Initial validation of a machine learning-derived prognostic test (KidneyIntelX) integrating biomarkers and electronic health record data to predict longitudinal kidney outcomes

Kinsuk Chauhan¹, Girish N. Nadkarni¹, Fergus Fleming², James McCullough², Cijiang J. He¹, John Quackenbush³, Barbara Murphy¹, Michael J. Donovan⁴, Steven G. Coca¹, Joseph V. Bonventre⁵

1. Department of Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
2. RenalytixAI, New York, NY
3. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
4. Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY
5. Renal Division, Brigham and Women's Hospital, Boston, Massachusetts

Correspondence:
Kinsuk Chauhan, MD, MPH
Division of Nephrology, Department of Medicine
Icahn School of Medicine at Mount Sinai
One Gustave L Levy Place, Box 1243
New York, NY 10029
Email: kinsuk.chauhan@mssm.edu
ABSTRACT

**Background:** Individuals with type 2 diabetes (T2D) or the Apolipoprotein L1 high-risk (APOL1-HR) genotype are at increased risk of rapid kidney function decline (RKFD) and kidney failure. We hypothesized that a prognostic test using machine learning integrating blood biomarkers and longitudinal electronic health record (EHR) data would improve risk stratification.

**Methods:** We selected two cohorts from the Mount Sinai BioMe Biobank: T2D (n=871) and African Ancestry (AA) with APOL1-HR (n=498). We measured plasma tumor necrosis factors 1/2 (TNFR1/2), and kidney injury molecule-1 (KIM-1) and used random forest (RF) algorithms to integrate biomarker and EHR data to generate a risk score for a composite outcome: RKFD (eGFR decline of ≥ 5 ml/min/year), or 40% sustained eGFR decline, or kidney failure. We compared performance to a validated clinical model and applied thresholds to assess the utility of the prognostic test (KidneyIntelX) to accurately stratify patients into risk categories.

**Results:** Overall, 23% with T2D and 18% of APOL1-HR experienced the composite endpoint over a median follow-up of 4.6 and 5.9 years, respectively. The AUC of KidneyIntelX was 0.77 (95% CI, 0.75-0.79) in T2D, and 0.80 (95% CI, 0.77-0.83) in APOL1-HR, out-performing the clinical models (AUC 0.66 [95% CI 0.65-0.67] and 0.72 [95% CI 0.71-0.73], respectively, p <0.001). The positive predictive values (PPV) for KidneyIntelX were 62% and 62% vs. 46% and 39% for the clinical models (p< 0.01) in high-risk (top 15%) strata for T2D and APOL1-HR, respectively. The negative predictive values (NPV) for KidneyIntelX were 92% in T2D and 96% for APOL1-HR vs. 85% and 93% for the clinical model, respectively, (p=NS) in low-risk strata (bottom 50%).

**Conclusions:** In patients with T2D or APOL1-HR, a prognostic test (KidneyIntelX) integrating biomarker levels with longitudinal EHR data, significantly improved prediction of a composite kidney endpoint of RKFD, 40% decline in eGFR or kidney failure over validated clinical models.
INTRODUCTION

Chronic kidney disease (CKD) affects over 35 million individuals in the United States. Diabetic kidney disease (DKD) due to type 2 diabetes (T2D) accounts for 44% of end-stage kidney disease (ESKD) patients and is a major independent risk factor for other complications. African Ancestry (AA) have higher rates of ESKD compared to European Americans (EAs) across all baseline estimated glomerular filtration rate (eGFR) levels. Of relevance, genetic studies demonstrated that two distinct alleles in the Apolipoprotein L1 (APOL1) gene confer increased risk for many kidney diseases in AAs. The APOL1 high-risk (HR) genotypes (i.e., two copies of risk allele) are associated with increased risk of ESKD, CKD incidence/progression, and eGFR decline.

Even though these populations are on average higher risk than the general population, accurate prediction of who will have rapid kidney function decline (RKFD, defined as eGFR decline > 5 ml/min/1.73m²/year, and worse kidney outcomes is lacking. A current standard for ESKD prediction in CKD Stages 3-5 is the kidney failure risk equation (KFRE), where clinical variables are assigned standard weights for a recursive score calculation. However, KFRE has not been validated in individuals with relatively preserved kidney function at baseline. Recent work has shown other recursive scores may aide in risk prediction of kidney outcomes in patients with preserved kidney function.

Several biomarkers have been investigated to aide in the prediction of kidney outcomes. Three of the most extensively studied and strongly associated biomarkers with kidney disease progression in several settings are soluble tumor necrosis factors 1 and 2 (TNFR1/2), and plasma kidney injury molecule-1 (KIM-1). While these markers have uniformly shown an independent association with kidney outcomes risk along with specific clinical variables such as eGFR and UACR, the implementation of accurate models which combine clinical data with these plasma biomarkers to predict progression of kidney disease is unavailable.

Widespread electronic health records (EHR) usage provides the potential to leverage thousands of clinical features. Standard statistical approaches are inadequate to leverage this data due to feature volume, unaligned nature of data, and correlation structure. However, contemporary machine learning approaches have improved the capacity of analytical model development to combine both biomarkers and longitudinal EHR data for improved prediction.

In the current study, we used retrospectively collected plasma samples linked to longitudinal clinical data from the Icahn School of Medicine at Mount Sinai (ISMMS) BioMe Biobank to examine the ability of a prognostic test (KidneyIntelX) which uses machine-learning algorithms to predict RKFD and kidney outcomes in two discrete, high-risk patient populations, T2D and APOL1-HR.
MATERIALS AND METHODS

The BioMe Biobank at Icahn School of Medicine at Mount Sinai (ISMMS)
The BioMe Biobank at ISMMS is an Institutional Review Board (IRB)-approved biorepository which includes consented access to the patients' EHR from a diverse community in New York City, New York. Operations were initiated in 2007 and include direct recruitment from over 30 broadly selected clinical sites. For the purpose of this study, we selected two subpopulations: 1) T2D, enrollment eGFR 45-90 ml/min, and ≥3 years of follow up data; 2) APOL1-HR, enrollment eGFR > 30 ml/min and ≥ 3 years of follow up data. We included all patients from these two biobanks meeting these criteria (n=1369).

Ascertainment and Definition of the kidney endpoint
We determined eGFR using the CKD-EPI equation and eGFR slope using a minimum of 3 values from baseline. The primary composite outcome was comprised of three components: rapid kidney function decline (RKFD), defined as an eGFR slope decline of ≥ 5 ml/min/1.73 m²/year, or a sustained (confirmed ≥3 months later) decline in eGFR of ≥40% from baseline, or “kidney failure” defined by sustained eGFR < 15 ml/min/1.73 m² confirmed at least 30 days later, or long-term maintenance dialysis or kidney transplant (i.e. ESKD).

Ascertainment of clinical variables in BioMe Biobank
Sex and race were obtained from an enrollment questionnaire. Clinical data were extracted for all continuous variables at the time of and prior to baseline from the EHR with concurrent time stamps. Hypertension and T2D were determined using phenotyping algorithms. Cardiovascular disease and heart failure were determined by a validated algorithm and ICD-9/10 codes respectively. We considered a participant to be on an angiotensin converting enzyme-inhibitor (ACE-I) or angiotensin receptor blocker (ARB) if they had a concurrent prescription at enrollment. We calculated follow-up time from enrollment to the latest visit. Only variables present in >70% of subjects (except uACR/blood pressure due to their established clinical importance) were included and used for training of the KidneyIntelX algorithm.

Biospecimen Storage and Analyte Measurement
Plasma specimens collected on the day of BioMe enrollment were stored continuously at -80°C. The biomarkers were measured in a multiplex format using the Mesoscale platform (Meso Scale Diagnostics, Gaithersburg, Maryland, USA), employing proprietary electrochemiluminescence detection methods combined with patterned arrays allowing for analyte multiplexing. The intra- and inter-assay coefficient of variation (CV) for quality control samples with known low, moderate and high-concentrations of each biomarker run on each plate were 3.5%, 3.9%, and 4.5%, and 12.4%, 10.8%, and 7.7%, for TNFR1, TNFR2, and KIM1, respectively. The laboratory personnel were blinded to clinical information.
Statistical Analysis

We expressed descriptive results for the participants’ baseline characteristics and biomarkers via means and standard deviations, or for skewed variables, medians and interquartile ranges.

For the RF model, we considered two inputs: 1. biomarker concentrations/ratios; and 2. EHR features including laboratory values, diagnosis/procedure codes, demographics (age, sex, and listed race), medications, and healthcare encounter history. Missing uACR values were imputed to 10 mg/g, missing blood pressure (BP) values were imputed using multiple predictors (age, sex, race, and antihypertensive medications), and median value imputation was used for other missing values. We then created meta-features from these variables including maximum, minimum, median, variability, and change over time to account for their longitudinal aspect and repeated nature. For model development, the clinical data was randomly and demographically split to create an 80:20% training and test set, respectively, with 10-fold cross-validation on all candidate models.

We then performed further iterations of the random forest model by tuning three hyperparameters. Hyperparameter ‘1’ is the number of decision trees, hyperparameter ‘2’, the number of variables randomly selected for splitting at each node and hyperparameter ‘3’, the minimum size of terminal nodes. The final model was chosen which had the best area under the receiver operator characteristic curve, AUC.

We generated risk probabilities for the composite kidney endpoint using the final model on all subjects from both cohorts (T2D and APOL1-HR) and then scaled to generate a continuous score. We compared KidneyInteRx to a published validated clinical model consisting of a regression equation for 40% eGFR decline prediction, including age, sex, race, eGFR, cardiovascular disease, smoking, hypertension, BMI, and UACR in non-diabetics and the aforementioned variables plus insulin, diabetes medications, HbA1c for patients with T2D (eTable11; Ref 9). We compared all differences between AUCs using DeLong’s test for comparisons.

We examined the thresholds of the risk score to define low, intermediate, and high-risk strata in each cohort. The “low-risk” strata was set to encompass 50% of the study population, and the thresholds for the “high-risk” strata were assessed to classify the top 10, 15, and 20% highest risk in each cohort. The remaining population was defined as the “intermediate-risk” group. We calculated sensitivity, specificity, and positive/negative predicted values (PPV/NPV) for the high-risk and low-risk cutoffs and compared these to the clinical model. The goodness of fit statistics (Hosmer-Lemeshow) was used to assess calibration.
We then conducted two subgroup analyses 1. individuals with existing CKD (eGFR <60 ml/min/1.73 m² and/or UACR>30 mg/g at baseline), and 2. using only data from ≤1 year prior to biomarker measurement (i.e., “contemporary data” to ensure that KidneyIntelX was robust in advanced stages of the disease and performed equally well with limited clinical data to a year prior to biomarker measurement. 3. Generation of a random forest model trained and tested that did not include any of the biomarkers (TNFR1, TNFR2, or KIM-1) in both cohorts. 4. Evaluation of the performance of the full KidneyIntelX model for the individual components of the composite kidney endpoint. 5. Kaplan Meier Survival analyses for time-dependent outcomes of 40% decline and kidney failure with hazard ratios using the Cox proportional hazards method for the high-risk (top 15%) vs. the intermediate and low-risk groups (bottom 50%). All analyses were performed with R software (www.rproject.org)
RESULTS

Baseline Characteristics of Cohorts
Patients with T2D (n=871)
The median age was 60 years, 507 (58%) were female, and the median eGFR was 68 ml/min/1.73 m² (Table 1). The most common comorbidities were hypertension (93%), coronary heart disease (50%), and heart failure (22%). The majority (77%) were on ACE inhibitors or ARBs. Patient characteristics including events between the training and test cohorts were balanced (Supplementary Table 1).

Patients with APOL1-HR (n=498)
The median age was 56 years, 337 (67.6%) were female, and the median eGFR was 83.3 ml/min/1.73 m² (Table 1). The prevalence of comorbidities were lower than the T2D cohort; hypertension (44%), coronary heart disease (8%), and heart failure (3%). Patient characteristics including events between the training and test cohorts were comparable (Supplementary Table 2).

Composite Kidney Endpoint
For participants with T2D, 201 of the 871 (23%) experienced the composite kidney endpoint (RKFD) over a median follow-up of 4.6 (IQR 3.4-5.6) years. In participants with APOL1 high-risk genotype, 90 of the 498 (18%) experienced the composite kidney endpoint over a median follow up of 5.9 (IQR 3.9-7.1) years.

Machine learning (RF) model for prediction of the composite kidney endpoint
The observed composite kidney event by deciles of risk with KidneyIntelX vs. the standard clinical model⁹ are shown in Figures 1a and 1b. For patients with T2D, applying 10-fold cross-validation, the KidneyIntelX AUC in the training set (80%, n=697) for the composite kidney endpoint was 0.81 (95% CI: 0.80-0.82) and 0.77 (95% CI: 0.75-0.79) in the test set (20%, n=174). By comparison, the clinical model⁹ had an AUC of 0.66 (95% CI: 0.65-0.67) in the entire T2D cohort (n=871).

For the patients with APOL1 high risk genotype, applying 10-fold cross validation, the AUC for KidneyIntelX in the training set (80%, n=398) was 0.86 (95% CI: 0.84-0.87) and 0.80 (95% CI: 0.77-0.83) in the test set (20%, n=99). The clinical model⁹ had an AUC of 0.72 (95% CI: 0.71-0.73) in the APOL1-HR cohort (n=498).

In both T2D and HR-APOL1 cohorts, the features noted to contribute most to performance were the three plasma biomarkers (TNFR1, TNFR2 and KIM1) or their ratios of individual biomarker values to each other (i.e., 3 ratios) and laboratory values or vital signs (either baseline or changes over time) that are linked to kidney disease (Supplementary Figure 1). The P values of the Hosmer-Lemeshow goodness-of-fit test for the prognostic models were 0.15 and 0.11, indicating there was no significant difference between the predicted and observed outcomes (Supplementary Figure 2).
KidneyIntelX cutoffs for the composite kidney endpoint (entire T2D (n=871) and APOL1 (n=498) cohorts)
The PPVs of KidneyIntelX were 58, 62, and 68% in the top 20, 15 and 10% highest risk of the T2D population
vs. 43, 46 and 54% in top 20, 15 and 10% of highest risk as classified by the clinical model (p<0.01 for all comparisons, Table 2). The PPVs of KidneyIntelX were 56, 62, and 66% in the top 20, 15 and 10% highest risk of APOL1-HR population vs. PPV of 38, 39, and 40% of the highest risk as classified by the clinical model, (p<0.01 for all comparisons). When applying cut-offs for the lowest 50% of risk in the T2D cohort, the NPV for KidneyIntelX compared to clinical model was 92% vs. 85% (p=0.76). Similarly, for the APOL1-HR cohort, the NPV for KidneyIntelX compared to the clinical model was 96% vs. 93% (p=0.93).

Supplementary and Sensitivity Analyses
Prevalent CKD
When we stratified the performance of KidneyIntelX by baseline CKD (i.e., eGFR ≤ 60 ml/min/1.73 m² and/or
UACR≥30 at baseline, n=366), 27.6% experienced the primary composite kidney endpoint during follow-up,
compared to 12.5% in those without baseline CKD (n=505). The AUC was 0.84, 95% CI: 0.81-0.87 in individuals with prevalent CKD vs. 0.79, 95% CI: 0.75-0.83 in those without CKD (Table 3). For APOL1-HR individuals, 112 had baseline prevalent CKD, of which 31.2% experienced the composite kidney endpoint. In this subgroup, the KidneyIntelX model produced an AUC of 0.88 (95% CI 0.84-0.92) vs. the 386 without baseline CKD, the AUC was 0.79 (95% CI 0.77-0.82; Table 3).

Contemporary Data
Using contemporary data only (data within 1 year prior to enrollment and biomarker measurement), the
discriminatory performance of the KidneyIntelX model in both the T2D (AUC 0.78, 95% CI 0.77-0.80) and
APOL1-HR (AUC 0.79, 95% CI 0.77-0.82) were similar when all available clinical data were available,
providing that KidneyIntelX is not dependent on multi-year history to provide accurate prognostic
information (Table 3).

Random Forest Model With and Without Biomarkers
A newly created random forest model with different clinical features that did not include any of the biomarkers
(TNFR1, TNFR2, or KIM-1) in both cohorts had lower training and test AUCs than the full KidneyIntelX model
with plasma biomarkers and their ratios (Supplementary Table 3).

Discrimination for Individual Components of the Composite Kidney Endpoint
The discriminatory performance of KidneyIntelX (trained for the entire composite endpoint) for the individual
components of the composite endpoint (RKFD alone, sustained 40% decline alone, or kidney failure alone) in
the test cohorts for T2D and APOL1-HR did not vary substantially (Supplementary Table 4).
Time to Event Analyses for 40% Sustained Decline in eGFR or Kidney Failure
Patients with high-risk KidneyIntelX scores (top 23% in T2D and top 18% in APOI1) had a greater risk of progression to time-to-event categorical outcomes of 40% sustained decline or kidney failure than patients in the low- or medium-risk strata combined (hazard ratio (HR) 9.9; 95% CI: 6.7-14.6) and 9.1, 95% CI 5.8-14.3), respectively. Separation of the high-risk strata Kaplan-Meier curve occurred within the first year and progressively declined over time (Figure 2).
DISCUSSION

Utilizing two large cohorts of patients at high-risk for progressive kidney function decline, (T2D and APOL1-HR) with banked plasma samples linked to the corresponding EHR data, we developed a prognostic model combining EHR data and three previously validated plasma biomarkers to predict a composite kidney endpoint, which included RKFD, 40% sustained decline, or kidney failure. The KidneyIntelX prognostic model was more accurate for predicting the risk of kidney function decline than a validated clinical model in this study population. The ability to identify a distinct patient group with the composite kidney endpoint with a PPV of over 55% allows for more appropriate future patient management including nephrologist referral, improved awareness of kidney health, and guidance towards more targeted, intensive therapies to slow progression. The demonstrated PPV in the high-risk strata represents a three-fold improvement over the observed baseline event rate in the two populations.

CKD is a complex, common problem challenging modern healthcare. In the absence of specific therapies, early identification of patients more likely to experience RKFD, and adverse kidney outcomes is paramount. Early identification would help in the allocation of limited resources as well as implementation or intensification of proven interventions to slow kidney function decline. In real world practice, the prediction of kidney disease progression in patients with T2D and or APOL1-HR are challenging, particularly in patients with largely preserved kidney function. There are two major problems contributing to the difficulties in early identification and prediction: 1) serum creatinine/eGFR and urine albumin creatinine are relatively insensitive and non-specific biomarkers, with significant fluctuations and variability in early stages of CKD, and 2) the prevalent standard includes recursive scores incorporating only a single (baseline) value of a selected predictive feature and does not include longitudinal data.

Recently, several biomarkers representing injury and inflammation have been the subject of an intense research focus. Among these, three biomarkers, soluble tumor necrosis factor receptor 1 and 2 (sTNFR1/2) and plasma kidney injury molecule-1 (KIM-1) have been extensively validated in multiple studies to support their translation to clinical use in a range of CKD settings, including patients with and without T2D, as well as APOL1-HR. The individual biomarkers have added a significant improvement to clinical metrics, and the combination of all three, perhaps because of different pathophysiologic pathways, appears to be synergistic. We have demonstrated that combining these biomarkers with clinical information using machine learning techniques can significantly improve the discrimination/prediction of composite kidney endpoints.

Biomarkers that can be measured during a routine clinical encounter can be combined with longitudinal EHR data present in most health care systems for optimal prediction. We have shown previously that the addition of
longitudinal data using supervised machine learning significantly outperforms ‘baseline’ clinical models and also has utility for subtyping disease trajectories. Thus, we hypothesized that combining biomarker information and extant longitudinal EHR data would improve prediction of future kidney progression.

This integrated approach has near-term clinical implications, especially when linked to clinical decision support (CDS) and embedded care pathways within the EMR. For example, patients with a high KidneyIntelX risk score, with a probability of > 50% for the kidney endpoint should be referred to a nephrologist, which has been associated with improved outcomes. In addition, referral to a dietician and delivery of educational materials regarding the importance and consequences of CKD to the high-risk patients should increase awareness and facilitate motivation for changes in lifestyles and behavior. Finally, the optimization of medical therapy including renin-angiotensin aldosterone system inhibitors, statins for cardiovascular risk management, and intensification of antihypertensive medication to meet guideline recommended blood pressure targets can be pursued. The application of sodium glucose transporter (SGLT)-2 inhibitors might also be advantageous in the high KidneyIntelX score group with T2D given recent data on robust renoprotection.

Alternatively, patients with a low-risk score could be clinically managed by their primary care provider and have a standard of care treatment with scheduled monitoring of their KidneyIntelX results. Finally, patients with an intermediate-risk score would be recommended for the standard of care and retesting longitudinally. Such patients may demonstrate changes in KidneyIntelX based on behavioral changes, clinical parameters, and treatment adjustments over time with appropriate clinical actions as necessary. This overall approach would not only benefit individual patient outcomes but also positively impact on health systems where there is uncertainty about which patients to refer to a limited number of subspecialists.

Our study is not without limitations. Although we used a multiethnic dataset, further validation in geographically diverse populations is necessary. Secondly, data structures/relationships change over time due to adjusted practice patterns and thus the algorithm may not perform similarly if the full complement of longitudinal data is not available or if the clinical practice is altered. We conducted a sensitivity analysis with only one year of data available prior to baseline, and the loss of performance was minimal; however, this should be evaluated further in additional validation studies. We have imputed the missing values for some features using different imputation methods which can have an effect on the correlations between features in the random forest model. Since the training and testing were performed within a single cohort, the potential for overfitting exists. However, we plan to expand our analysis to independent cohorts for external validation before prospective testing. Finally, this analysis does not address implementation or utility. Therefore, a prospective clinical utility trial to assess the decision impact of the KidneyIntelX risk score when provided to primary care physicians and patients is underway.
In conclusion, we have demonstrated that machine learning techniques (random forest models) combining longitudinal EHR information with three plasma biomarkers improved prediction of rapid kidney function decline or kidney failure over validated clinical models in two distinct clinical settings and populations. With the advent of advanced high-performance computing, validated biomarkers, and an integrated EHR, the ability to improve outcomes with the integration of the KidneyIntelX into routine patient care should be assessed.

Author Contributions: Drs. Chauhan, Nadkarni, and Coca had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

K Chauhan: Formal analysis; Methodology; Writing - original draft; Writing - review and editing
G Nadkarni: Conceptualization; Formal analysis; Investigation; Methodology; Supervision; Writing - original draft; Writing - review and editing
F Fleming: Conceptualization; Formal analysis; Funding acquisition; Supervision; Writing - original draft; Writing - review and editing
J McCullough: Conceptualization; Supervision
J He: Writing - review and editing
J Quackenbush: Writing - review and editing
B Murphy: Writing - review and editing
M Donovan: Formal analysis; Funding acquisition; Supervision; Writing - original draft
S Coca: Conceptualization; Formal analysis; Funding acquisition; Supervision; Writing - original draft; Writing - review and editing
J Bonventre: Writing - review and editing

Disclosures:

GNN, MB, CJH, JQ, and SGC receive financial compensation as consultants and advisory board members for RenalytixAI, Inc., and own equity in Renalytix. JBV is a non-paid advisory board member for Renalytix. GNN and SGC are scientific co-founders of RenalytixAI. FF, JRM, BM, MD are officers of RenalytixAI.

SGC has received consulting fees from CHF Solutions, Takeda Pharmaceuticals, Relypsa, Bayer, and Boehringer-Ingelheim in the past three years. GNN has received operational funding from Goldfinch Bio and consulting fees from BioVie Inc and GLG consulting in the past three years. JVB is a co-inventor on KIM-1 patents assigned to Partners Healthcare. He is a consultant to Cadent, Praxis and Seattle Genetics, Aldeyra, Angion and owns equity in Goldfinch, Innoviva, MediBeacon, DxNow, Verinano, Sensor Kinesis, and Sentien.

Funding:

This research was supported, in part, by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; R01DK096549 to SGC). GNN is supported by a career development award from the National Institutes of Health (NIH) (K23DK107908) and is also supported by R01DK108803, U01HG007278, U01HG009610, and 1U01DK116100. SGC and JVB are members and are supported in part by the Chronic Kidney Disease Biomarker Consortium (U01DK106962). SGC is also supported by the following grants: R01DK115562, R01HL85757, R01DK112258, U01OH011326, and R01DK126477. JVB is also supported by NIH R37 DK39773 and RO1 DK072381.
REFERENCES:


<table>
<thead>
<tr>
<th>Table 1. Clinical Characteristics at Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Clinical Characteristics</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age in years, Median (IQR)</td>
</tr>
<tr>
<td>Female, n (%)</td>
</tr>
<tr>
<td>Race</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>African Ancestry</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Body Mass Index in kg/m², Median [IQR]</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
</tr>
<tr>
<td>Coronary Artery Disease, n (%)</td>
</tr>
<tr>
<td>Heart Failure, n (%)</td>
</tr>
<tr>
<td>Systolic BP in mm Hg, Median [IQR]</td>
</tr>
<tr>
<td>Diastolic BP in mm Hg, Median [IQR]</td>
</tr>
<tr>
<td>Follow up Time in years, Median [IQR]</td>
</tr>
<tr>
<td><strong>Laboratory Characteristics</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Baseline eGFR in ml/min/1.73 m², Median [IQR]</td>
</tr>
<tr>
<td>Baseline UACR in mg/g, Median [IQR]</td>
</tr>
<tr>
<td>UACR available, n (%)</td>
</tr>
<tr>
<td>Baseline Hemoglobin A1C, Median [IQR]</td>
</tr>
<tr>
<td><strong>Medications</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ACE/ARB at baseline, n (%)</td>
</tr>
<tr>
<td><strong>Plasma Biomarker Concentrations</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>TNFR1, in pg/ml, Median [IQR]</td>
</tr>
<tr>
<td>TNFR2, in pg/ml, Median [IQR]</td>
</tr>
<tr>
<td>KIM1, in pg/ml, Median [IQR]</td>
</tr>
</tbody>
</table>
Table 2. KidneyIntelX Thresholds for the Composite Kidney Endpoint with Sensitivity, Specificity, PPV and NPV for Type 2 DM and APOL1 High-Risk Populations in High- and Low-Risk Strata

<table>
<thead>
<tr>
<th></th>
<th>Risk Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T2D KidneyIntelX</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom 50%</td>
<td>0.192</td>
<td>0.82</td>
<td>0.59</td>
<td>0.38</td>
<td>0.92</td>
</tr>
<tr>
<td>Top 20%</td>
<td>0.444</td>
<td>0.50</td>
<td>0.89</td>
<td>0.58</td>
<td>0.86</td>
</tr>
<tr>
<td>Top 15%</td>
<td>0.555</td>
<td>0.40</td>
<td>0.93</td>
<td>0.62</td>
<td>0.84</td>
</tr>
<tr>
<td>Top 10%</td>
<td>0.707</td>
<td>0.29</td>
<td>0.96</td>
<td>0.68</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>T2D Clinical Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom 50%</td>
<td>0.148</td>
<td>0.68</td>
<td>0.55</td>
<td>0.31</td>
<td>0.85</td>
</tr>
<tr>
<td>Top 20%</td>
<td>0.240</td>
<td>0.38</td>
<td>0.85</td>
<td>0.43</td>
<td>0.82</td>
</tr>
<tr>
<td>Top 15%</td>
<td>0.278</td>
<td>0.30</td>
<td>0.89</td>
<td>0.46</td>
<td>0.81</td>
</tr>
<tr>
<td>Top 10%</td>
<td>0.338</td>
<td>0.23</td>
<td>0.94</td>
<td>0.54</td>
<td>0.80</td>
</tr>
<tr>
<td><strong>APOL1-HR KidneyIntelX</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom 50%</td>
<td>0.209</td>
<td>0.88</td>
<td>0.58</td>
<td>0.32</td>
<td>0.96</td>
</tr>
<tr>
<td>Top 20%</td>
<td>0.438</td>
<td>0.60</td>
<td>0.89</td>
<td>0.56</td>
<td>0.91</td>
</tr>
<tr>
<td>Top 15%</td>
<td>0.489</td>
<td>0.52</td>
<td>0.93</td>
<td>0.62</td>
<td>0.90</td>
</tr>
<tr>
<td>Top 10%</td>
<td>0.546</td>
<td>0.36</td>
<td>0.96</td>
<td>0.66</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>APOL1-HR Clinical Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom 50%</td>
<td>0.151</td>
<td>0.79</td>
<td>0.57</td>
<td>0.29</td>
<td>0.93</td>
</tr>
<tr>
<td>Top 20%</td>
<td>0.322</td>
<td>0.42</td>
<td>0.85</td>
<td>0.38</td>
<td>0.87</td>
</tr>
<tr>
<td>Top 15%</td>
<td>0.387</td>
<td>0.32</td>
<td>0.87</td>
<td>0.39</td>
<td>0.85</td>
</tr>
<tr>
<td>Top 10%</td>
<td>0.448</td>
<td>0.22</td>
<td>0.93</td>
<td>0.4</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Table 3. AUCs (95% CI) for KidneyIntelX vs. Clinical Model for Subgroups

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>KidneyIntelX</th>
<th>Clinical Model</th>
<th>Subgroup</th>
<th>KidneyIntelX</th>
<th>Clinical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM2</td>
<td></td>
<td></td>
<td>APOL1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalent CKD (n=366)</td>
<td>0.84 (0.81 - 0.87)</td>
<td>0.70 (0.69 - 0.71)</td>
<td>Prevalent CKD (n=112)</td>
<td>0.88 (0.84 - 0.92)</td>
<td>0.59 (0.56 - 0.61)</td>
</tr>
<tr>
<td>No CKD (n= 505)</td>
<td>0.79 (0.75 - 0.83)</td>
<td>0.63 (0.61 - 0.64)</td>
<td>No CKD (n= 386)</td>
<td>0.79 (0.76 - 0.83)</td>
<td>0.74 (0.73 - 0.75)</td>
</tr>
<tr>
<td>Contemporary Data (n=871)</td>
<td>0.78 (0.77 - 0.80)</td>
<td>0.66 (0.65 - 0.67)</td>
<td>Contemporary Data (n=498)</td>
<td>0.79 (0.77 - 0.82)</td>
<td>0.72 (0.71 - 0.73)</td>
</tr>
</tbody>
</table>
Figure 1a. Proportion with the Composite Kidney Endpoint by Deciles of Predicted Risk via KidneyIntelX vs. Clinical Model in Type 2 DM
Figure 1b. Proportion with the Composite Kidney Endpoint by Deciles of Predicted Risk via KidneyIntelX vs. Clinical Model in APOL1
Figure 2a. Kaplan-Meier Curves by KidneyIntelX Risk Strata for the Endpoint of Sustained 40% Decline in eGFR or Kidney Failure in Type 2 DM

Risk Category | Low Risk | Intermediate Risk | High Risk
---|---|---|---
No. at Risk  | 436 | 435 | 432 | 424 | 412 | 373 | 332
Low Risk     | 304 | 300 | 287 | 258 | 246 | 220 | 181
Intermediate Risk | 131 | 117 | 102 | 90 | 73 | 65 | 43

Proportion of population: Low risk: 50%, Intermediate risk: 35%, High Risk: 15%
Hazard ratio for high vs. low risk: 16.8, 95% CI 9.7-29.3
Hazard ratio for high vs. low and intermediate risk combined: 9.9; 95% CI: 6.7-14.6
Figure 2b. Kaplan-Meier Curves by KidneyIntelX Risk Strata for the Endpoint of Sustained 40% Decline in eGFR or Kidney Failure in APOL1 High-Risk

![Graph showing Kaplan-Meier curves for different risk categories.](image)

### No. at Risk

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Risk</td>
<td>249</td>
<td>238</td>
<td>233</td>
<td>222</td>
<td>214</td>
<td>203</td>
<td>189</td>
</tr>
<tr>
<td>Intermediate Risk</td>
<td>174</td>
<td>169</td>
<td>164</td>
<td>155</td>
<td>144</td>
<td>133</td>
<td>126</td>
</tr>
<tr>
<td>High Risk</td>
<td>75</td>
<td>67</td>
<td>61</td>
<td>48</td>
<td>43</td>
<td>36</td>
<td>31</td>
</tr>
</tbody>
</table>

Proportion of population: Low risk: 50%, Intermediate risk: 35%, High Risk: 15%

Hazard ratio for high vs. low risk: 20.2, 95% CI 9.8-41.2

Hazard ratio for high vs. low and intermediate risk combined: 9.1, 95% CI 5.8-14.3