

Using individual level Real World Data to inform disease area data models employed for clinical trial simulation and design

Presented by:

Sheila McCartan

Data Strategy Team Lead

sheila.mccartan@exploristics.com

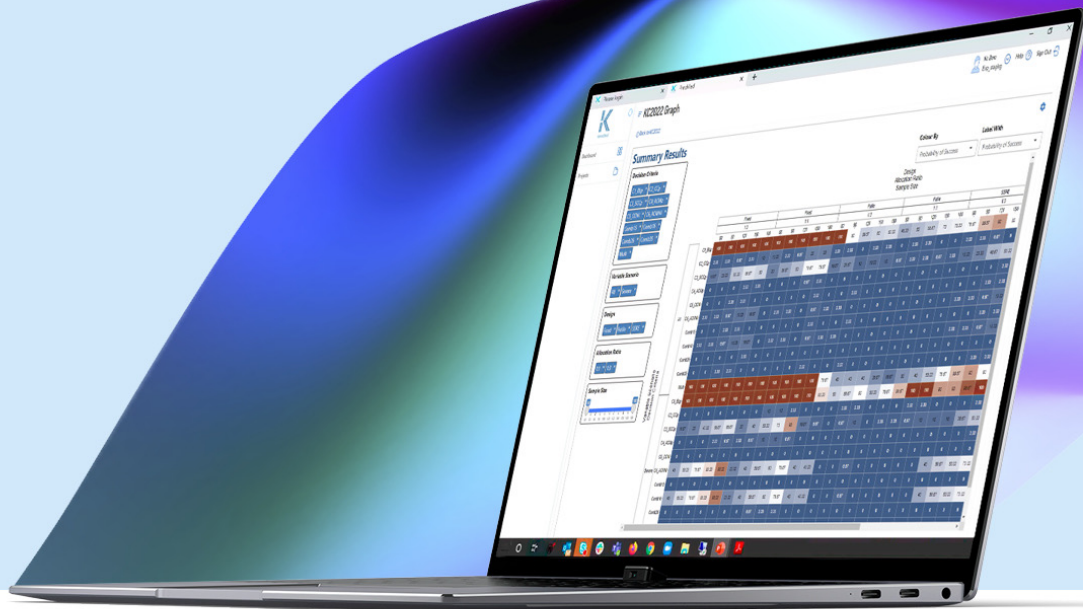
Work by:

Data Strategy Team: Antonia Popescu, Aisling Kearney, Oliver Dewhirst, Sheila McCartan

Strategic Consulting: Alecia Nickless, Andrew Mills



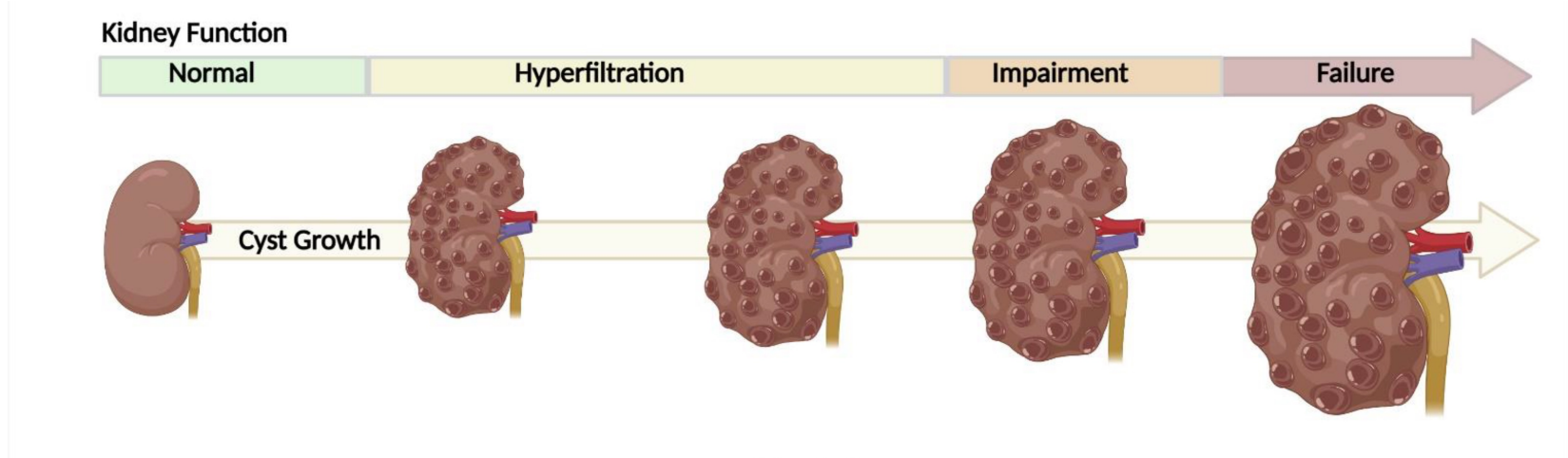
The Disease, the Data Source & the Dataset



exploristics

INNOVATIVE ANALYTICS SOLUTIONS

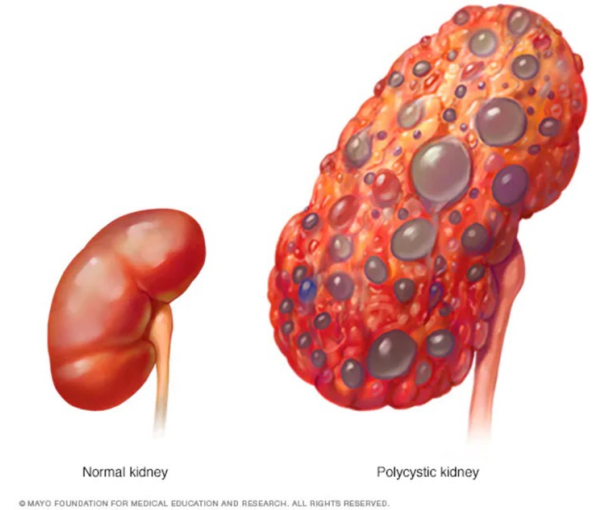
The Disease: Autosomal Dominant Polycystic Kidney Disease (ADPKD)



- Cysts develop in kidneys and effect kidney function
- Genetic disease
- Progresses through the individuals lifetime
- Can result in renal failure and death in absence of renal replacement therapy (RTT)

Clinical Presentation & Prognosis

- Pain in back, side and abdomen
- High blood pressure (hypertension)
- Blood in the urine (haematuria)
- Potentially serious upper urinary tract infections (UTIs)
- Kidney stones
- Headaches
- Kidney function will gradually deteriorate until so much is lost that kidney failure occurs.



Increasing cyst size causes a gradual loss in kidney function, that can result in **End Stage Renal Disease (ESRD)**, requiring dialysis or kidney transplantation.

Outlook is variable - can have kidney failure soon after diagnosis, others may have functional kidneys for most of their life.

On average, around half of people with ADPKD require treatment for kidney failure by the time they're 60.

Current standard of care: Tolvaptan, a vasopressin V₂-receptor antagonist.

The ADPKD population

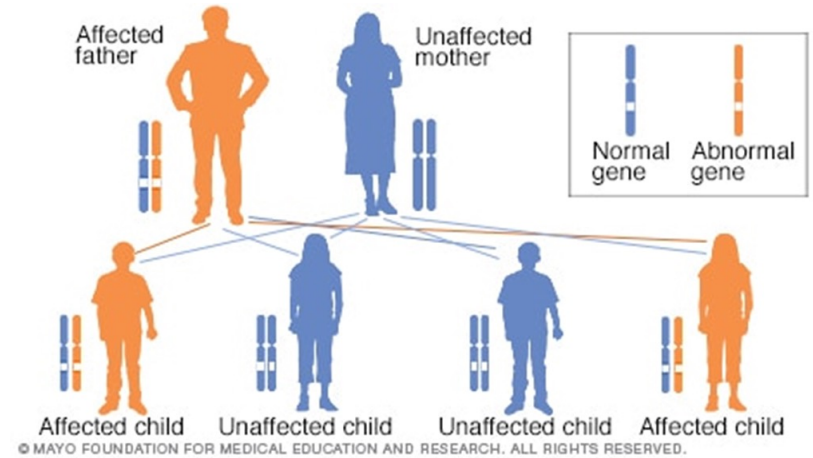
- Genetically inherited disease, caused by mutations in *PKD1* or *PKD2*.
- Effects more than 10 million people globally.
- Effects all ethnic groups, with higher prevalence (per 100,000) in Black (73.0) and White (63.2) individuals compared to Asian (48.9) and Hispanic individuals (39.9; Aung *et al*, 2021).
- Women are diagnosed at an earlier age (Willey *et al*, 2019).

The incidence within the 20- to 34-year age group was 6.19 for females, versus 3.52 per 100,000 for males.

Males are more likely to be diagnosed at older ages; the incidence is 18% higher in males aged 50–64 years than in females of the same age and 80% higher in males aged 65 years or older than in older females.

- Men have higher rates of End Stage Kidney Disease (Bergmann *et al*, 2018).

were receiving RRT¹⁴. ESRD occurs as a result of ADPKD in up to 75% of patients by 70 years of age^{14,19–21}. Age-adjusted sex ratios suggest that ADPKD is more progressive in men than in women (1.2–1.3 men to women) on the basis of rates of ESRD. A Canadian

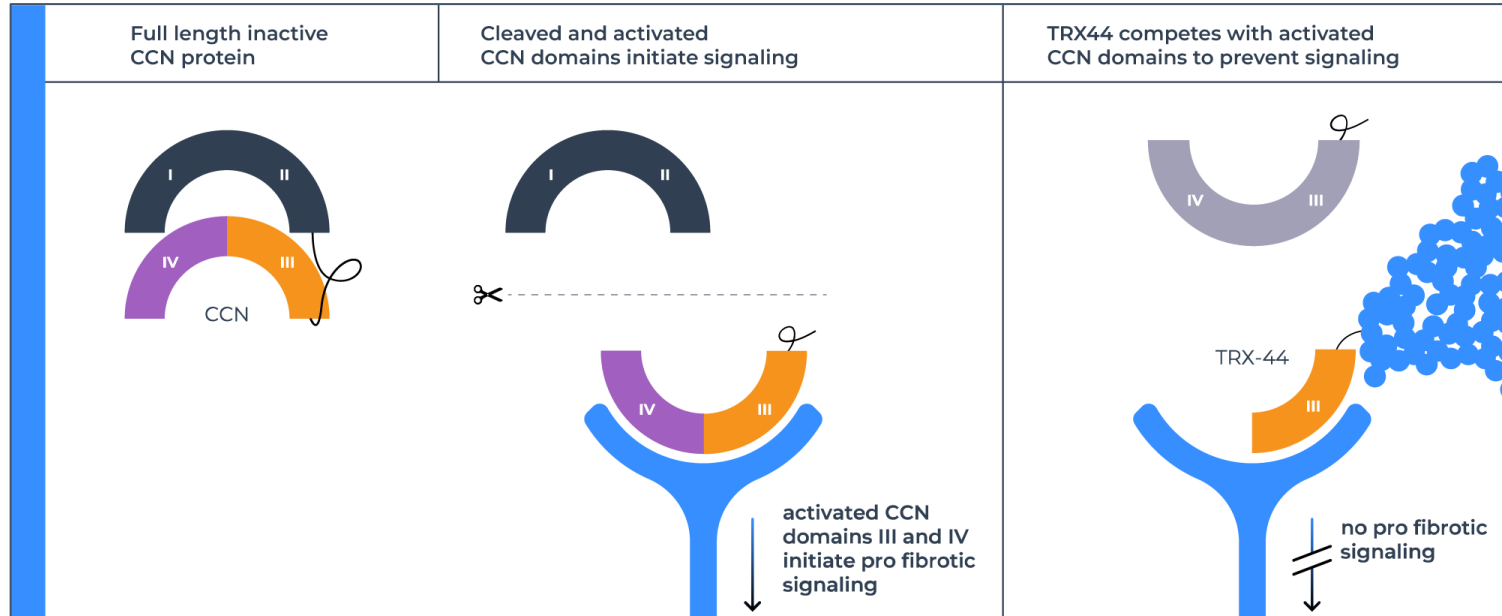


The Data Providers



Tribune Therapeutics.

A preclinical biopharmaceutical company, founding in 2020, with aim of building a pipeline of novel medicines to treat fibrotic diseases, founded on research from the Dr. Håvard Attramadal lab on CCN signalling.



The Data Source

C-Path's polycystic kidney disease (PKD) Outcomes Consortium.

Collaboration between C-Path, leading academic medical centres, pharmaceutical companies, patient organisations and international regulatory agencies.

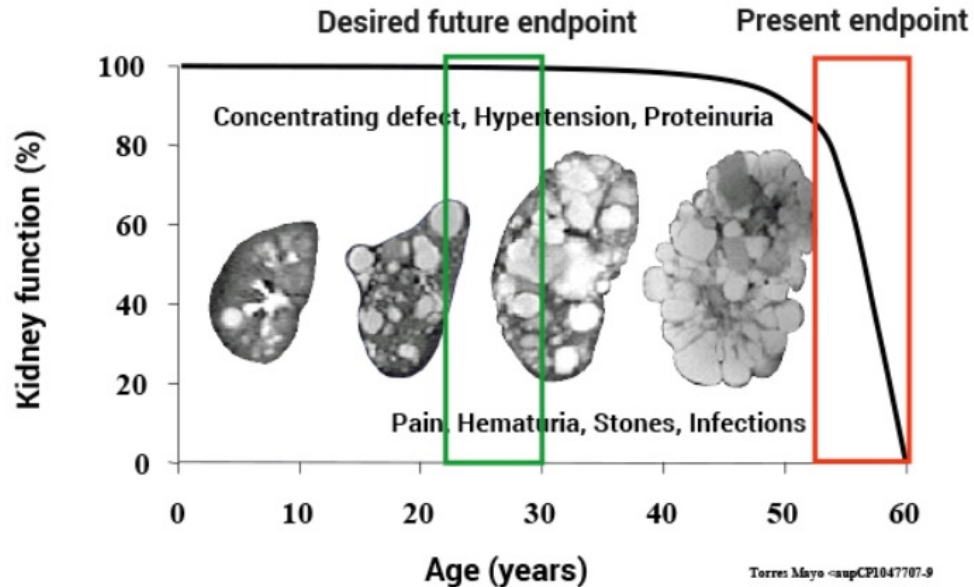


Fig.1: Increase in kidney size and change in kidney function with age



Mission: Develop tools and processes and promote research that will lead to development and regulatory approval of new treatments for PKD, and improve lives it affects.

C-Path data provided by:



The Dataset



N=2498 patient entries from 3 registries:
I0-003385 (N=1010); PKD-EM (N=376) & UCPKDI (N=1112).

Data from 2498 subjects with ages ranging from 0 to 84 years at study entry (mean 35.9; median 37).
Majority (81%) of the subjects are Caucasian, and 59% are female.

SDTM format

File	Contents
ce	Clinical events, including urological events and ESRD
cm	Co-medications, includes Tolvaptan/Samsca (521) described as 'diuretic: vasopressin antagonist'. (Jynarque not listed.)
dm	Demographics including age, gender and ethnicity
fa	Findings about interventions or events
fh	Familial history
ho	Healthcare encounters/hospitalisations
lb	Lab test results including creatinine levels

File	Contents
mh	Medical history
mp	Medical procedures including measurement of Kidney Volume
pf	Pharmacogenomics/genetic findings including polymorphism type, interpretation (PKD1, PKD2), genetic change or amino acid change.
pr	Procedures, e.g. kidney stone removal.
rs	Reproductive system findings
su	Substance use
vs	Vital signs including blood pressure, height and weight.

The Clinical Endpoints

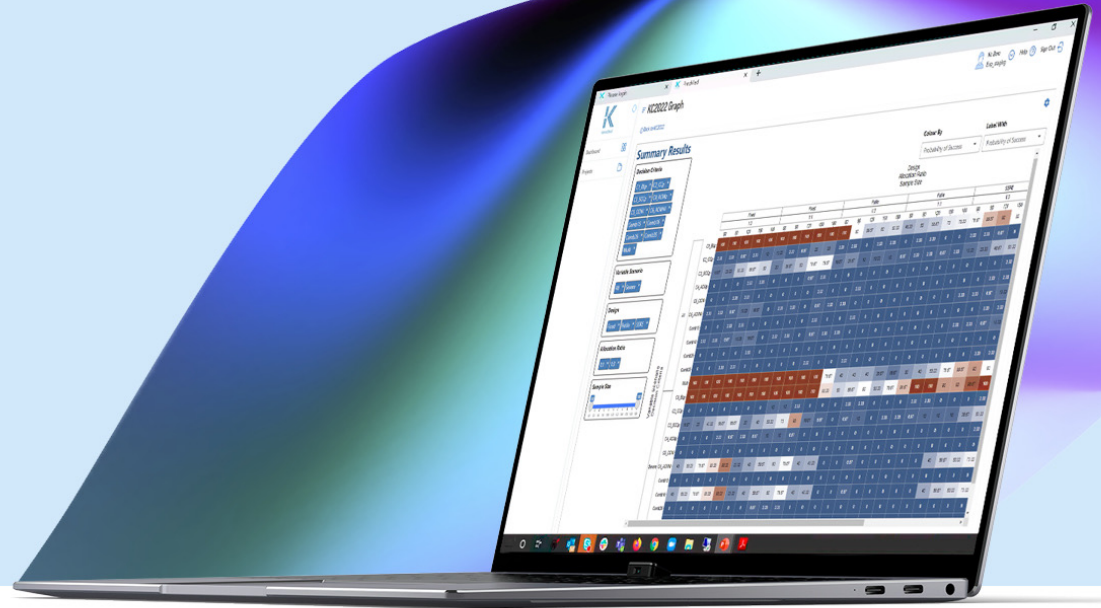
Frequency of top 10 end points across **126 ADPKD trials** in the clinical trial registry (clinicaltrials.gov).

	Frequency ↕
Total Kidney Volume (TKV)	68
Glomerular Filtration Rate (GFR)	49
Estimated Glomerular Filtration Rate eGFR	47
Quality of Life	40
Adverse Events	36
Hypertension/Blood Pressure	28
Urine Osmolality	23
Pain	20
Liver Volume	14
Alanine Transaminase	12

Endpoints and information available in the C-PATH disease registry dataset:

- Demographic information (sex, age, ethnicity)
- Total Kidney Volume
- Measured Glomerular Filtration rate (mGFR).
- Laboratory results including **Creatinine** levels and Albuminuria
- Blood pressure (systolic/diastolic).
- Medical history/clinical events including ESRD.
- Medications taken.
- Pharmacogenomic/Genetic information.
- **No estimated Glomerular Filtration rate (eGFR).**

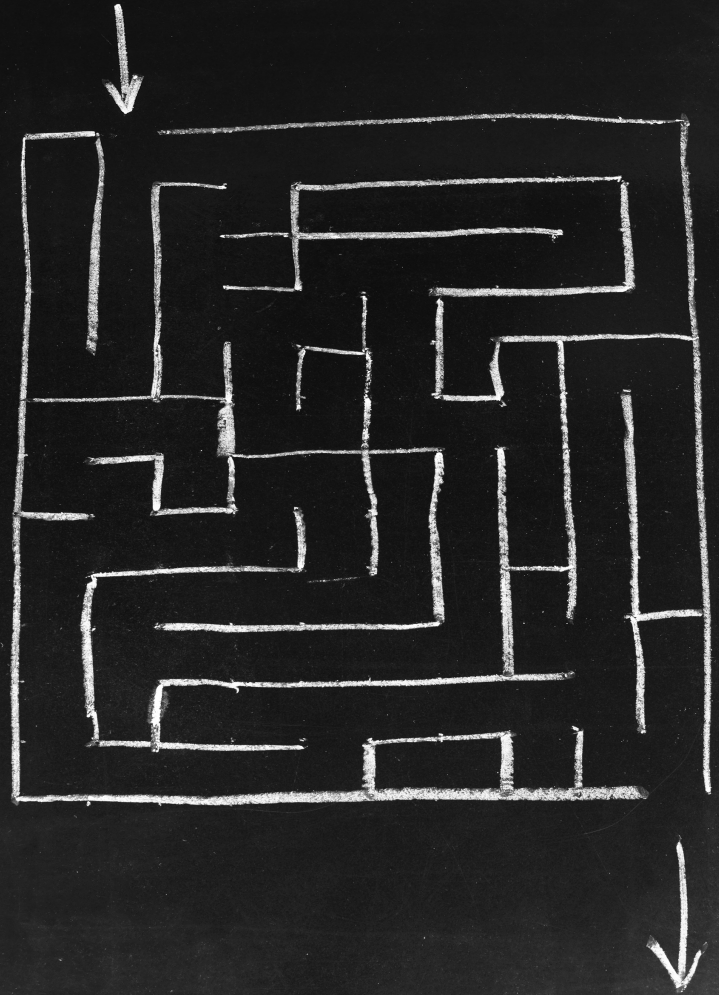
The Approach



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

What did we want to do?

- Compare disease registry data to clinical trial data.
- Use the disease registry data to run a clinical trial simulation to help inform optimal clinical trial design
- Compare a simulation informed by individual participant level registry data with one informed by summary clinical trial data.
- Go through the process and see what we find out along the way.



The Approach

- Mimic a key ADPKD trial using the disease registry data: Disease Registry IPD (DR IPD).
- Run KerusCloud Simulations with the original Tolvaptan trial summary data and with the DR IPD.
- Potential final figures will have comparison of:

Simulation:
Disease registry IPD
data.

Simulation:
Tolvaptan trial
summary data

- Investigating subgroups, correlations, distributions and variability.

Hypothesis:

Disease Registry mimic data will allow investigation of disease subgroups such as gender, TKV cut-offs and ESRD, and also exploration of variable relationships such as between eGFR and TKV.

<https://www.nejm.org/doi/full/10.1056/NEJMoa1205511>

Tolvaptan in Patients with Autosomal Dominant Polycystic Kidney Disease

Vicente E. Torres, M.D., Ph.D., Arlene B. Chapman, M.D.,
Olivier Devuyst, M.D., Ph.D., Ron T. Gansevoort, M.D., Ph.D.,
Jared J. Grantham, M.D., Eiji Higashihara, M.D., Ph.D., Ronald D. Perrone, M.D.,
Holly B. Krasa, M.S., John Ouyang, Ph.D., and Frank S. Czerwiec, M.D., Ph.D.,
for the TEMPO 3:4 Trial Investigators*

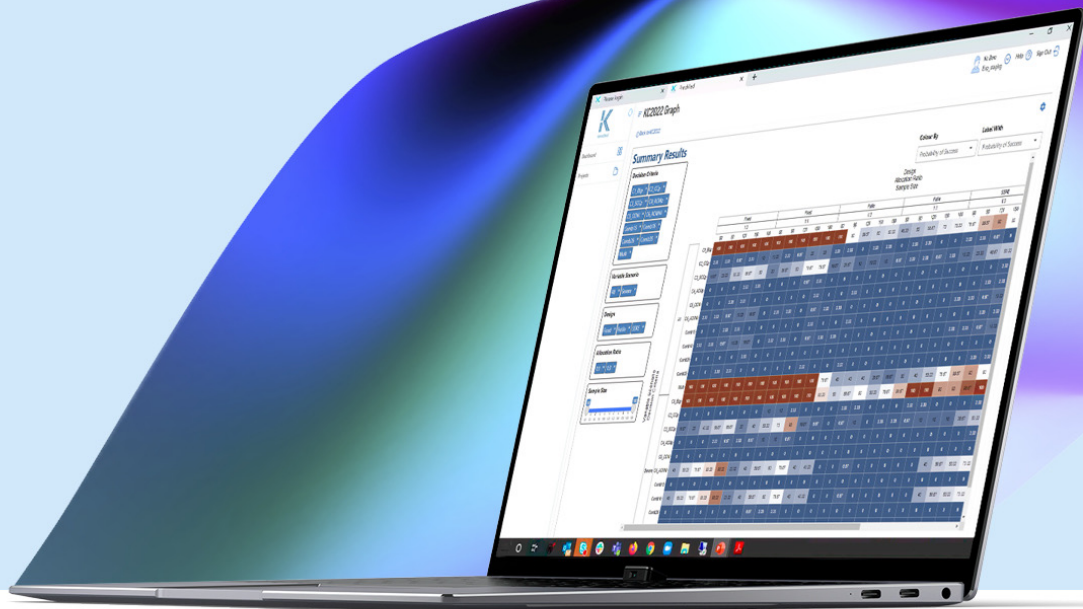
METHODS

In this phase 3, multicenter, double-blind, placebo-controlled, 3-year trial, we randomly assigned 1445 patients, 18 to 50 years of age, who had ADPKD with a total kidney volume of 750 ml or more and an estimated creatinine clearance of 60 ml per minute or more, in a 2:1 ratio to receive tolvaptan, a V₂-receptor antagonist, at the highest of three twice-daily dose regimens that the patient found tolerable, or placebo. The primary outcome was the annual rate of change in the total kidney volume. Sequential secondary end points included a composite of time to clinical progression (defined as worsening kidney function, kidney pain, hypertension, and albuminuria) and rate of kidney-function decline.

Key ADPKD trial

December 20, 2012
N Engl J Med 2012; 367:2407-2418
DOI: 10.1056/NEJMoa1205511

Disease Registry & Data Wrangling Challenges



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Clinical Trial Criteria

- **Two visits** - Individuals with a visit at baseline and at 3 years.
- **Age** - between 18 and 50 years old.
- **Total Kidney Volume** - 750mL or more.
- **eGFR** (same as estimated creatinine clearance) - 60 mL per minute or more.

Primary outcome: Annual rate of change for the TKV.

Secondary outcome: Composite time to clinical progression (defined as worsening kidney function, kidney pain, hypertension, and albuminuria) and rate of kidney-function decline.

Tolvaptan in Patients with Autosomal Dominant Polycystic Kidney Disease

Vicente E. Torres, M.D., Ph.D., Arlene B. Chapman, M.D.,
Olivier Devuyst, M.D., Ph.D., Ron T. Gansevoort, M.D., Ph.D.,
Jared J. Grantham, M.D., Eiji Higashihara, M.D., Ph.D., Ronald D. Perrone, M.D.,
Holly B. Krasa, M.S., John Ouyang, Ph.D., and Frank S. Czerwiec, M.D., Ph.D.,
for the TEMPO 3:4 Trial Investigators*

METHODS

In this phase 3, multicenter, double-blind, placebo-controlled, 3-year trial, we randomly assigned 1445 patients, 18 to 50 years of age, who had ADPKD with a total kidney volume of 750 ml or more and an estimated creatinine clearance of 60 ml per minute or more, in a 2:1 ratio to receive tolvaptan, a V_2 -receptor antagonist, at the highest of three twice-daily dose regimens that the patient found tolerable, or placebo. The primary outcome was the annual rate of change in the total kidney volume. Sequential secondary end points included a composite of time to clinical progression (defined as worsening kidney function, kidney pain, hypertension, and albuminuria) and rate of kidney-function decline.

How hard
could it be?



Challenges and Considerations for Disease Registry Data

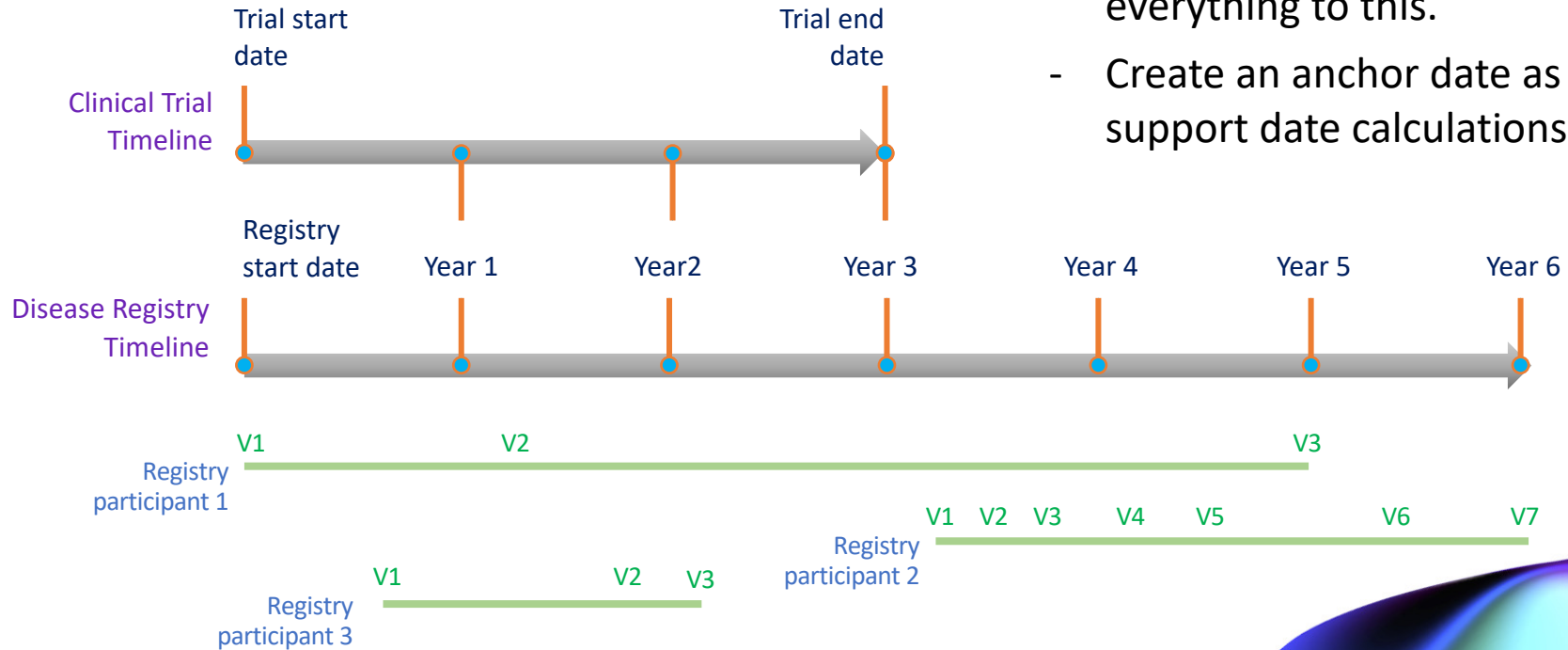
- Understanding and wrangling data in the SDTM format.
- Dealing with the non-standardised visit pattern, likely due to hospitalisations or disease related events.
- Aligning the registry mimic endpoints with the Tolvaptan clinical trial endpoints including deriving eGFR from creatinine.
- Aligning the endpoints within the disease registry, i.e. TKV and eGFR.
- Dwindling N numbers.



Disease Registry Timelines

Approach:

- Total kidney volume as an anchor endpoint – match everything to this.
- Create an anchor date as well to support date calculations



Time standardisation

- Example of visit days (MPDY) for one individual

MPDY	FirstDay	newDay	Anchor_date	False_date	Months	Years
5600	5212	388	1992-09-20	1993-10-13	12	1
5615	5212	403	1992-09-20	1993-10-28	13	1
6329	5212	1117	1992-09-20	1995-10-12	36	3
5212	5212	0	1992-09-20	1992-09-20	0	0
6028	5212	816	1992-09-20	1994-12-15	26	2
5663	5212	451	1992-09-20	1993-12-15	14	1

- Identify lowest visit day and set as baseline.
- Subtract from remaining visit days.
- Add difference to anchor date to get a new false date.
- Use anchor date and false date to calculate difference in months and years via *Lubridate* R package

Derivation of eGFR from creatinine levels

National Institute of Diabetes and Digestive and Kidney Diseases advises either the Modification of Diet in Renal Disease (MDRD) Study equation or the Chronic Kidney Disease Epidemiology Collaboration (CKD – EPI) equation for estimating GFR in individuals aged 18 years and above.

MDRD equation include age, gender and race as well as serum creatinine concentration.

$$\text{GFR (mL/min/1.73 m}^2\text{)} = 175 \times (S_{\text{cr}})^{-1.154} \times (\text{Age})^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if African American})$$

S_{cr} = serum creatinine concentration

Age

Gender

Ethnicity

The equation does not take height or weight variables because the results are reported normalised to 1.73 m² body surface area, an accepted adult surface area.

Dwindling N numbers

N = 2496

All N numbers

N = 760

Two visits available
for TKV

N = 93

With 3 year
timepoint

N = 21

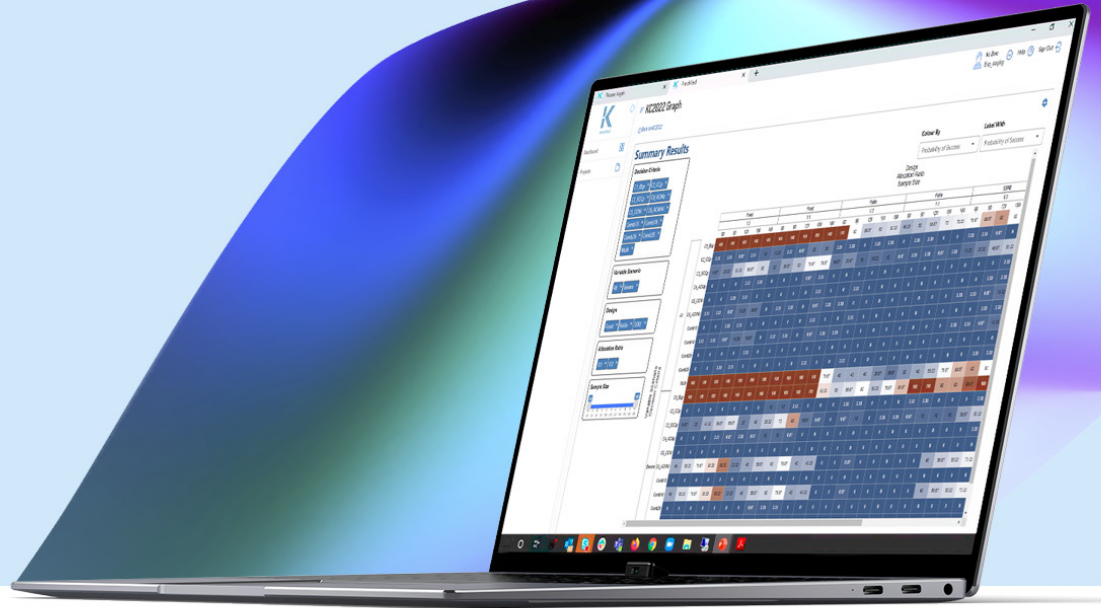
Match
Trial



What?

Baseline age of 18 – 50 years;
TKV $\geq$ 750 mL; eGFR $\geq$ 60 mL

Disease Registry Data versus the Clinical Trial



exploristics

INNOVATIVE ANALYTICS SOLUTIONS

Comparing the baseline summary data for each simulation input

Tolvaptan Clinical Trial Data Summary Torres *et al*, 2016

Characteristic	Tolvaptan (N = 961)	Placebo (N = 484)
Male sex — no. (%)	495 (51.5)	251 (51.9)
Age — yr	39±7	39±7
Race — no. (%)†		
White	810 (84.3)	408 (84.3)
Asian	121 (12.6)	62 (12.8)
Other	30 (3.1)	14 (2.9)
Total kidney volume — ml	1705±921	1668±873
Estimated GFR — ml/min/1.73 m ² ¶	81.35±21.02	82.14±22.73

The values for age and sex are not greatly different; however Race does differ.

The TKV and eGFR values are comparable; and the disease registry IPD data does not have any individuals with Tolvaptan co-medication.

Disease Registry IPD Summary

Characteristic (N=21)	
Male sex (%)	66.6
Age — yr	31.6±7
Race (%): White	95.2
Black	4.8
Total Kidney Volume (mL)	1412 ± 608
eGFR (mL/min/1.73m ²)	83.9 ± 23.92
Tolvaptan (%)	0
End Stage Renal Disease — no. (%)	2 (9.5)

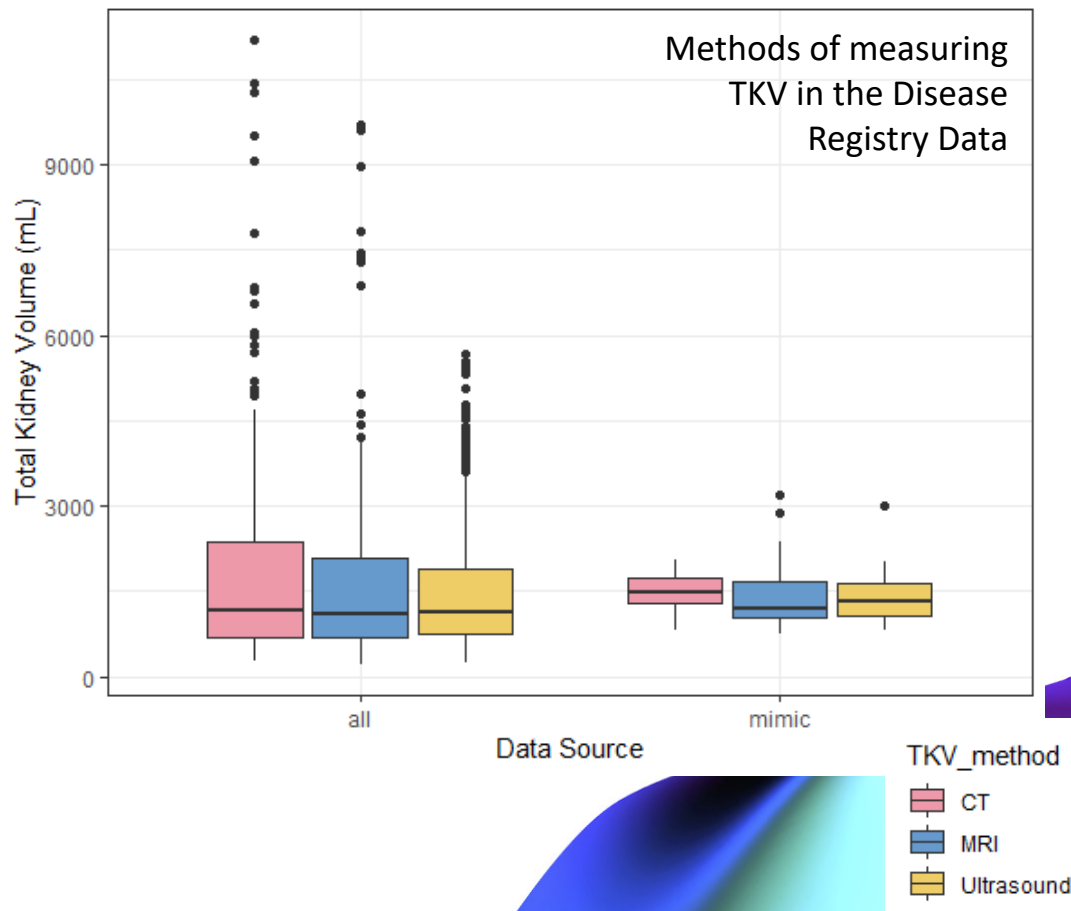
Disease Registry IPD is more messy than trial data

The Disease Registry Data records are likely to be associated with a clinical event.

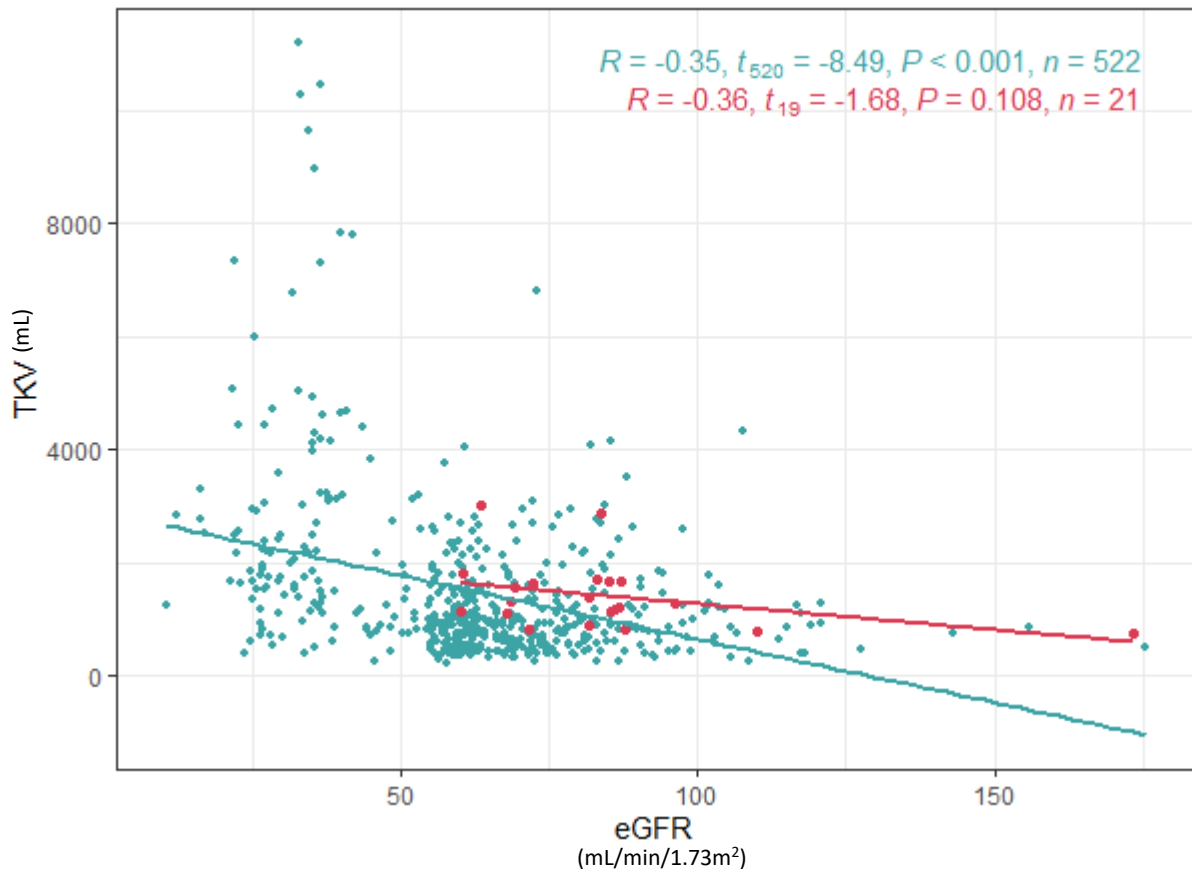
In the Registry data there are:

- all ages
- a broad range of disease severity
- comorbidities
- comedications
- variability in measurement methods e.g. procedure, laboratory

There are three methods for measuring TKV in the registry data (CT, MRI and Ultrasound), in the original clinical trial measured TKV via MRI only.



Disease Registry Data can provide correlations between core endpoints

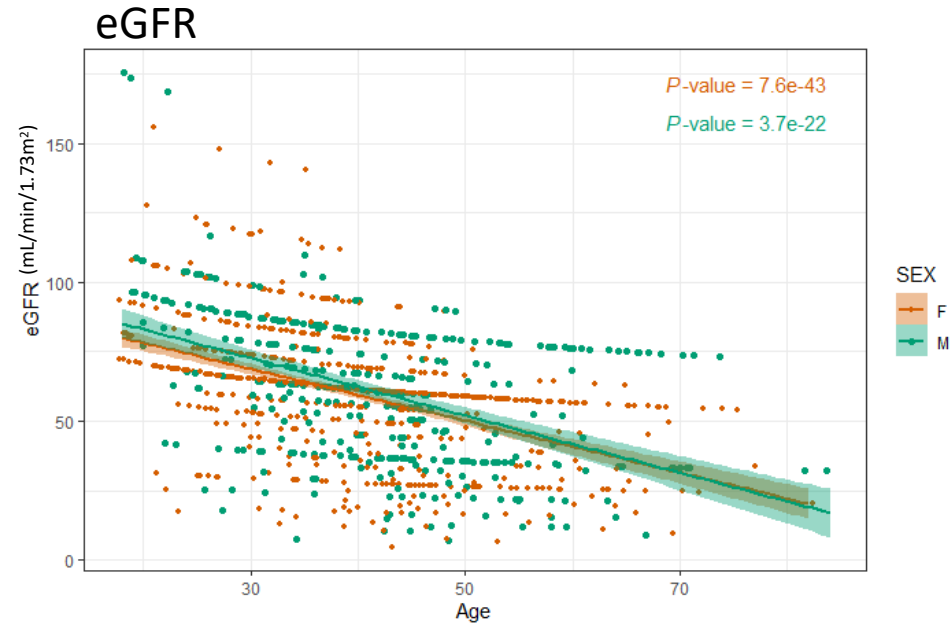
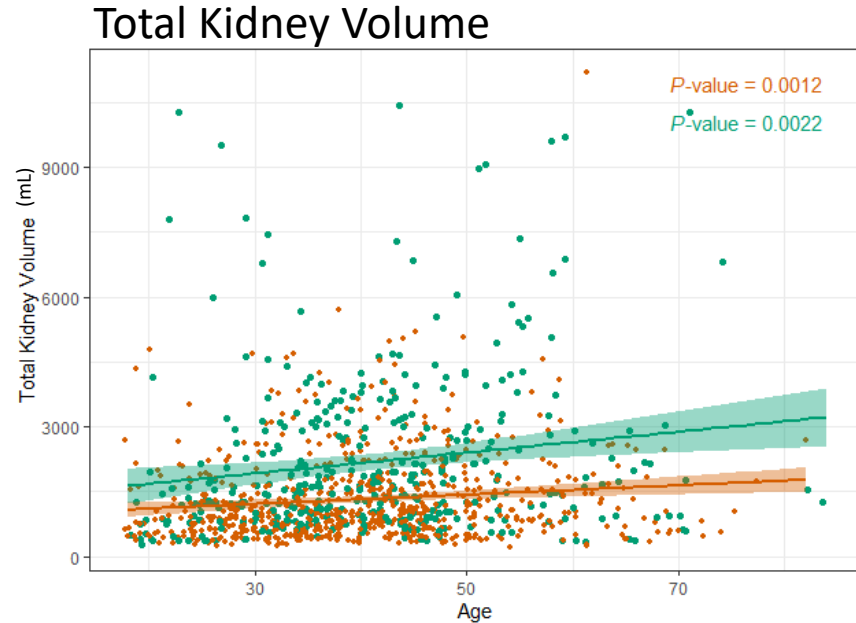


Using **Disease Registry IPD** a value for the correlation between TKV and eGFR can be obtained for both:

- disease registry data mimicking the Tolvaptan Clinical Trial eligibility criteria (mimic, $R = -0.36$, $p = 0.108$)
- And for the entire disease registry (all, $R = -0.35$, $p < 0.001$).

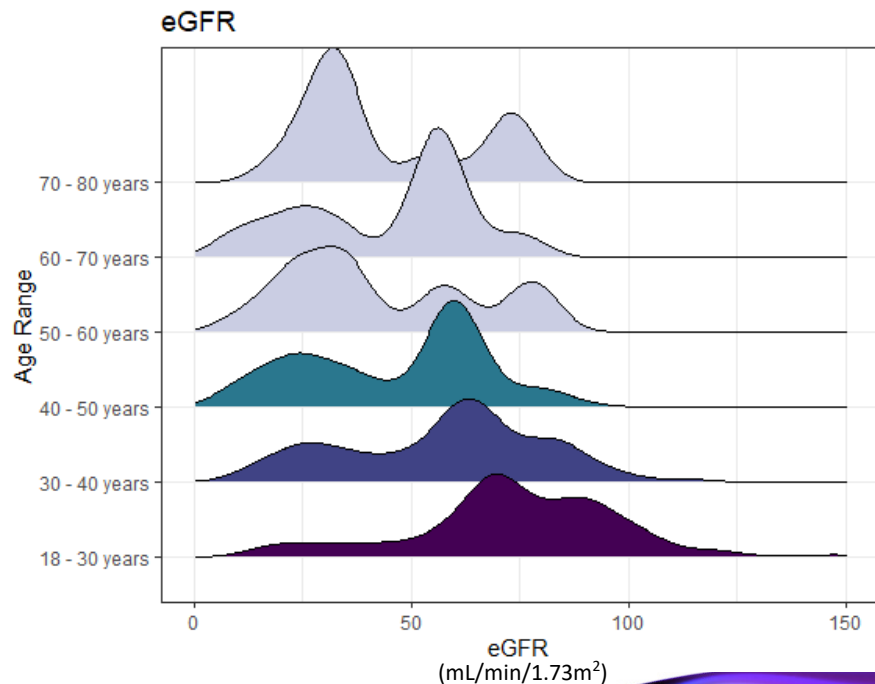
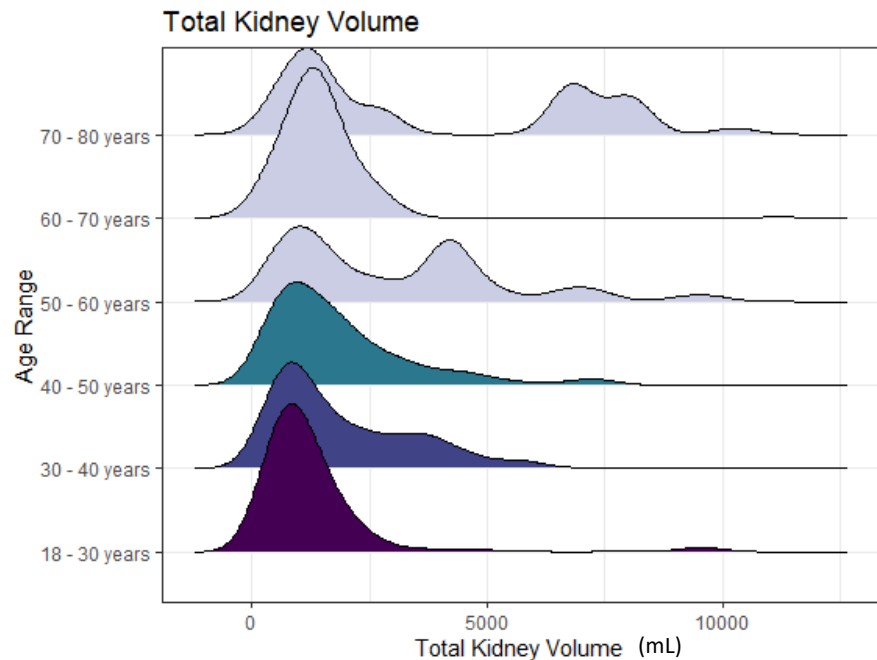
With the **Tolvaptan Clinical Trial Summary Data** it is not possible to know the correlation between endpoints.

Disease Registry IPD can inform on endpoint and subgroup relationships.



With **Disease Registry IPD** the relationship between either Total Kidney Volume or eGFR with Age and Sex can be investigated (Pearson's correlation). These relationships are not captured in the **Tolvaptan Clinical Trial Summary Data**.

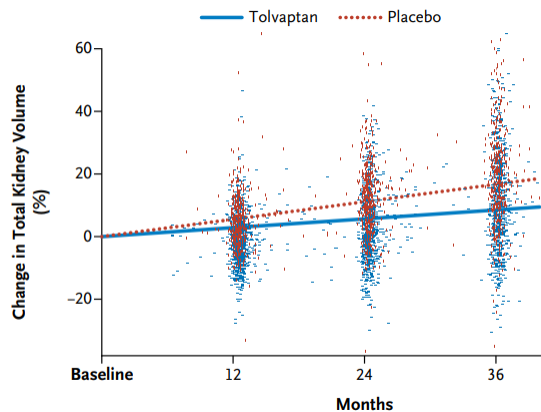
Disease Registry data can show information beyond restrictions of trial eligibility criteria



The Tolvaptan Clinical Trial eligibility criteria restrict age range to 18 – 50 years at baseline (dark ridges). The Disease Registry IPD can show endpoint information beyond these age ranges (light ridges).

Disease Registry IPD analysis can span greater time range

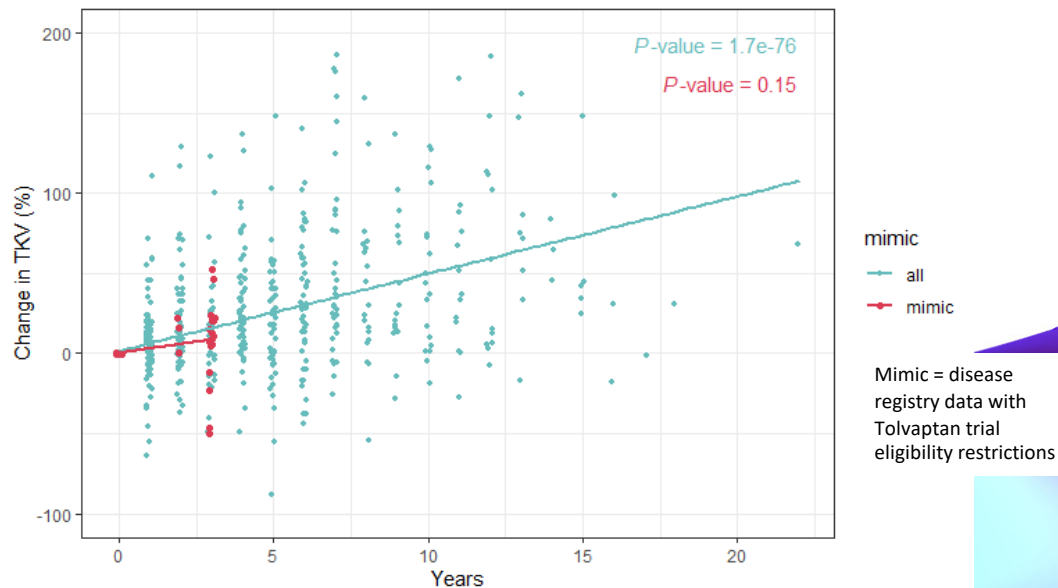
From Torres *et al*, 2012: 0 – 36 months (3 years)



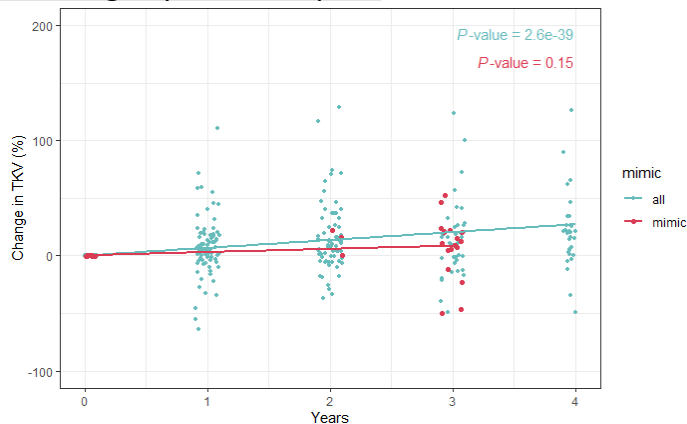
One of the main figures from the Tolvaptan Trial showed an increase in TKV in the tolvaptan group of **2.8%** per year (CI: 2.5 to 3.1) versus **5.5%** per year (CI: 5.1 – 6.0) in placebo group (figure from Torres *et al*, 2012).

The overall increase in TKV Disease registry IPD was **5.11%** per year from data ranging right up to 22 years.

Disease Registry IPD: 0 – 22 years



Disease Registry IPD: 0 – 4 years



Disease Registry IPD can incorporate retrospectively informative variables

Disease Registry data enables investigation of disease progression outcomes including End-stage Renal Disease (ESRD), and how they're influenced by endpoints at baseline and across the individual's lifetime.

C-Path consortium reported TKV levels of ≥ 1000 mL as a predictor of kidney failure later in life.

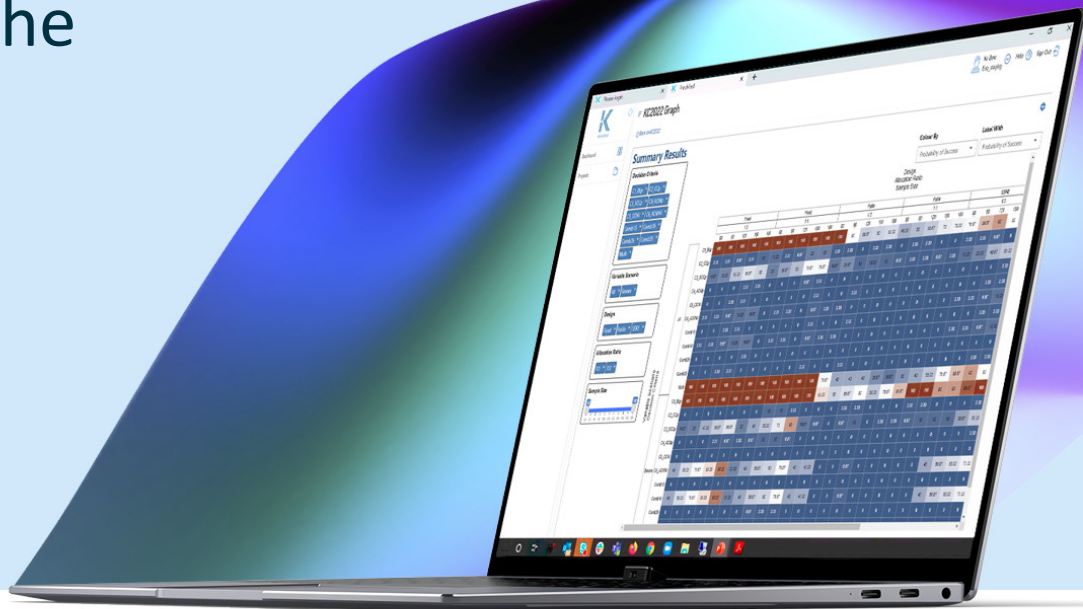
This is borne out by the number of individuals who had TKV ≥ 1000 mL and eventually had ESRD.

This information can be incorporated into simulations.

Baseline TKV & End-stage Renal Disease



KerusCloud Simulation and assessment of the Virtual Cohort



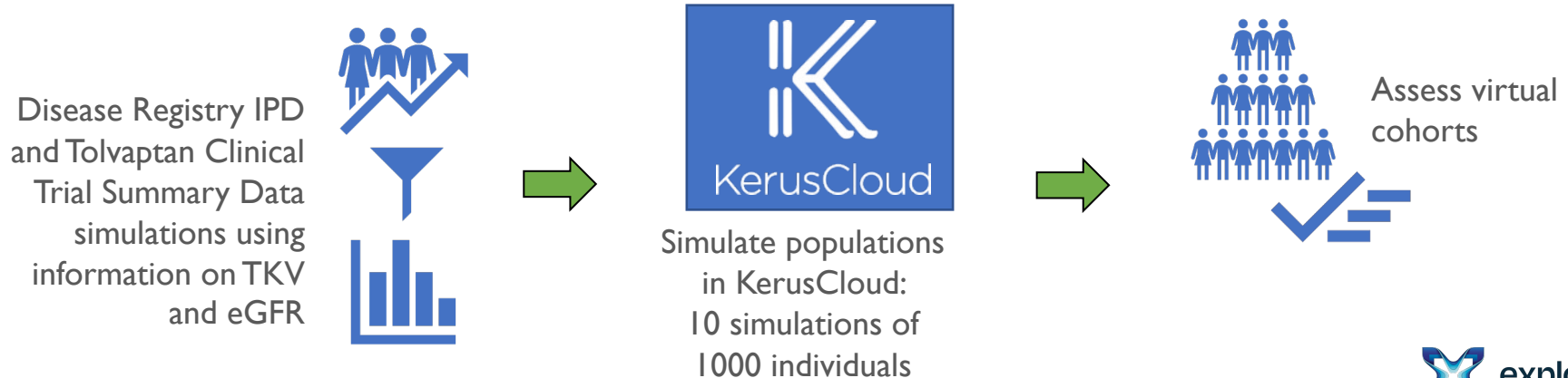
exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The KerusCloud Simulation

The KerusCloud software:

- Is a cloud-based study simulation platform used to statistically optimise clinical studies from the outset to maximise their likelihood of success
- Generates realistic IPD, using it to run virtual clinical trials
- Allows exploration of a wide range of study design and analysis options in a virtual environment

KerusCloud Simulation Workflow:



The KerusCloud Input

Log Total Kidney Volume (TKV)

Timepoint/Measure	Population/Arm	Tolvaptan Clinical Trial Summary Data		Disease Registry IPD Data	
		Mean	SD	Mean	SD
Baseline	Population	7.31	0.51		
	Male			7.20	0.38
	Female			7.14	0.43
Three Year	Population	7.41	0.47		
	Male			7.21	0.35
	Female			7.25	0.42
Percent change	Population	2.8	4.4		
	Male			1.99	8.87
	Female			4.56	7.53

Tolvaptan Clinical Trial Summary data input is limited to what is reported in the paper (Torres *et al*, 2012), whereas the Disease Registry IPD allows for subgroup analysis and exploration.

The KerusCloud Input

Log estimated Glomerular Filtration Rate (eGFR)

Timepoint/Measure	Population/Arm	Tolvaptan Clinical Trial Summary Data		Disease Registry IPD Data	
		Mean	SD	Mean	SD
Baseline	Population	4.38	0.27		
	Male			4.47	0.25
	Female			4.27	0.11
Three year	Population	4.26	0.30		
	Male			4.31	0.38
	Female			4.11	0.35

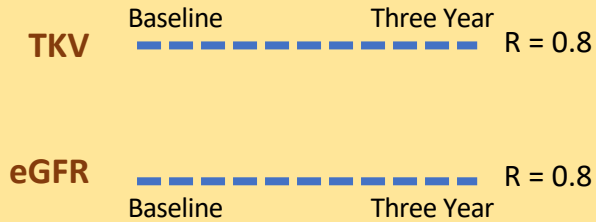
The KerusCloud Input

Endpoint Correlations

Endpoint 1	Timepoint 1	Endpoint 2	Timepoint 2	Tolvaptan Clinical Trial Summary Data R	Disease Registry IPD Data
Log TKV	Baseline	Log TKV	Three years	0.8	0.73
Log eGFR	Baseline	Log eGFR	Three years	0.8	0.71
Log TKV	Baseline	Log eGFR	Baseline		-0.45
Log TKV	Three years	Log eGFR	Three years		-0.20
Log TKV	Baseline	Log eGFR	Three years		-0.46
Log eGFR	Baseline	Log TKV	Three years		-0.21

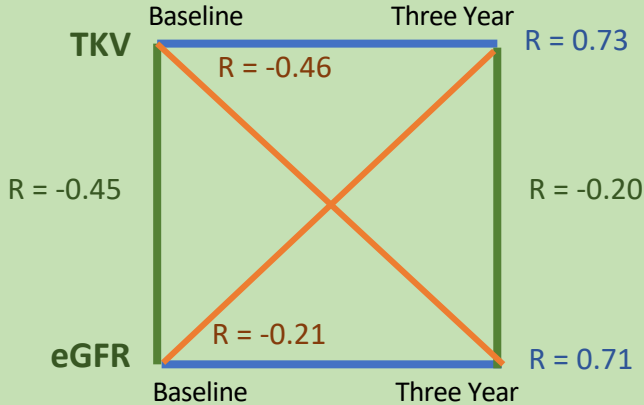
Simulations derived from disease registry data contain more information.

Tolvaptan Clinical Trial Summary Data



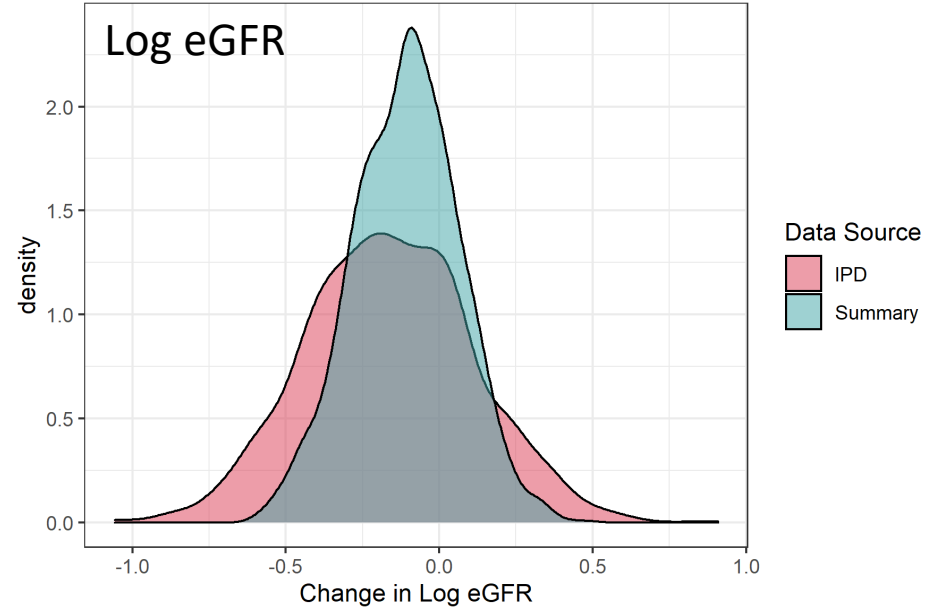
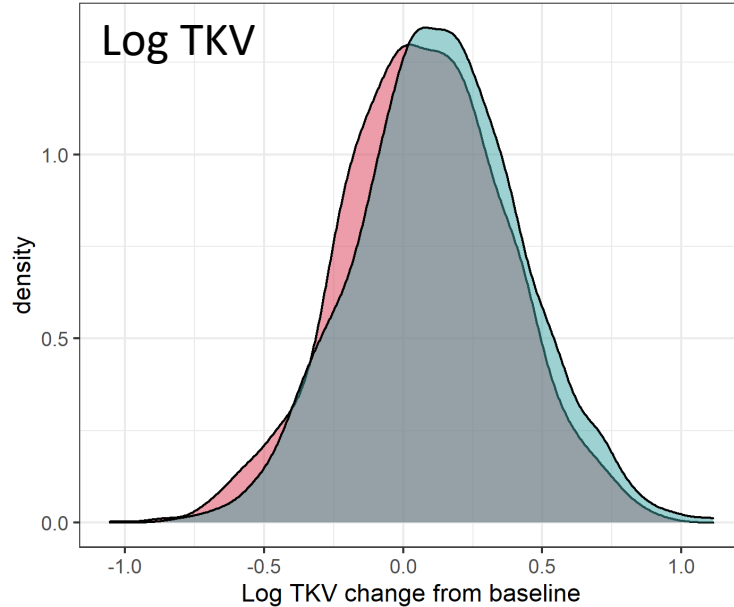
Correlations are **assumed**

Disease Registry IPD



Correlations are **observed**

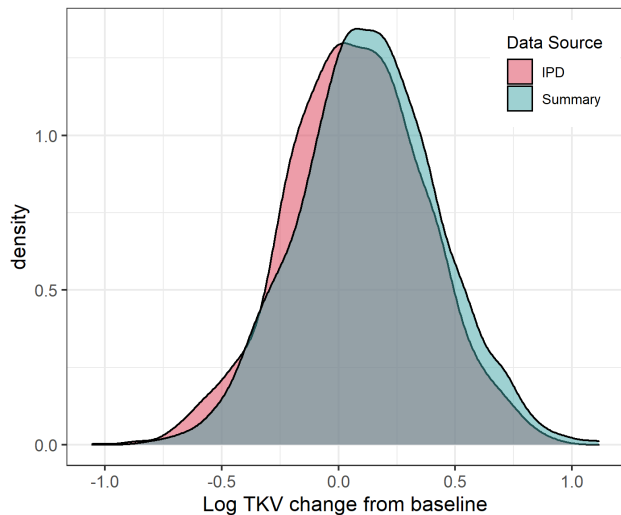
Simulations from IPD show details for data spread and distribution



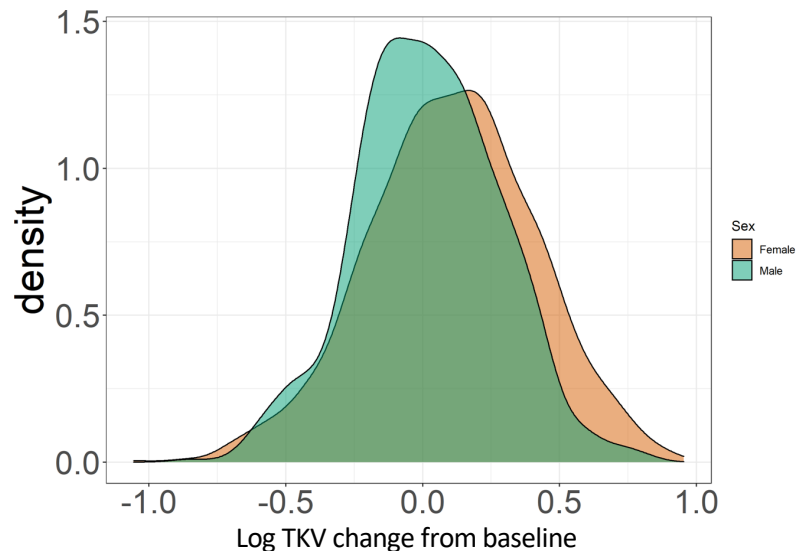
The Disease Registry IPD simulation data (red) has a slightly different spread than the Tolvaptan Clinical Trial summary data (blue) for both Log TKV and Log eGFR.

With Disease Registry IPD simulated cohorts pre-existing subgroups distributions can be investigated

Disease Registry IPD cohort vs Tolvaptan Clinical Trial Summary cohort

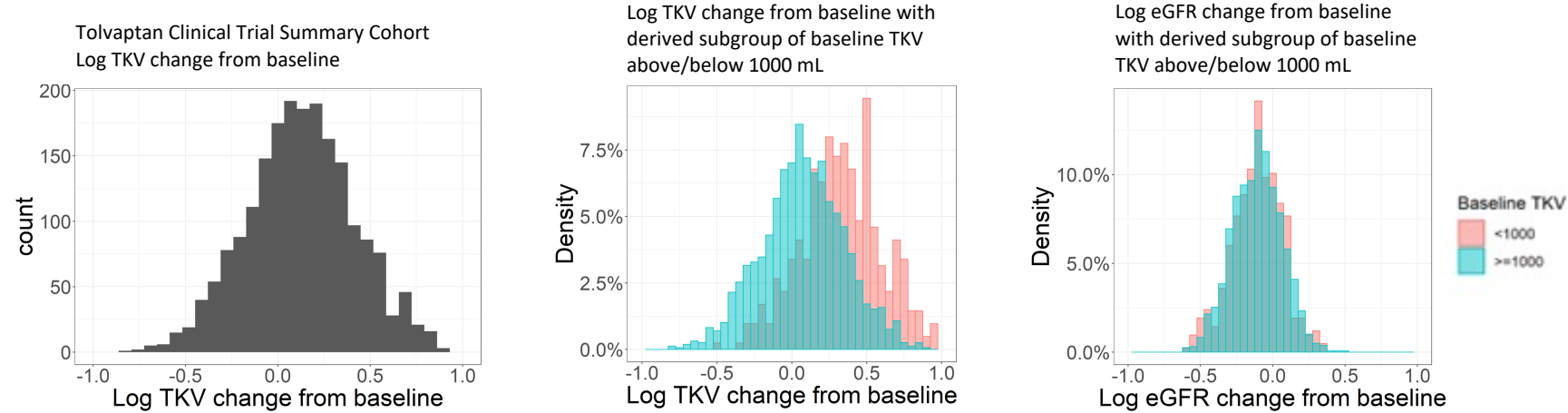


Disease Registry IPD cohort: Male vs Female

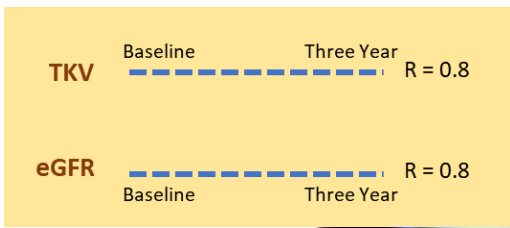


With Disease Registry IPD derived cohorts it's possible to investigate influence of subgroup differences on distribution, for example, if we further investigate the IPD distribution (pink) from the first plot, we find more females (orange) have greater log TKV change from baseline than males (green). It is not possible to look deeper into the Tolvaptan Clinical Trial Summary cohort (blue).

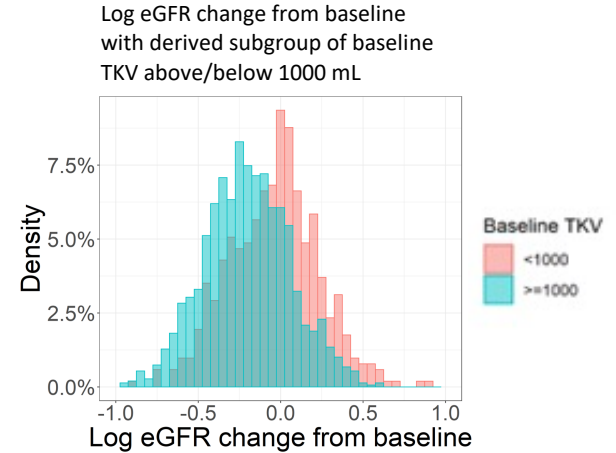
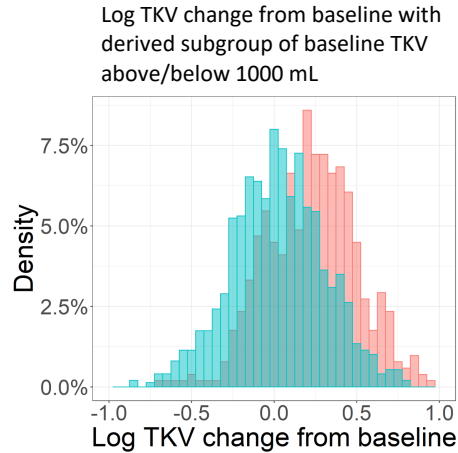
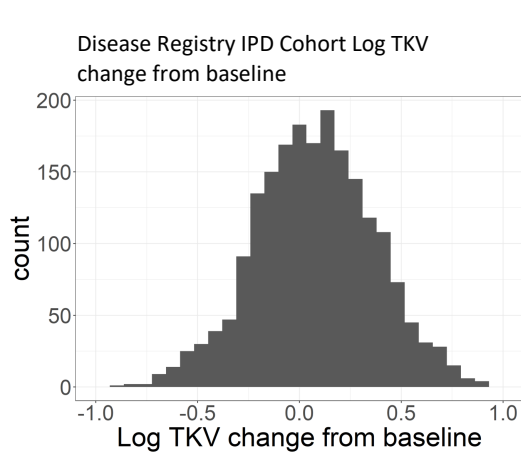
Assumed correlations within endpoints can provide limited information on derived subgroups



Simulation of the Tolvaptan Clinical Trial Summary data can assume a correlation between endpoints at baseline and at three years, e.g. TKV correlation is assumed to be $R = 0.8$. This is reflected in the TKV change from baseline values, providing a distribution for derived subgroups such as baseline TKV above/below 1000 mL; however because there is no information about the correlation between TKV and eGFR, the distribution of Log eGFR change from baseline for the derived subgroups is absent.

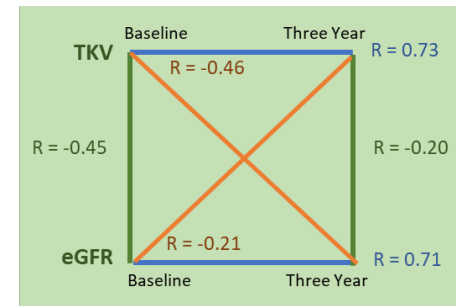


Observed correlations within and between endpoints can provide information on derived subgroups

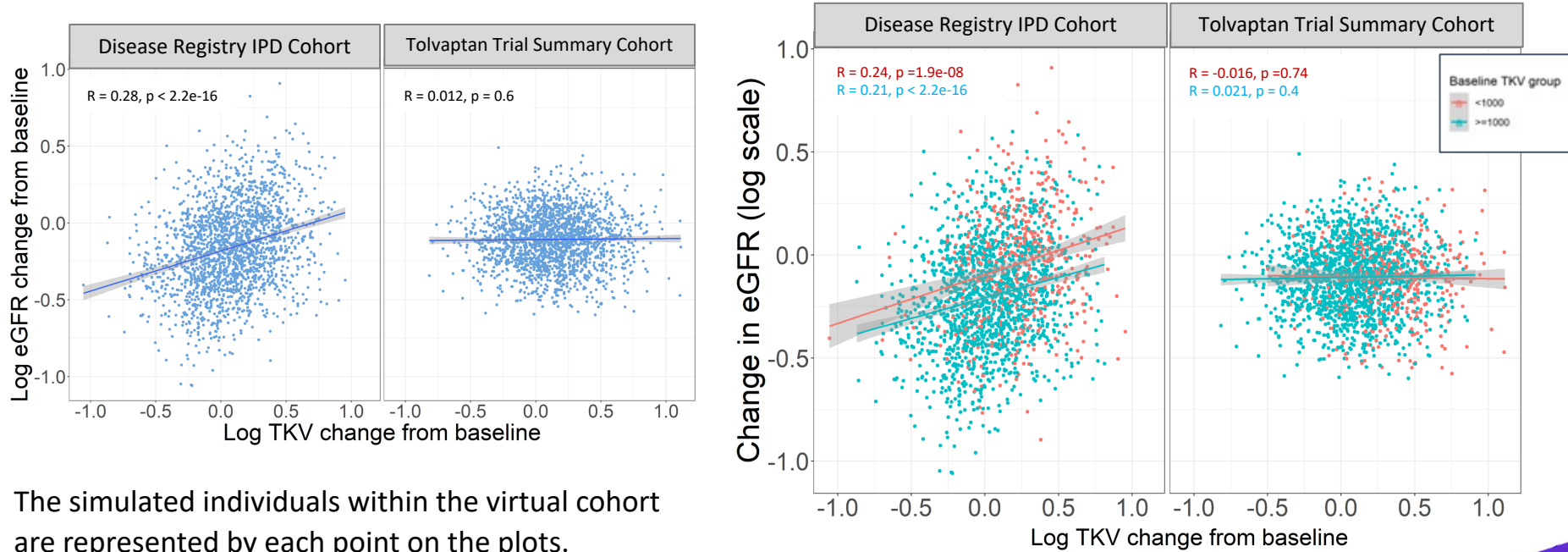


The simulation of Disease Registry IPD had input of observed correlations (green box). This is reflected not only in the Log TKV change from baseline values, providing a distribution for derived subgroups such as baseline TKV above/below 1000 mL; but also in the Log eGFR change from baseline values as well.

The C-PATH announcement of TKV as a prognostic marker for ADPKD disease progression was in August 2015. The Tolvaptan Clinical Trial Results were published in 2012. Disease Registry IPD makes it possible to incorporate new endpoint information into a historical trial framework.



The simulated virtual cohorts are comprised of participants retaining endpoint relationships



The simulated individuals within the virtual cohort are represented by each point on the plots.

The Disease Registry IPD cohorts retain information about the endpoint correlations, absent in the cohorts from Tolvaptan Clinical Trial Summary Data.

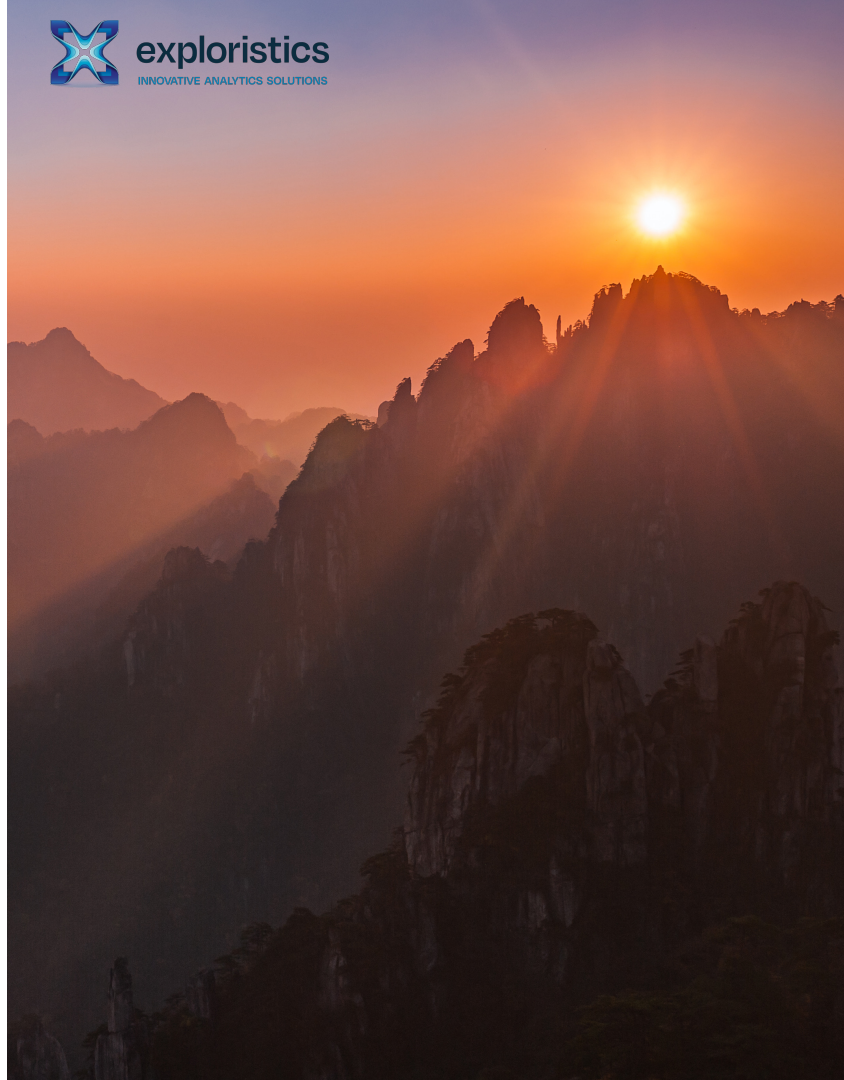
The simulated virtual cohorts are synthetic data, with characteristics of real data but without privacy issues, and can be used for clinical trial design optimisation.

Summary

- **The value of Disease Registry IPD:**
 - Information on Data distributions, endpoint relationships and subgroups
 - New information in the disease area literature can be combined with historical trial findings
- **Difficulties or drawbacks:**
 - The unstandardised nature of registry data
 - Wrangling timeframes
 - Dwindling N-numbers

Conclusion & Next Steps

- IPD registry data adds valuable detail to simulations, but requires consideration of caveats above coupled with time requirements.
- However, assess on a case by case basis, trying different trial parameters, different time scales, and with different disease registries in different disease areas.



References

Aung TT, Bhandari SK, Chen Q, Malik FT, Willey CJ, Reynolds K, Jacobsen SJ, Sim JJ. Autosomal Dominant Polycystic Kidney Disease Prevalence among a Racially Diverse United States Population, 2002 through 2018. *Kidney360*. 2021 Sep 22;2(12):2010-2015. doi: 10.34067/KID.0004522021. PMID: 35419536; PMCID: PMC8986058.

Bergmann C, Guay-Woodford LM, Harris PC, Horie S, Peters DJM, Torres VE. Polycystic kidney disease. *Nat Rev Dis Primers*. 2018 Dec 6;4(1):50. doi: 10.1038/s41572-018-0047-y. PMID: 30523303; PMCID: PMC6592047.

Mayo Clinic. "Polycystic Kidney Disease - Symptoms and Causes." Accessed June 19, 2022. [Polycystic kidney disease - Symptoms and causes](#)

National Institute of Diabetes and Digestive and Kidney Diseases, <https://www.niddk.nih.gov/health-information/professionals/clinical-tools-patient-management/kidney-disease/laboratory-evaluation/glomerular-filtration-rate/estimating>

Steele, C.; Nowak, K. Obesity, Weight Loss, Lifestyle Interventions, and Autosomal Dominant Polycystic Kidney Disease. *Kidney Dial*. **2022**, 2, 106-122. <https://doi.org/10.3390/kidneydial2010013>

Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Grantham JJ, Higashihara E, Perrone RD, Krasa HB, Ouyang J, Czerwiec FS; TEMPO 3:4 Trial Investigators. Tolvaptan in patients with autosomal dominant polycystic kidney disease. *N Engl J Med*. 2012 Dec 20;367(25):2407-18. doi: 10.1056/NEJMoa1205511. Epub 2012 Nov 3. PMID: 23121377; PMCID: PMC3760207.

Willey C, Kamat S, Stellhorn R, Blais J. Analysis of Nationwide Data to Determine the Incidence and Diagnosed Prevalence of Autosomal Dominant Polycystic Kidney Disease in the USA: 2013-2015. *Kidney Dis (Basel)*. 2019 Mar;5(2):107-117. doi: 10.1159/000494923. Epub 2019 Jan 9. PMID: 31019924; PMCID: PMC6465773.

Acknowledgements



Ole J. Kaasbøll
Georg Vo Beiske



Antonia Popescu
Aisling Kearney
Oliver Dewhirst

Kim Hacquoil
Sam Miller



Data used in the preparation of this presentation were obtained from the Polycystic Kidney Disease Outcomes Consortium (PKDOC). As such, the investigators within PKDOC contributed to the design and implementation of the PKDOC database and/or provided data but did not participate in the analysis of the data or the writing of this report.

Alecia Nickless
Andrew Mills
Jamie Inshaw