Early Detection of Chronic Kidney Disease in Patients with Obesity: A Predictive Modeling Approach for Preventive Care

Shashi Khan, Vikash Verma, Louis Brooks Jr, Marissa Seligman, Abhimanyu Roy, Abhinav Nayyar, Ankit Arora, Ankita Misra, Shruti Chaudhary, Pooja Sharma, Varshith Gandla, Vishan Khatavkar, Vaibhav Bansal, Anuj Gupta, Kavita Karayat, Srishti Motila, Tisha Sharma

Background

- Chronic kidney disease (CKD) is a major contributor to morbidity, healthcare utilization and costs.[1] Obesity is a well-recognized clinical characteristic and independent risk factor of patients with CKD and is associated with metabolic and hemodynamic changes that may influence disease risk and clinical recognition.[2]
- Predictive models applied to real-world data may help identify patients with obesity who are more likely to have CKD; however, evidence on obesity-specific risk factor patterns remains limited.[3]

Objective

- Develop and evaluate a machine-learning (ML) model to identify patients with obesity at risk for CKD using real world claims and EHR data.
- Exploratory: Explore obesity-specific risk indicators by comparing CKD risk factor importance in a non-obese cohort.

Methodology

- **Study Design:** Retrospective matched case–control study using Optum® Market Clarity claims and EHR data (study period: January 2016–June 2025).
- **Study Population and cohorts**: Adults aged 18–65 years with obesity (ICD-10 codes) and ≥1 renal laboratory measurement during the 12-month baseline period. CKD cases were defined as a documented CKD diagnosis (stages 1–5) at index, and no known baseline CKD.
- **Case–Control Definition and Matching: Cases:** Patients with a CKD diagnosis indexed on the first observed CKD date. **Controls:** Patients without CKD, assigned an index date aligned to the matched case. Cases and controls were matched 1:1 using propensity scores based on age, sex, and Charlson Comorbidity Index (CCI) and assigned a matching index date. Covariate balance was assessed using standardized mean differences.
- **Cohorts**: Two analytic cohorts (CKD vs non-CKD) were evaluated separately: **Primary Model 1**: Obese patients, **Exploratory Model 2**: Non-obese patients
- **Baseline Predictors:**
  - Clinical factors, including comorbidities such as hypertension (HTN), type 2 diabetes (T2D) etc.
  - Healthcare utilization: prior nephrologist visits, medication use – obesity management medications, diagnostic testing etc. in the 12-month baseline period.
- **Modeling**: Logistic Regression (LR) and XGBoost models were developed separately for each cohort to predict CKD status vs. matched non-CKD controls, using a 75/25 train–test split. Random Forest models were explored during preliminary analyses but were not included in the final results.
- **Model Evaluation:** Assessed using area under the ROC curve (AUC), F1-score, precision, and recall. Odds ratios (ORs) from LR were examined to identify key CKD risk indicators among obese patients, and to assess differences by obesity status.

Results

Study Population: A total of 22,982 patients were included in the study.

- Obese cohort: 10,088 patients (Cases: 5,044; Controls: 5,044)
- Exploratory non-obese cohort: 12,894 patients (Cases: 6,447; Controls: 6,447)

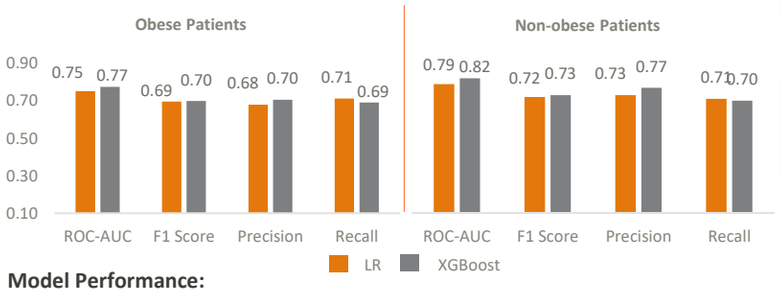
Patient characteristics:

- In both CKD and non-CKD cohorts were well balanced on age, sex, and CCI, post matching.

Feature Selection

- Variance Inflation Factor analysis was used to assess multicollinearity among candidate predictors, retaining 22 features with acceptable collinearity for CKD prediction modeling.

Figure 1. Comparison of Model Performance



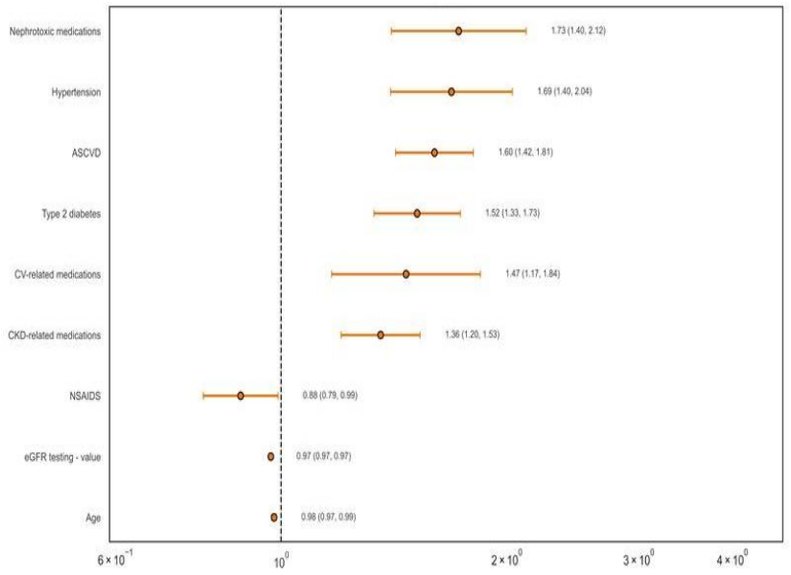
Model Performance:

- For patients with obesity, LR yielded AUC of 0.75 (F1 score=0.69), while XGBoost achieved AUC of 0.77 (F1 score=0.70). Precision, recall and accuracy were balanced (0.68–0.75) across models. (Table 1).
- In the non-obese population, performance was higher: LR AUC of 0.79 (F1 score=0.72) vs XGBoost AUC of 0.82 (F1 score= 0.73).
- Both LR and XGBoost models achieved moderate discriminative performance in distinguishing CKD cases from non-CKD controls, with XGBoost performing slightly better.

Top CKD risk indicators

- **Obese patients:** Top CKD predictors of CKD included HTN, T2D, ASCVD, and nephrotoxic medication use, CKD-related medications. CV medication use also showed a moderate association. African American race was associated with lower odds (OR 0.73). Other factors, including NSAID use (OR 0.87) and laboratory measures, showed smaller effects.
- **Obese vs. Non-obese patients (Exploratory):** Across obese and non-obese cohorts, common predictors included hypertension (OR 1.74), type 2 diabetes (OR 1.82), ASCVD (OR 1.58), CKD-related medications (OR 1.30), and nephrotoxic medication use (OR 1.89), indicating consistent clinical drivers. In the non-obese cohort, NSAID use (OR 0.78) and geographic region (e.g., Midwest OR ~0.80) were more prominent, suggesting potential heterogeneity. Cardiovascular medication use showed relatively stronger associations in non-obese patients (OR 1.44).

Figure 2. Forest Plot for Obese Cohort (p<0.05)



Odds Ratio (Log Scale)

Conclusions

- Machine learning and traditional regression models demonstrated moderate and broadly comparable discrimination for predicting CKD risk, with relatively improved performance observed for XGBoost. Comorbidities such as hypertension and type 2 diabetes, along with CKD-specific clinical indicators and healthcare utilization patterns, were among the most informative predictors, supporting their relevance for early identification in the obese population. Observed differences in predictor odds ratios between obese and non-obese cohorts may reflect variation in predictor patterns by obesity status.
- This retrospective matched case–control study identified baseline factors associated with CKD identification rather than causal effects. These study findings may be influenced by healthcare utilization and diagnostic practices proximal to diagnosis, residual confounding, and limited generalizability.
- Future research should focus on prospective validation and earlier prediction of CKD onset and progression over time.

References: 1. GBD 2023 IHD & Dietary Risk Factors Collaborators. Nat Med. 2026;32:1454–1478. 2. Kounatidis D et al. Curr Obes Rep. 2024;13:680–702. 3. Zhang H et al. Lipids Health Dis. 2024;23:57. CCI: Charlson Comorbidity Index, LR: Logistic regression, XGBoost: extreme gradient boosting

Optum