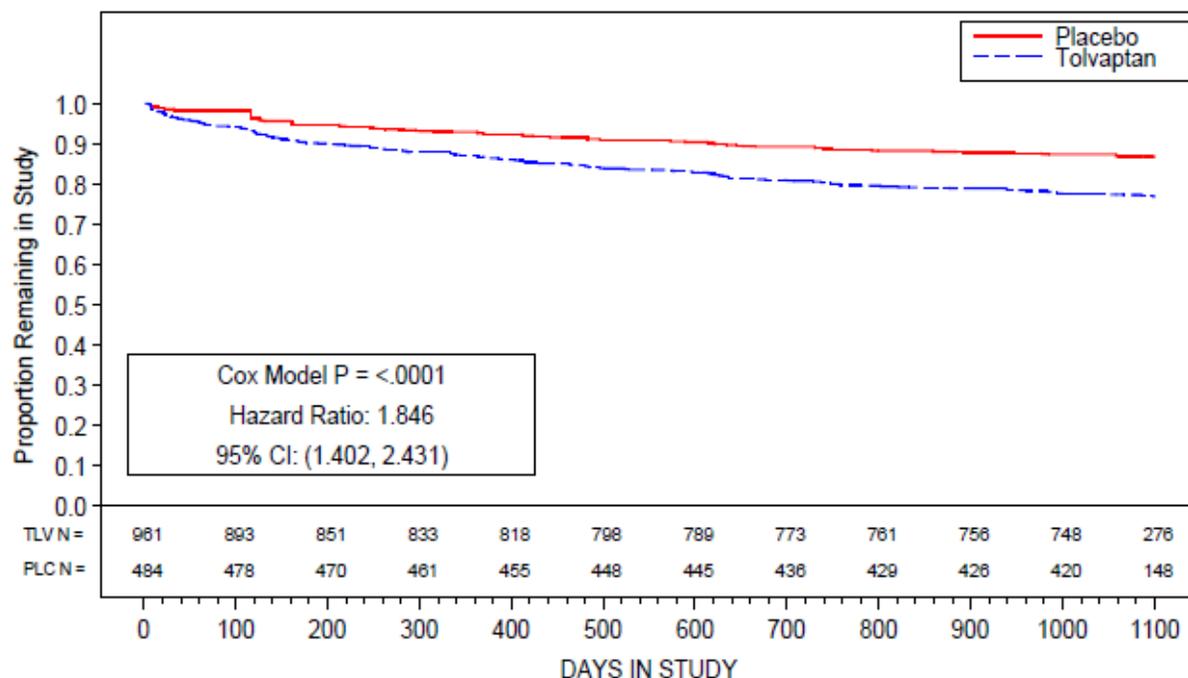


Figure 1 Kaplan-Meier plot of time to discontinuation for all reasons (Figure 8.1-1 of Study Report).

The distribution of the number of reliable eGFR measurements per subject used in the sponsor's analysis by treatment arm is shown in Table 2. In this table, only subjects and measurements used in the sponsor's longitudinal eGFR analysis (Table 9.5.1.1-1 in the Study Report) are included. This includes measurements from end of titration through month 36 for subjects with at least 4 months of follow-up and at least 2 measurements and only counting measurements labeled as reliable. There are only 10 possible visits: End of titration/week 3, Months 4, 8, 12, 16, 20, 24, 28, 32, 36. However, a few subjects had two measurements that fell within a single visit window and both measurements were included. One subject had 11 measurements included in this analysis because they had two measurements in the Month 24 window and one measurement at every other possible visit. Figure 2 shows the cumulative distribution plot of time to last eGFR used in the sponsor's analysis. It can be seen that a relatively high proportion of subjects in the tolvaptan arm were not used in the analysis at all. More than 10% of the subjects in the tolvaptan arm were not included at all and more than 20% had no measurements beyond 1 year from randomization. Table 3 shows the distribution of number of eGFR measurements. The difference between this and the previous table is that it includes measurements labeled unreliable, subjects with less than 4 months follow-up, off-treatment measurements, and measurements from subjects with only one valid measurement. Of note, the sponsor's analysis used 11,785 measurements from 1306 subjects while there were 16,197 measurements from 1445 subjects in the full dataset. If every patient randomized had 13 measurements, there would have been 18,785 measurements.

Number of observations	Tolvaptan n (%)	Placebo n (%)	Total
2	33 (4)	9 (2)	42 (3)
3	23 (3)	7 (2)	30 (2)
4	19 (2)	7 (2)	26 (2)
5	19 (2)	11 (2)	30 (2)
6	15 (2)	12 (3)	27 (2)
7	16 (2)	4 (1)	20 (2)
8	27 (3)	20 (4)	47 (4)
9	105 (12)	57 (12)	162 (12)
10	584 (69)	337 (73)	921 (71)
11	1 (0)	0	1 (0)

Table 2 Distribution of number of eGFR measurements per subject used in sponsor's analysis.

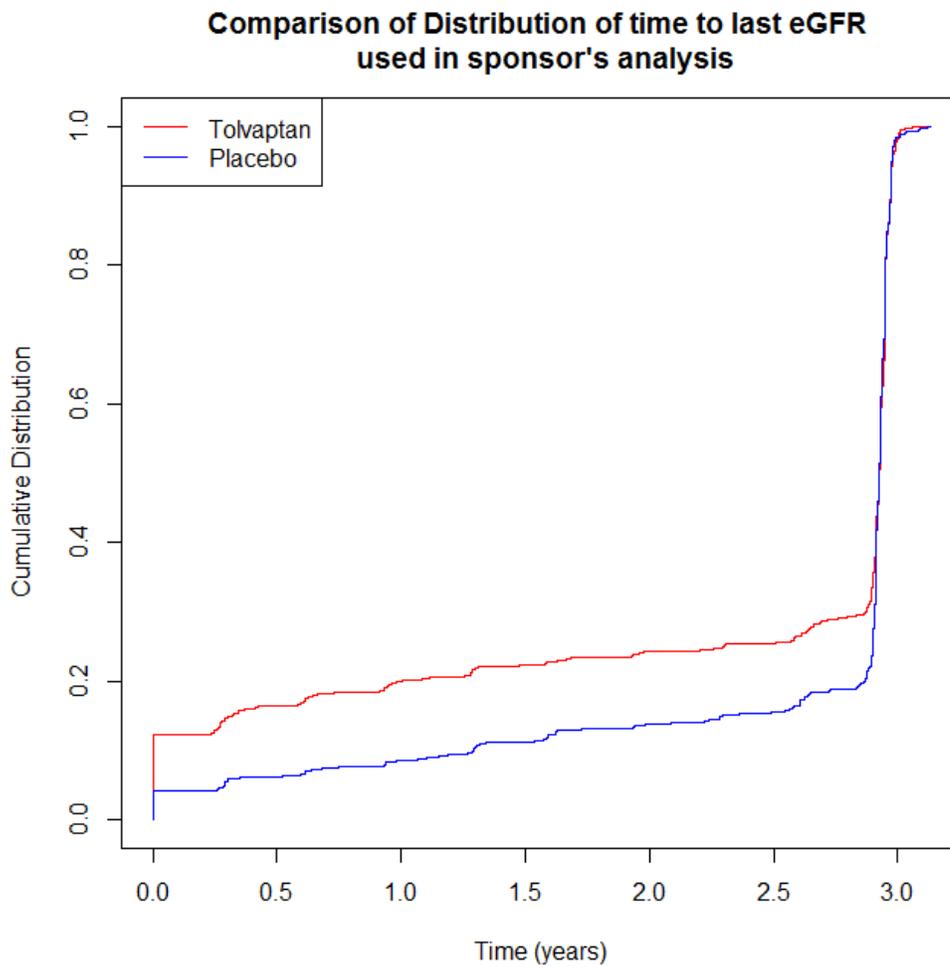


Figure 2 Kaplan-Meier plot of time to last eGFR measurement used in sponsor's analysis (Source: FDA)

Number of observations	Tolvaptan n (%)	Placebo n (%)	Total
1	7 (1)	3 (1)	10 (1)
2	45 (5)	5 (1)	50 (3)
3	56 (6)	5 (1)	61 (4)
4	23 (2)	10 (2)	33 (2)
5	19 (2)	8 (2)	27 (2)
6	22 (2)	9 (2)	31 (2)
7	13 (1)	9 (2)	22 (2)
8	13 (1)	9 (2)	22 (2)
9	7 (1)	4 (1)	11 (1)
10	13 (1)	4 (1)	17 (1)
11	23 (2)	19 (4)	42 (3)
12	110 (11)	49 (10)	159 (11)
13	609 (63)	349 (72)	958 (66)
14	1 (0)	1 (0)	2 (0)

Table 3 Distribution of number of eGFR measurements per subject actually measured.

The patient demographic characteristics are shown in Tables 2 and 3. The demographics were comparable between the two groups.

Demographic Characteristic	Tolvaptan			Placebo			Total		
	Male (N = 495)	Female (N = 466)	Total (N = 961)	Male (N = 251)	Female (N = 233)	Total (N = 484)	Male (N = 746)	Female (N = 699)	Total (N = 1445)
Age (years)									
Number of subjects	495	466	961	251	233	484	746	699	1445
Mean	38.2	38.9	38.6	38.3	39.4	38.8	38.3	39.1	38.7
SD	7.1	7.1	7.1	7.3	7.0	7.1	7.1	7.1	7.1
Median	39.0	40.0	39.0	39.0	40.0	39.0	39.0	40.0	39.0
Minimum	18	19	18	18	18	18	18	18	18
Maximum	51	50	51	50	50	50	51	50	51
Height (cm)									
Number of subjects	495	465	960	251	232	483	746	697	1443
Mean	180.4	166.2	173.5	180.0	166.6	173.6	180.3	166.4	173.6
SD	7.9	7.3	10.4	7.4	6.5	9.7	7.8	7.0	10.2
Median	180.0	166.0	173.0	180.0	167.0	173.0	180.0	167.0	173.0
Minimum	150	143	143	159	150	150	150	143	143
Maximum	210	192	210	201	188	201	210	192	210
Weight (kg)									
Number of subjects	495	466	961	251	233	484	746	699	1445
Mean	87.68	70.74	79.47	86.13	70.30	78.51	87.16	70.59	79.15
SD	15.80	16.59	18.27	16.96	16.04	18.31	16.21	16.4	18.28
Median	86.00	66.50	78.20	82.70	66.90	75.50	85.00	66.70	77.50
Minimum	50.6	40.6	40.6	54.7	46.0	46.0	50.6	40.6	40.6
Maximum	160.6	133.6	160.6	151.8	135.2	151.8	160.6	135.2	160.6
Race, ^a n (%)									
Caucasian	418 (84.4)	392 (84.1)	810 (84.3)	204 (81.3)	204 (87.6)	408 (84.3)	622 (83.4)	596 (85.3)	1218 (84.3)
Black	7 (1.4)	9 (1.9)	16 (1.7)	3 (1.2)	0	3 (0.6)	10 (1.3)	9 (1.3)	19 (1.3)
Hispanic	8 (1.6)	5 (1.1)	13 (1.4)	6 (2.4)	3 (1.3)	9 (1.9)	14 (1.9)	8 (1.1)	22 (1.5)
Asian	61 (12.3)	60 (12.9)	121 (12.6)	37 (14.7)	25 (10.7)	62 (12.8)	98 (13.1)	85 (12.2)	183 (12.7)
Other	1 (0.2)	0	1 (0.1)	1 (0.4)	1 (0.4)	2 (0.4)	2 (0.3)	1 (0.1)	3 (0.2)

Table 4 Patient demographic characteristics (Table 8.2-1 of Study Report)

Parameter	Tolvaptan (N = 961)	Placebo (N = 484)	Total (N = 1445)
1/serum creatinine ($[\text{mg/mL}]^{-1}$)			
Number of subjects	958	482	1440
Mean	102.27	104.30	102.95
SD	27.21	33.87	29.61
Median	100.00	100.00	100.00
Minimum	43.7	35.5	35.5
Maximum	263.2	500.0	500.0
eGFR_{CKD-EPI} (mL/min/1.73 m²)			
Number of subjects	958	482	1440
Mean	81.35	82.14	81.61
SD	21.02	22.73	21.60
Median	80.76	80.40	80.74
Minimum	32.3	26.4	26.4
Maximum	132.8	186.7	186.7
TKV (mL)			
Number of subjects	961	483	1444
Mean	1704.8	1667.5	1692.3
SD	921.27	873.11	905.31
Median	1456.7	1468.5	1458.8
Minimum	750.0	751.1	750.0
Maximum	7555.4	6751.1	7555.4

Table 5 Patient demographic characteristics (Table and 8.2-3 of Study Report)

1.4.4 Results and Conclusions

The drug had an effect on the primary endpoint, TKV. The effect is not linear over time, but rather there is a large initial drop in TKV in the tolvaptan arm and that difference is maintained for up to 3 years.

The sponsor used the log-transformation in their words, "to reduce heterogeneity in variance and achieve linearity over time" (Study Report). The residual variance was approximately homogeneous (see Figure 3). They were not normally distributed (skewness 2.22, excess kurtosis 6.3). In addition, $\log_{10}(\text{TKV})$ was not linear over time. One simple way to see this is to include a second degree term for time in the model (two extra fixed effects, one for each treatment group). When I did that, the log-likelihood improved by about 300 (note that an improvement of 3 in the log-likelihood with two extra parameters would be a significant improvement) and the AIC improved by almost 600.

As in the sponsor's eGFR analysis, their analysis of TKV did not include all the subjects. All 1445 subjects randomized had a baseline TKV measurement, but only 1277 were included in the sponsor's analysis.

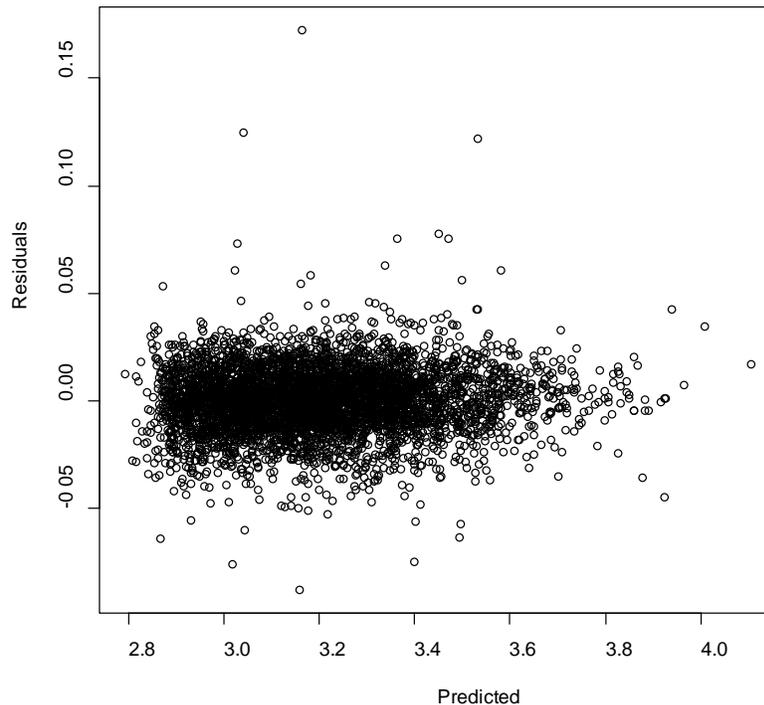


Figure 3 Scatterplot of residuals versus predicted log(TKV) from sponsor's model.

Like I did with the longitudinal change in eGFR, I tried fitting a model with an acute effect and a chronic effect on $\log_{10}(\text{TKV})$ and I included a quadratic function of time. I used all TKV measurements in the dataset. My estimated parameters were:

Intercept	3.18
time	0.0158
time ²	0.00258
acute treatment effect	-0.0223
chronic treatment*time interaction	-0.00433

Treatment*time² was not included because it did not improve the fit. The estimated standard deviations were: random intercept = 0.194, random slope = 0.0183, residual error = 0.0180. The estimated correlation between the random effects was 0.23.

According to the Study Report, the study was designed based on an assumption of a 7% annual increase in the placebo arm (this model actually estimates $10^{(0.0158+0.00258)}=1.043$, or a 4.3% rate of increase in the placebo group during the first year. The sponsor's analysis estimates a mean increase of 5.6% in the placebo group. Either way, the rate of growth was slower than expected. Also, the study design assumed a standard deviation of 0.017 for the residual error and

a standard deviation of 0.0184 for the random effect of slope (on the \log_{10} scale). Those values turned out to be almost exactly what the estimates are for those parameters in the FDA model.

Figure 4 shows the mean of observed $\log_{10}(\text{TKV})$ at each visit, as well as the predictions from the sponsor's model and the FDA model. I transformed everything back to the original scale of TKV by using the exponential function. There is one point for each treatment group at each of four visits. The x-coordinate is at the mean of the times when the observations happened and the y-coordinate is the geometric mean of the observations. The acute effect in the blue curve. It looks in the figure like the solid blue curve doesn't fit the points as well as it could, but that's OK because what the mixed effects model is doing is more complicated than just trying to come close to these group means.

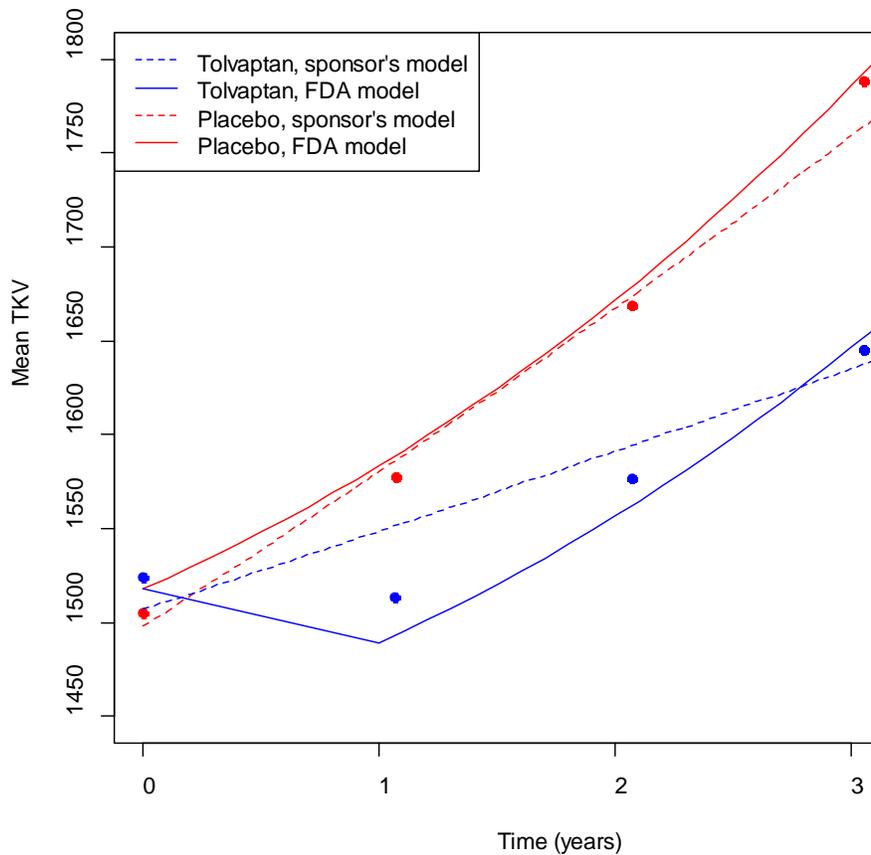


Figure 4 Mean observed and predicted TKV over time.

The trial did not confirm the drug has an effect on the key composite endpoint at a significance level of 0.01. The sponsor's analysis used the Anderson-Gill method for recurrent events. The sponsor's results were a hazard ratio estimate of 0.865 and p-value of 0.0095 using the

investigator reported events. I found the same total number of events (1049 vs. 665) in each group when I tried to repeat the sponsor's analysis and I found the same total number of years of follow-up in both groups (2387 vs. 1329). However, my estimates and p-values were slightly different (hazard ratio of 0.860 and p-value of 0.010).

The statistical issues with this analysis are three-fold. Missing data and ITT analysis, post-randomization baseline used for creatinine component, standard error is estimated under the alternative.

There were more subjects with missing values in the tolvaptan group, as discussed for other endpoints. The way that censoring was done in the sponsor's analysis used the last censoring time for all events, i.e. if there was follow-up on any of the four events, then the subject was not censored for the composite. Handling censoring for a composite endpoint with different censoring times for the components is not straightforward. The sponsor's sensitivity analyses are taken from the Study Report p. 208:

"

In response to a regulatory request to examine the effect of partial missing data on the result of the SAP-defined primary analysis of the key secondary composite endpoint, subjects could only contribute to the treatment group denominator at the last visit where an event occurred or where all 4 components were evaluated. This analysis also favored tolvaptan (HR 0.878, 95% CI 0.787 to 0.979, $p = 0.0194$) (CT-5.2.16.2).

...

Less restrictive ITT analyses were used to include data collected off treatment up to Month 36. Time to multiple composite ADPKD events analyses used a nonrestricted ITT approach (regardless of treatment period) using either predose baseline (CT-5.2.6.1; HR 0.874, 95% CI 0.784 to 0.974, $p = 0.0147$) or Week 3/EOT as baseline (CT-5.2.6.2; HR 0.889, 95% CI 0.797 to 0.992, $p = 0.0354$). Both of these analyses maintained statistical significance.

"

The use of post-randomization baseline for the definition of the creatinine event component and the subjects who dropped out in the first 4 months (and a much higher percentage in the tolvaptan group) complicate the interpretation of this analysis. There is no good way to handle this. It would be better to continue to collect data from subjects after they discontinue study drug. As long as I continue to see studies with a large amount of missing data, I think the best way to handle it is to put some kind of penalty in the analysis whereby subject from the placebo group with missing data are imputed with some kind of neutral or good value, but subjects from the treatment group are given a worse value. Because of the amount of missing data here, that kind of imputation will undoubtedly raise the p-value above 0.05.

Finally, the Anderson-Gill analysis uses a Wald-type estimate of the variance of the treatment effect estimate. That means, the variance is estimated under the alternative hypothesis. For a

clinical trial, when testing the null hypothesis, it is best to calculate the variance using the design-based method. That means, in part, that the variance should be estimated under the null hypothesis. I permuted the treatment assignments 10,000 times and found the variance of those 10,000 estimates. This does not fix the problems with missing data or anything else, it's only an attempt to find the correct variance of the estimate under the null hypothesis. That standard deviation was 0.0617 compared to the estimate in the sponsor's analysis of 0.0558. That may not seem like a big difference, but that is sufficient to change the p-value from $2\Phi(-2.57) \approx 0.010$ to $2\Phi\left(-2.57 \frac{0.0558}{0.0617}\right) \approx 0.020$.

The remainder of this section discusses changes in eGFR using the CKD-EPI equation.

The longitudinal analysis of eGFR is complicated because of acute and chronic effects. Many interventions that have effects on creatinine have different acute and chronic effects. This was anticipated and was the reason that the study was designed to have follow-up visits off treatment. The sponsor's analysis attempted to look only at the chronic effect by eliminating the measurements before titration and the measurements off treatment as well as the measurements that were labeled unreliable. However, besides throwing away a large amount of data, the sponsor's analysis had some other drawbacks. Their model assumes that eGFR changes in a linear way over time. Also, their model assumes the residual errors are independent, normally distributed, with a homogeneous variance. The data actually show that all these assumptions are false.

In the tolvaptan arm: the mean change in the 3 week titration phase was $-3.9 \text{ mL/min/1.73 m}^2$ and 71% of the subjects had a drop in eGFR. The mean change in the placebo arm was $-0.1 \text{ mL/min/1.73 m}^2$ and 47% had a drop in eGFR. These means and percentages are using the observed cases and the data from baseline and end of week 3 only (not based on any model). The estimated densities of the change in eGFR during the titration phase for both groups are shown in Figure 2.

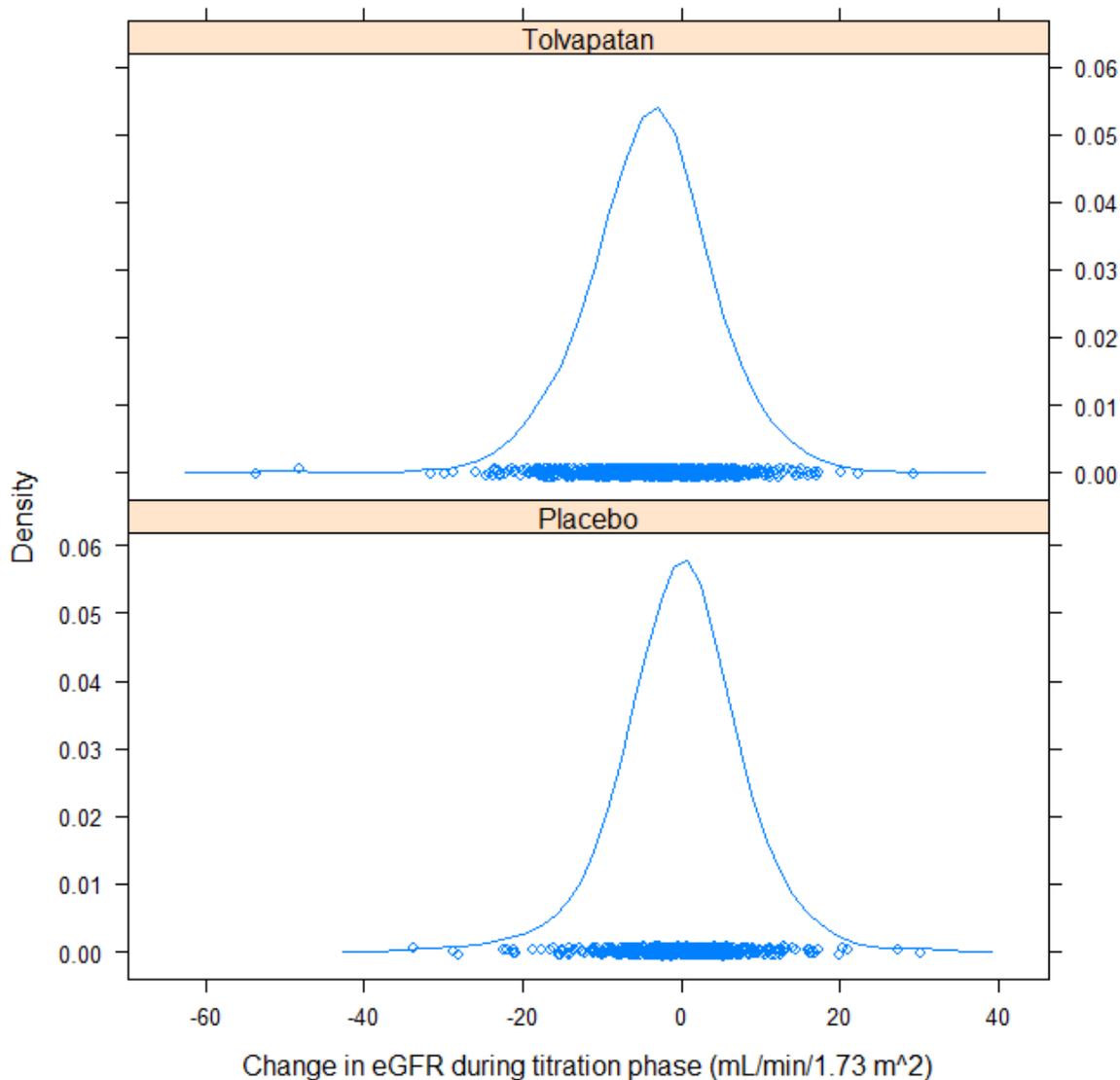


Figure 5 Estimated density of change in eGFR during initial 3 week titration period.

The sponsor's model for longitudinal changes of eGFR over time includes an intercept and terms for baseline, time, treatment group, treatment by time interaction. There are random effects within subject of intercept and time with unstructured covariance matrix. The estimates from the sponsor's model using the sponsor's data are:

Intercept	3.096
treatment	0.749
time	-3.700
baseline	0.954
treatment*time interaction	0.977

The estimated standard deviation of the random intercept is 3.227 and the estimated standard deviation of the random slope is 2.479 and their correlation is 0.663. The residual error standard deviation is 5.560.

The within subject residuals are shown in Figure 4. The red curves show the upper and lower 2.5 percentiles of the distribution as a function of the predicted value. These percentiles are estimated by quantile regression (I used the algorithm from <http://www.e-publications.org/ims/submission/index.php/AOAS/user/submissionFile/4295?confirm=37ca4b7>) and give some sense of whether the variance is homogeneous. In addition, one can divide the graph into 5 parts from left to right with equal number of points in each part and then calculate the sample variance of the residuals in each of the 5 sections. Doing that, I found variances (from left to right) of 11.5, 19.6, 39.9, 32.8, and 32.3. The three on the right are all significantly larger than the two on the left using the F-test for the ratio of the variances. Therefore, the variance is not homogeneous. Figure 5 shows the normal probability plot for the residuals which confirms they are not normal.

To investigate the linearity assumption, one way is to fit a more complicated model and compare the AIC and/or the likelihood ratio if the models are nested. For example, I tried a slightly more complicated model that includes a quadratic term for time and the interaction with treatment (same random effects as before). This more complicated model (with 2 additional parameters) fits the data better than the linear time model; the AIC improves by 40, minus $2 \cdot \log$ -likelihood ratio is 44, which has a p-value of close to 10^{-10} . Also, the model using $\log(\text{eGFR})$ as the response and replaces the covariate baseline by $\log(\text{baseline})$ fits the data better. It is more complicated to compare these two models and it cannot be done by comparing AIC or likelihood ratios. Instead, to account for the transformation, we have to add the sum of the $\log(\text{eGFR})$ to the likelihood in the first model to compare it with the likelihood of the second model. After accounting for the transformation of the response variable, the log-likelihood of the second model is larger by almost 71. The models have the same number of parameters and clearly the second model (using $\log(\text{eGFR})$) fits much better and so is the preferred model between the two.

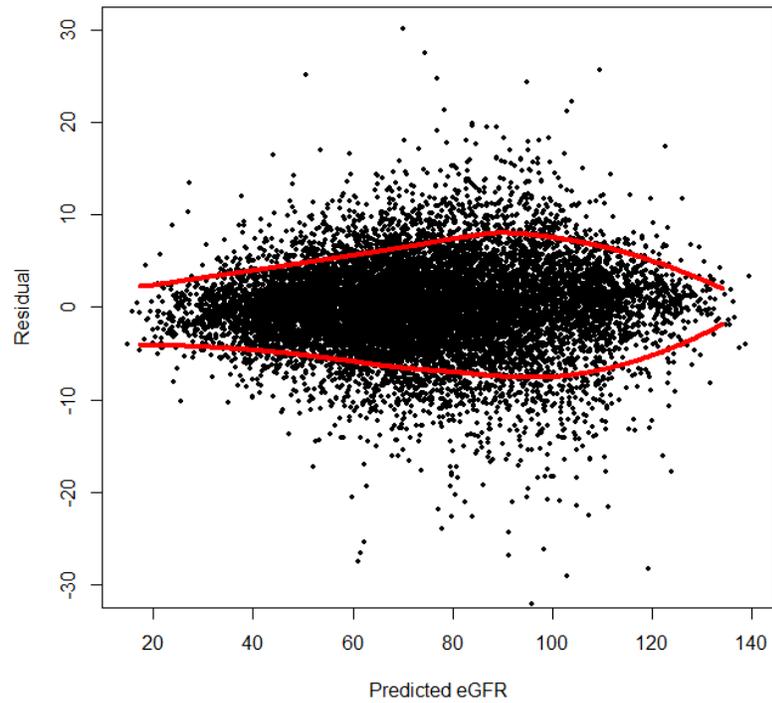


Figure 6 Residuals versus predicted scatterplot from sponsor's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 14 residuals with magnitude larger than 30 not shown.

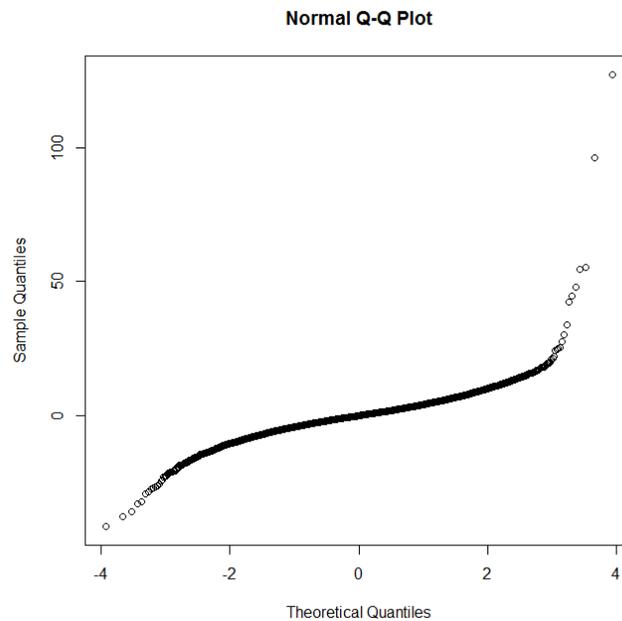


Figure 7 Normal probability plot for residuals from sponsor's eGFR model.

I tried to build a model I thought was reasonable for eGFR that: a) uses all of the measurements and b) accounts for possible acute and chronic effects. Using the log-transformation makes more sense over a long period of time if for no other reason because using a straight line without the transformation will eventually cross into the region where y is negative, but negative values of eGFR are not possible. Since using $\log(\text{eGFR})$ fit the data used in the sponsor's analysis better than eGFR confirms this intuition, I used that as a starting point for a model. Next, I included terms for acute drop of eGFR at the start of treatment and for acute rise of eGFR after stopping treatment. Finally, I considered other covariates, but I found that only baseline eGFR (at randomization), baseline $\log(\text{TKV})$, and age improved the fit significantly among the covariates I tried. Five people had missing baseline eGFR. Since I used baseline $\log(\text{eGFR})$ as a covariate in the model I needed to impute values for those 5 subjects. I tried values that were the same as the subjects' observed data at a nearby timepoint and I also tried other values that were biased against any treatment effect (adding 10 to the reasonable baseline for the two placebo subjects and subtracting 10 to the reasonable baseline from the 3 tolvaptan subjects to make it appear tolvaptan was not effective). However, the estimates in the model were essentially identical in both imputations.

The estimated fixed effects coefficients are:

Intercept	0.0852
$\log(\text{baseline TKV})$	-0.0382
$\log(\text{baseline eGFR})$	0.884
age	-0.00172
time	-0.358
treatment*time interaction	0.0204
$\log(\text{baseline TKV}) \cdot \text{time}$	-0.0205
$\log(\text{baseline eGFR}) \cdot \text{time}$	0.102
acute treatment effect at start	-0.0458
acute effect of withdrawal	0.0415

The estimated standard deviation of the random intercept is 0.0882 and the estimated standard deviation of the random slope is 0.0479 and their correlation is -0.052. The residual error standard deviation is 0.0804.

The residuals from this model are shown in Figure 6. The variance looks homogeneous up to the predicted $\log\text{-eGFR}$ of about 4.5 (eGFR of about 90). The normal probability plot shown in Figure 7 demonstrates that the residuals are not normally distributed. See the appendix for more details about this model including the distribution of the residual errors and the random effects. Also, see the appendix for examples of predictions of GFR for individual subjects based on this model and future predictions for the population based on this model.

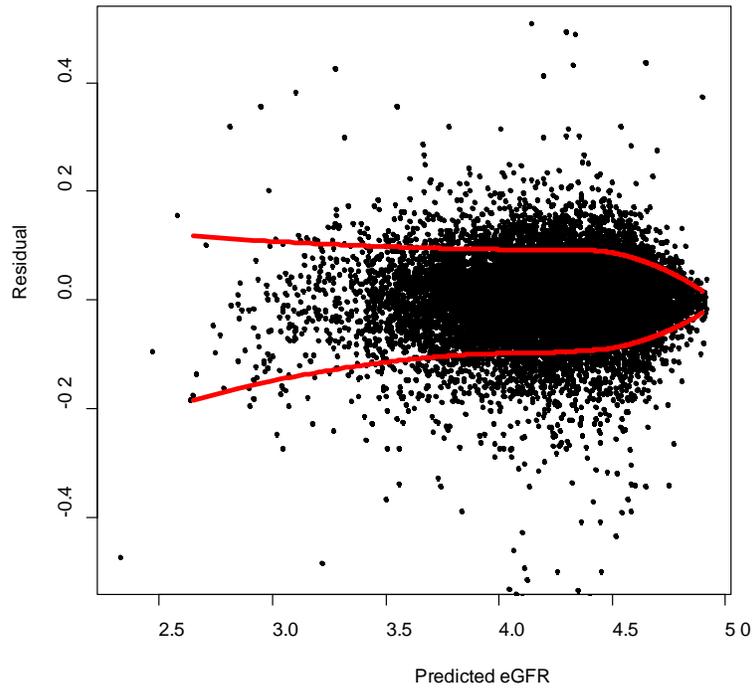


Figure 8 Residuals versus predicted scatterplot from FDA's eGFR analysis. Red curves are the estimated upper and lower 2.5 percentiles of the distribution. 19 residuals with magnitude greater than 0.5 not shown.

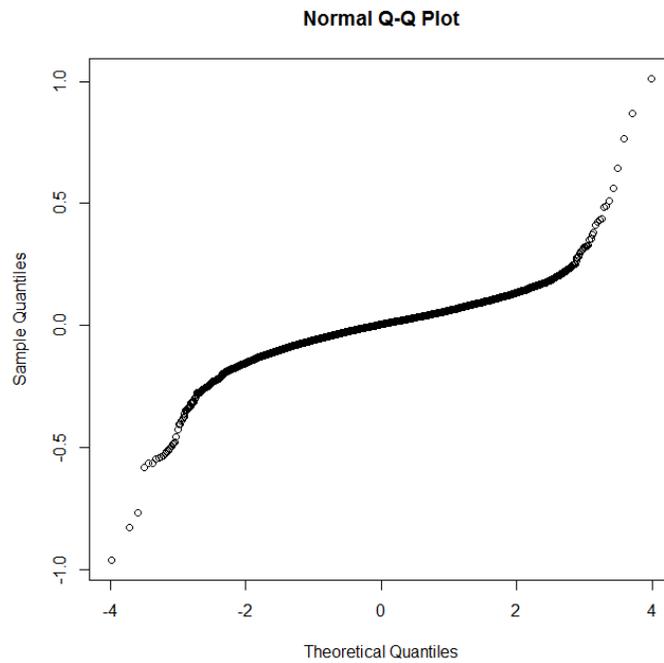


Figure 9 Normal probability plot for residuals from FDA's log-eGFR model.

1.5 Evaluation of Safety

See clinical review.

1.6 Benefit-Risk Assessment (Optional)

See clinical review.

FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

1.7 Gender, Race, Age, and Geographic Region

The sponsor's results for the key secondary endpoint are shown in Figure 8.

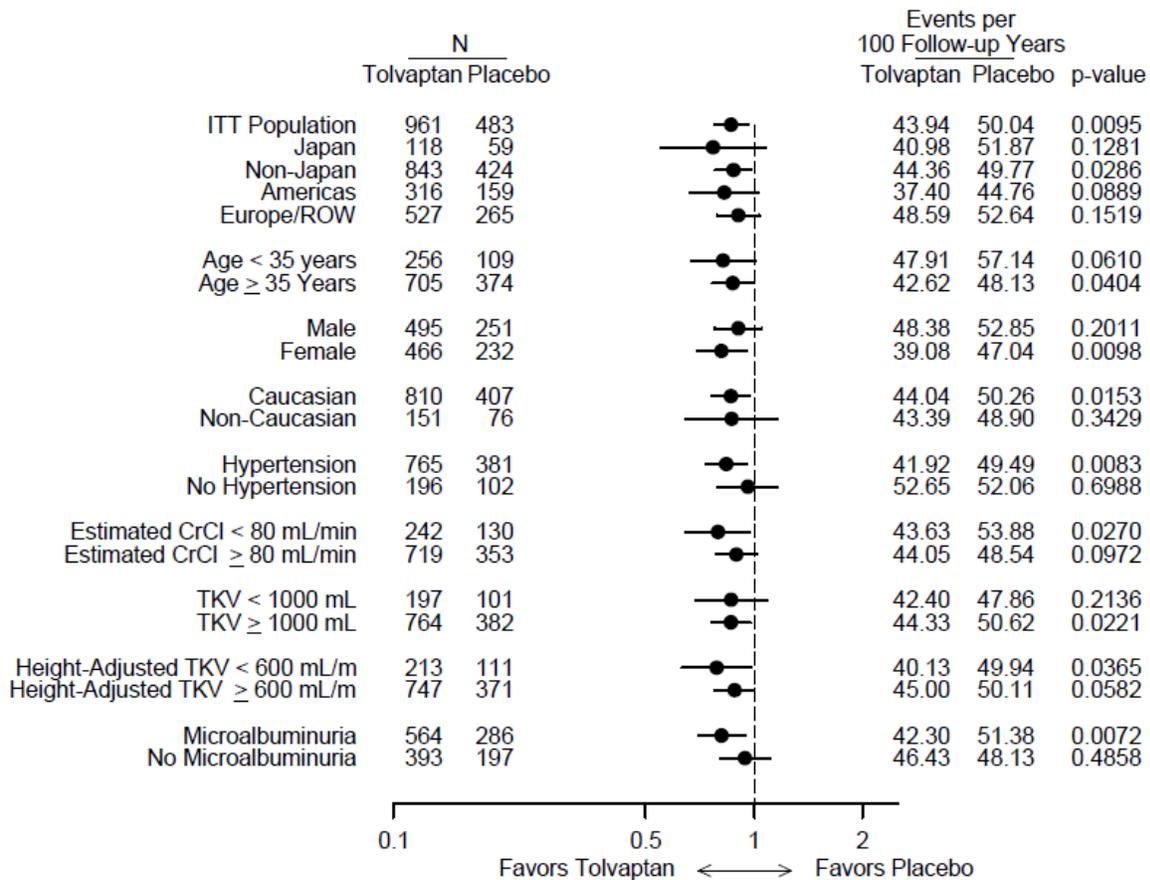


Figure 10 Sponsor's results of key secondary endpoint within subgroups (Study Report Figure 9.4.2-1)

1.8 Other Special/Subgroup Populations

About 4/5 of the subjects were taking an ACE inhibitor or ARB at randomization. The subjects taking those drugs had lower starting eGFR and higher TKV on average (76.4 vs. 89.8 mL/min/1.73 m² and 1598 vs. 1200 mL).

In the subgroup not taking ACEi/ARB, the average number of years of follow-up per subject were 2.12 years (tolvaptan) and 2.65 years (placebo). There were 41.0 events per 100 follow-up years (tolvaptan) and 46.6 events/100 follow-up years (placebo). The estimated hazard ratio for the key secondary endpoint was 0.82 in this subgroup.

In the subgroup taking ACEi/ARB, the average number of years of follow-up per subject were 2.57 years (tolvaptan) and 2.77 years (placebo). There were 44.5 events per 100 follow-up years (tolvaptan) and 50.7 events/100 follow-up years (placebo). The estimated hazard ratio for the key secondary endpoint was 0.86 in this subgroup.

SUMMARY AND CONCLUSIONS

1.9 Statistical Issues

The trial should have been planned with a type 1 error rate of 0.01 (two-sided) for a clinically meaningful endpoint, but was not.

Although the trial was technically blinded, the treatment assignment could have been guessed from effects on dehydration and water intake.

There were a high percentage of dropouts, particularly in the tolvaptan arm. Missing values were not imputed, but many subjects were not included at all in the sponsor's analyses. Other subjects were included with missing values but that is always raises problems, even without imputation.

Endpoints that used change in eGFR defined the baseline using a post-randomization value (post-titration) and a high percentage of subjects (particularly from the tolvaptan arm) had no post-titration value.

The analyses used assumptions that in some cases could be shown false with the data.

1.10 Collective Evidence

There was only one phase 3 trial in the submission.

1.11 Conclusions and Recommendations

The results on the clinical composite endpoint from the phase 3 trial, based on the sponsor's analysis, are just below the level they were told would be needed for approval ($p=0.0095$ when they were told they needed $p<0.01$ for approval). There is a large amount of missing data and use of a post-randomization baseline for change in eGFR. The Anderson-Gill method for recurrent events analysis estimates the variance under the alternative hypothesis. If we do nothing about the missing data or the post-randomization baseline, but just replace the variance estimate with an estimate under the null hypothesis, the p-value from the recurrent events analysis is 0.02. Other analyses by the sponsor of the eGFR and TKV endpoints have the same problems related to missing data, but also use unverified model assumptions and in some cases use assumptions that can be demonstrated to be false.

1.12 Labeling Recommendations (as applicable)

NA.

APPENDICES

Distribution of residuals from FDA model of eGFR

The normal probability plot and any test of normality (Anderson-Darling, etc.) show the residuals are not normally distributed. The skewness is 2.88 and the excess kurtosis is 13.9. The empirical cumulative distribution function is shown in Figure A1. Also, the figure shows best fitting normal and Laplace distribution with parameters estimated by maximum likelihood. Neither fits very well, but I believe the Laplace distribution fits a little better. It is not easy to fit mixed effects models outside of the common assumptions of normally distributed errors. However, I think it may still be useful as far as modeling the mean true GFR, at least within the range of the time frame of 3 years from baseline. It may or may not be a reasonable model for extrapolation beyond 3 years.

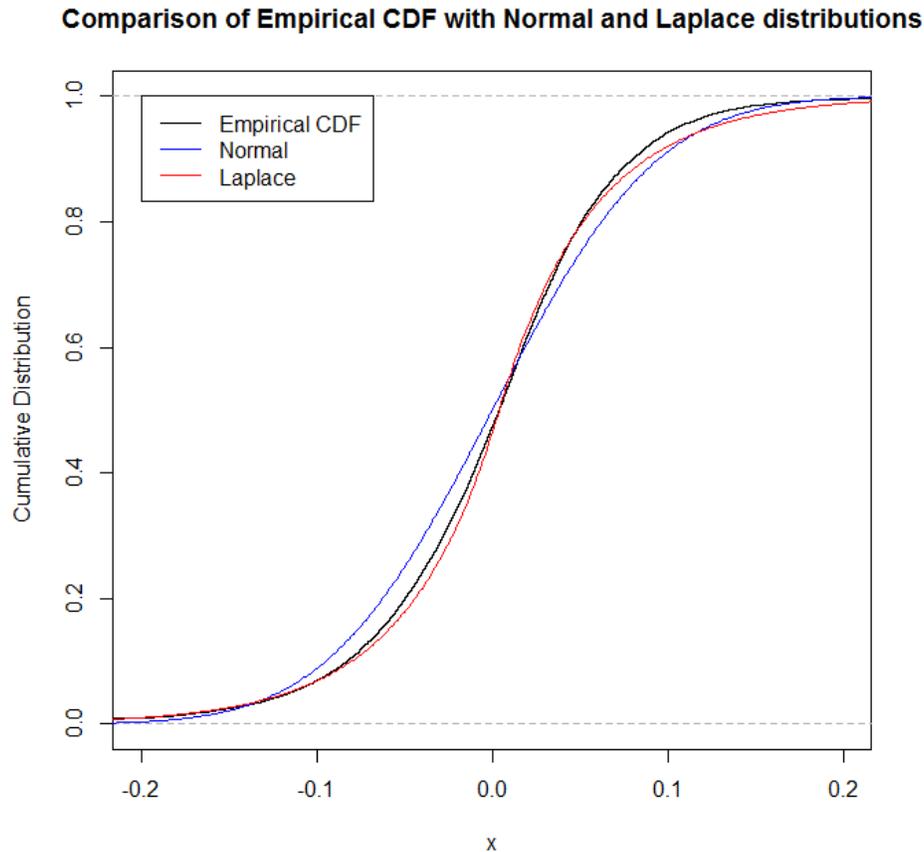


Figure A1 Comparison of Empirical CDF of within subject residuals from model described in the appendix with Normal distribution and Laplace distribution distributions with maximum likelihood estimates of parameters.

Distribution of the random effects

The estimated random effects are also not normally distributed. The scatterplot of the bivariate random effects is shown in Figure A2. Also, some of the estimated slopes are positive. The estimated slopes depend on the random effect for slope, but also on baseline TKV and eGFR. In more than 100 subjects, the estimated slope is still positive after subtracting the estimated chronic effect of treatment on the slope (the estimated slope would be positive for those subjects with no treatment). This doesn't seem to be biologically possible.

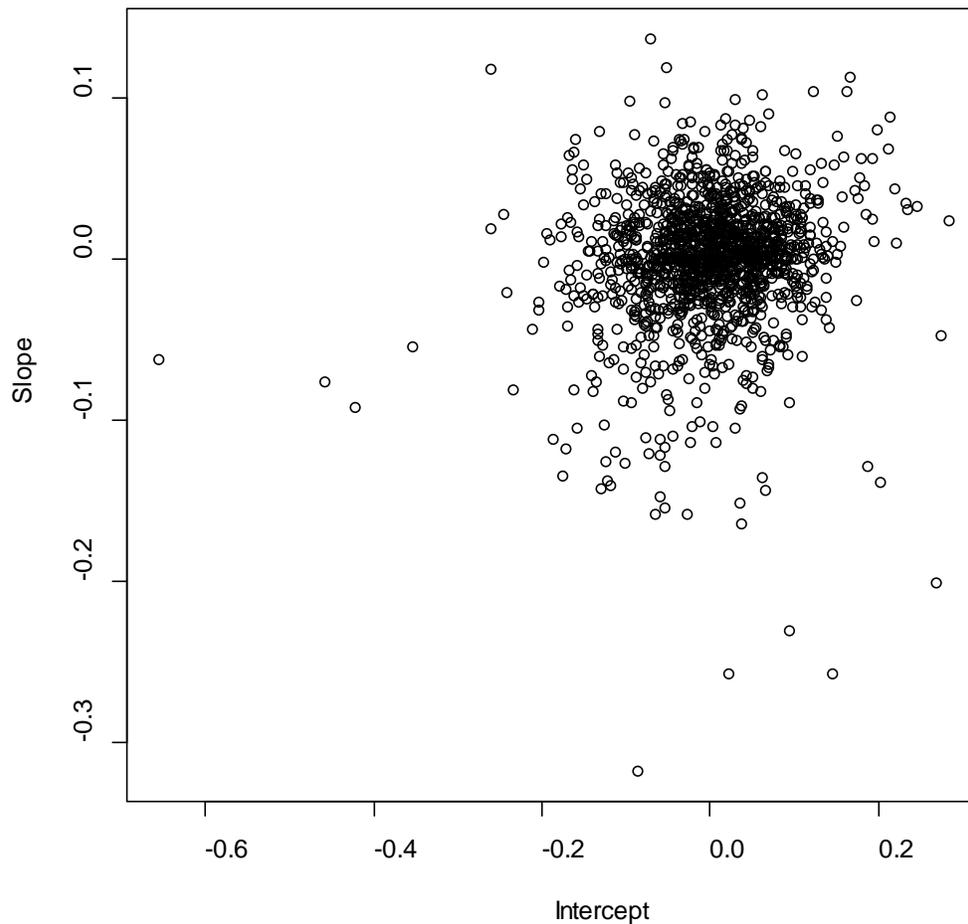


Figure A2 Scatterplot of estimated random intercept and random slope effects.

Model for eGFR between 30 and 45 mL/min/1.73 m²

The function $f(x)$ is the log-eGFR at time x , $y_1 = \log(45)$, x_1 is the time when the log-eGFR is y_1 , d_1 is the slope before reaching a eGFR of 45, $y_2 = \log(30)$, a is the acute effect of the drug withdrawal, d_2 is the slope after reaching the eGFR of 30. We want the acute treatment effect

and the chronic treatment effect to both disappear in a uniform way during the time interval between the eGFR of 30 and 45.

Problem. Suppose that $x_1, y_1, d_1, y_2, a,$ and d_2 are given and that $f(x_1) = y_1, f'(x) = d_1$ for x in a neighborhood to the left of x_1 . Can we define a continuous extension of f onto the interval $(x_1, x_2]$ for some $x_2 > x_1$ such that the following two conditions hold: i) $f(x_2) = y_2$, and ii) $f'(x) = d_1 + (x - x_1) \frac{d_2 - d_1}{x_2 - x_1} + \frac{a}{x_2 - x_1}$ for all $x \in (x_1, x_2)$?

Solution. By taking the anti-derivative of both sides of the equation in the second condition, we find

$$f(x) = c_0 + \left(\frac{a - d_2 x_1 + d_1 x_2}{x_2 - x_1} \right) x + \left(\frac{d_2 - d_1}{x_2 - x_1} \right) \frac{1}{2} x^2$$

Now, use the conditions $f(x_1) = y_1$ and $f(x_2) = y_2$ and solve those two equations simultaneously for the two unknowns c_0 and x_2 to find

$$x_2 = x_1 - \frac{2(a + y_1 - y_2)}{d_1 + d_2}$$

and

$$c_0 = \frac{(d_1 + d_2)x_1(2a + (d_1 - d_2)x_1)}{4(a + y_1 - y_2)} - d_1 x_1 + y_1$$

Examples

Start with one example from the dataset, the first subject in the dataset. This subject was 46 years old with a baseline TKV of 2343.9, baseline eGFR of 70.0 mL/min/1.73 m² and was randomized to tolvaptan. He completed the trial and had 13 total eGFR measured including both follow-up visits. Those two follow-up visits are included in the Figure A3 below using filled circles. There are three scenarios shown, one (in red) assumes he never took the drug, the second (in blue) is where tolvaptan is assumed to always have the same effect. In those first two scenarios, log-eGFR after baseline is a straight line with a constant slope, but the slope is different in the two scenarios. The actual slopes (for log-eGFR) in those two scenarios are estimated from the mixed effects model. The third scenario is shown in brown. This follows the blue curve exactly until GFR hits 45, then uses the solution to the equation shown above for times between GFR of 45 and 30, then has a constant slope identical to the slope of the red curve (on the log scale). It can be seen that during this time period of losing drug effects, the recapture of the acute effect makes the brown curve rise above the blue curve, but later, the blue curve is on top again.

The predicted eGFR shown on the y-axis is a prediction in this sense. For log-eGFR, the prediction is the expected value of an observation at that time point assuming the model with the estimated parameters and the empirical Bayes estimate of the random effects for this subject. It is the mean and median of an observation at that time with those assumptions. I transformed this prediction to the original scale of eGFR by evaluating the exponential function at that prediction. This is no longer the expected value of an observation on the original scale, but it is the median of the distribution of those values. Other ways of handling the transformation in the prediction

may be better. As already noted, the residuals on the log-scale are not normally distributed or even symmetric, so methods based on that assumption might not be adequate.

46 yr old, baseline TKV = 2343.9, 1

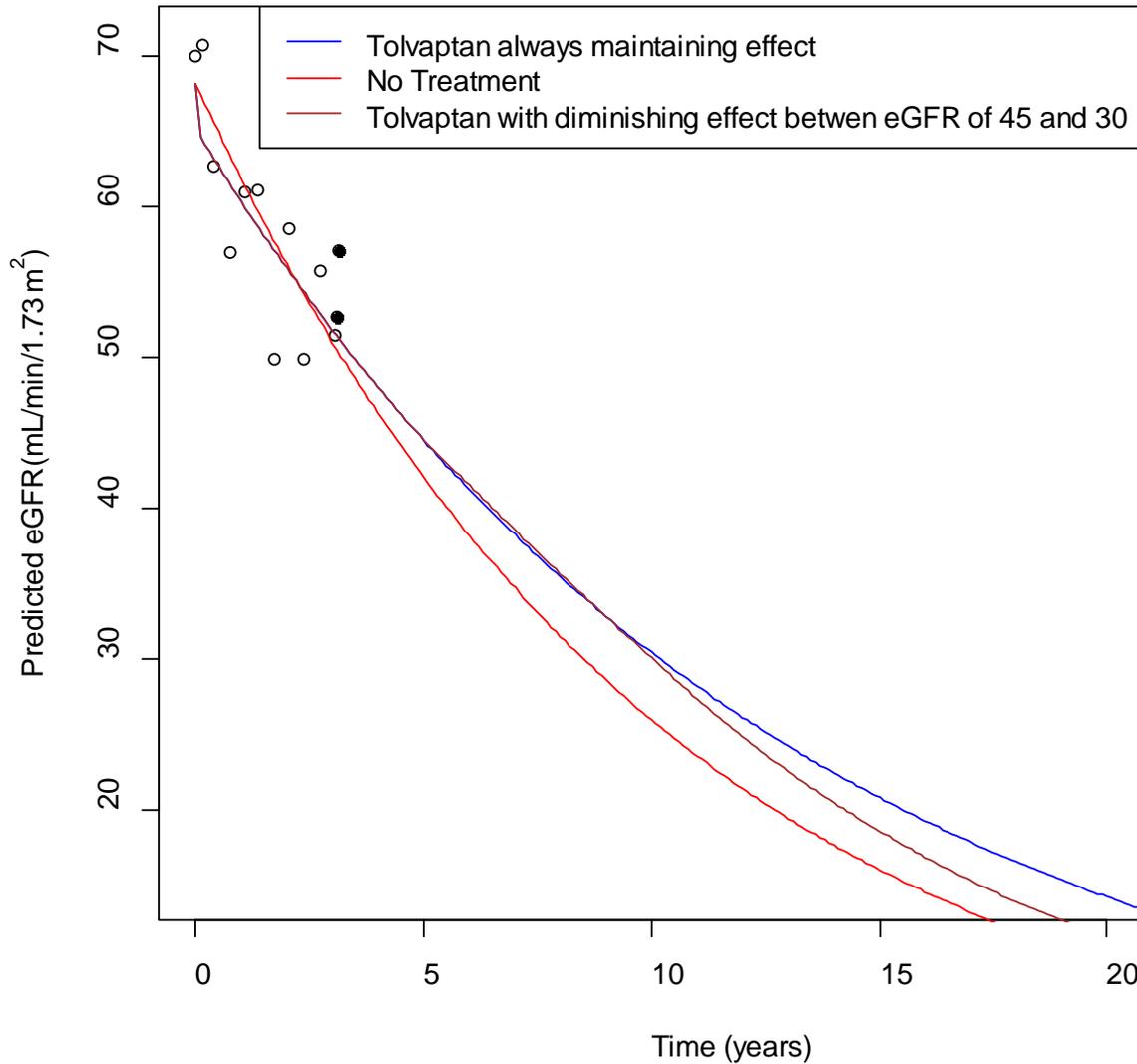


Figure A3 Predicted eGFR for one subject.

A second example from the dataset, a 20 year old subject with baseline eGFR of 88.3 and TKV of 5546.1 is shown in Figure A4. For almost the entire time shown in the figure, the brown curve is on top of the blue curve (the blue curve is there, but hidden under the brown curve). This figure shows that the model predicts an enormous benefit in terms of delaying ESRD. In fact, even reaching a GFR of 60 (CKD stage 3) is delayed by the treatment for a good 15 years in this figure. However, the figure also shows that we only have data for a short period of time. It is left

to the reader to decide how much faith to put into extrapolations of treatment effects 40 years into the future even if they are totally convinced that there is a beneficial effect on eGFR over 3 years.

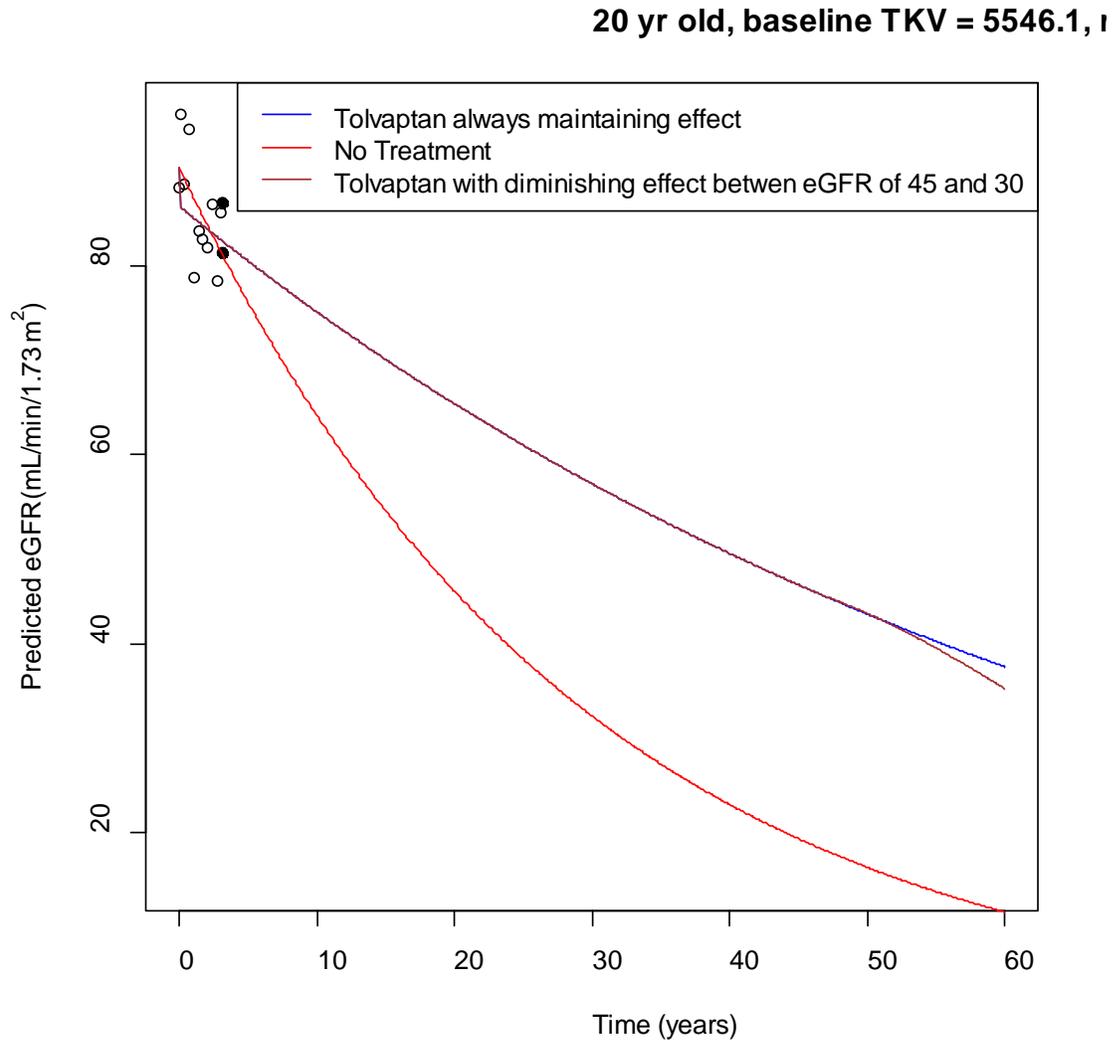


Figure A4 Predicted eGFR for one subject.

Using the model to predict into the future

Taking the subjects in the trial and using their age, baseline TKV, baseline eGFR and estimated random effects, the model allows us to extrapolate into the future and also allows us to estimate what would happen for each subject if they took tolvaptan or did not. We can then estimate for each time point in the future what proportion of subjects would have GFR < 15 mL/min/1.73 m² and compare those proportions assuming all the subjects took tolvaptan and had that estimated

treatment effect versus not taking tolvaptan. These are big assumptions about what will happen in the future. Figure A5 shows that the treatment effect could be somewhere around a 4 year delay in the time to $GFR < 15 \text{ mL/min/1.73 m}^2$.

To be more specific, start with the estimated coefficients in the model:

Intercept	0.0852
log(baseline TKV)	-0.0382
log(baseline eGFR)	0.884
age	-0.00172
time	-0.358
treatment*time interaction	0.0204
log(baseline TKV)*time	-0.0205
log(baseline eGFR)*time	0.102
acute treatment effect at start	-0.0458
acute effect of withdrawal	0.0415

For subject i , let α_i and β_i be their estimated random effects. If they took no drug, the predicted log-GFR at time t years from randomization is:

$$\log(GFR(t)) = 0.0852 + \alpha_i - 0.0382 \log(\text{baseline TKV}_i) + 0.884 \log(\text{baseline eGFR}_i) - 0.00172 \text{ age}_i + \{-0.358 + \beta_i - 0.0204 \log(\text{baseline TKV}_i) + 0.102 \log(\text{baseline eGFR}_i)\} * t$$

and their estimated time when their GFR is 15 is:

$$\tau_i = \{\log(15) - (0.0852 + \alpha_i - 0.0382 \log(\text{baseline TKV}_i) + 0.884 \log(\text{baseline eGFR}_i) - 0.00172 \text{ age}_i)\} / \{-0.358 + \beta_i - 0.0204 \log(\text{baseline TKV}_i) + 0.102 \log(\text{baseline eGFR}_i)\}$$

The estimated proportion of subjects with $GFR < 15$ at time t is then

$$\frac{1}{n} \sum_{i=1}^n I(\tau_i < t)$$

The red curve in the figure is a graph of this for t between 0 and 40.

The blue curve is more complicated because there is no fixed τ_i for each subject assuming they take the drug. The time to reach $GFR < 15$ depends now on how long they take the drug, which is a random variable. I assumed the time to withdrawal had a constant hazard in the first 4 months and another different constant hazard beyond 4 months. The hazards were defined to make it so they had a 10% chance of withdrawal during the first 4 months and, if they passed that point, a 5% chance of withdrawal each year thereafter.

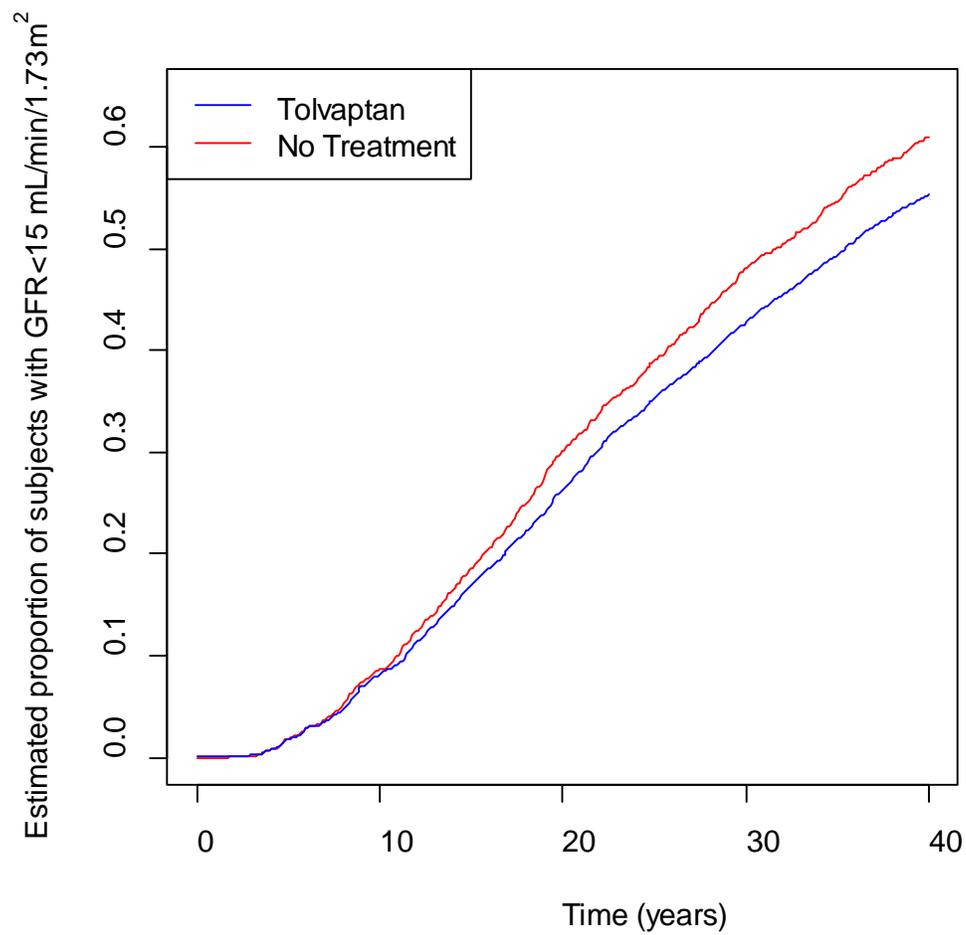


Figure A5 Estimated proportion of subjects with $\text{GFR} < 15 \text{ mL/min/1.73 m}^2$ using FDA's model and extrapolating from 3 years of data into 40 years into the future.

The next figure is based on a similar model, but assumes that after 3 years there is no treatment effect.

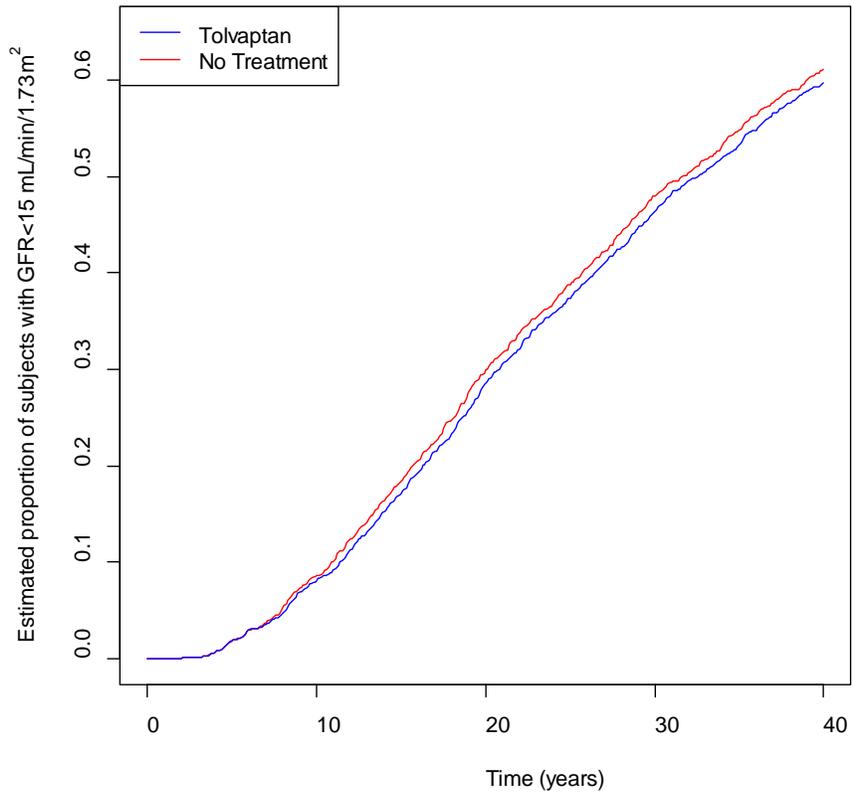


Figure A6 Estimated proportion of subjects with $\text{GFR} < 15 \text{ mL/min/1.73 m}^2$ using FDA's model, assuming no treatment effect beyond 3 years.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

JOHN P LAWRENCE
06/25/2013

HSIEN MING J HUNG
06/25/2013