

Development and Validation of a Predictive Model to Identify Patients With Undiagnosed Chronic Kidney Disease

Steven M. Brunelli MD, MSCE; Adam G. Walker PhD; Kathryn Gray, MS
DaVita Clinical Research, Minneapolis, Minnesota, USA

Disclosures

- SMB, AGW, and KS are employees of DaVita Clinical Research
 - SMB's spouse is an employee of AstraZeneca

Background

- Chronic kidney disease (CKD) is a common condition and often goes undiagnosed.
 - Unmanaged CKD can progress rapidly, resulting in poor clinical outcomes and increased health care costs.
 - The ability of health systems to identify individuals with undiagnosed CKD could allow for implementation of CKD management practices in order to slow progression, potentially improving outcomes and reducing health care costs.
- Here, we report the development and validation of a claims-based algorithm to identify CKD.

Methods: Data Source and Patients

- This model was developed using Medicare Part A and B claims from calendar year 2017.
- This model was developed and tested using data from **378,460** patients split among two sets:
 - **189,203** patients were used to *derive the model*
 - **189,257** patients were used to *validate the model*
- Inclusion:
 - ≥ 18 years old
 - Medicare Part A and B insurance coverage
 - No diagnosis of ESRD or claims for dialysis between 01 January- 30 April 2017

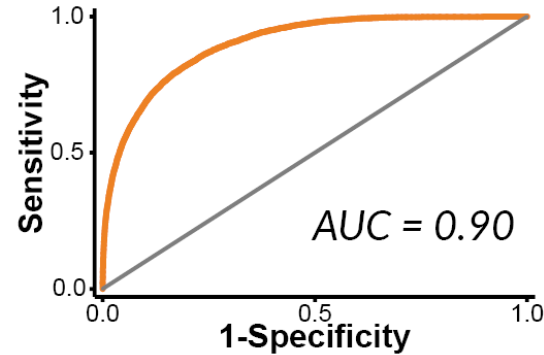
Methods: Modeling Approach



- The predicted outcome was the presence of a diagnosis code for CKD stages 3 to 5, which occurred in 4.4% of patients within the data.
 - To simulate the use case, codes for kidney disease were not eligible as predictors in the model.
- Area under the curve (AUC) of the receiver operating curve and positive predictive value (PPV) were used to assess the performance of candidate models.

Results

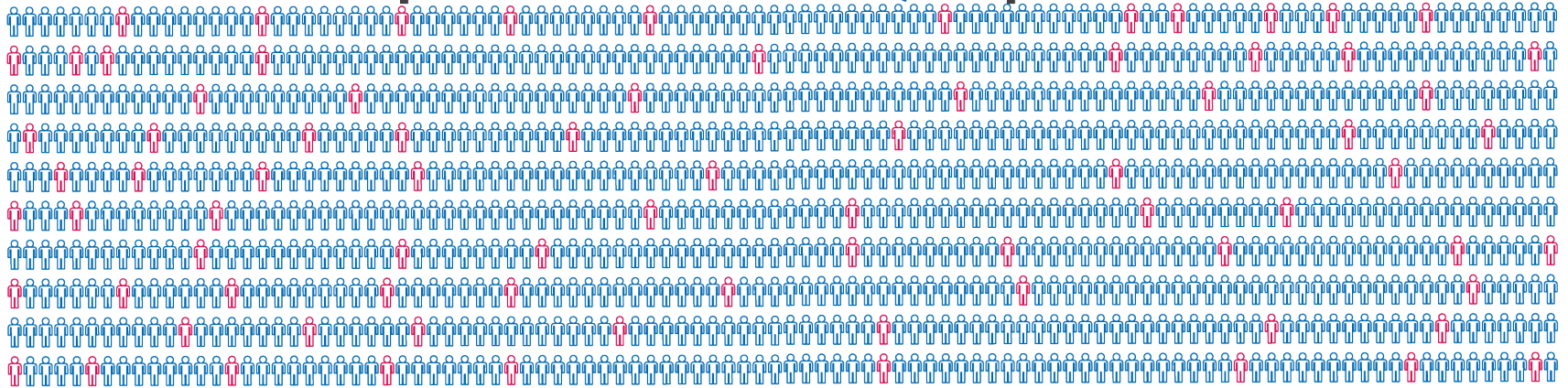
	Derivation Set	Validation Set
Patients	189,203	189,257
CKD 3-4 Prevalence	4.3%	4.4%
AUC	0.90	0.90



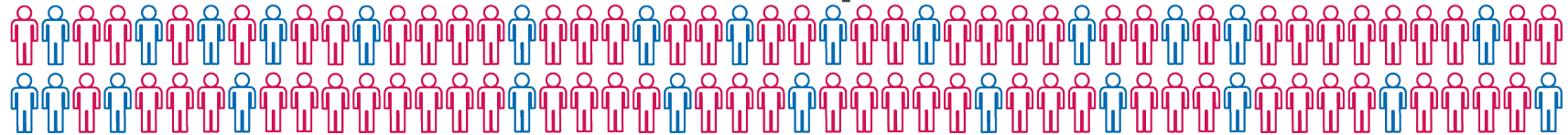
- The best model was a logistic regression algorithm based on 94 input terms derived from 13 clinical constructs.
- The model demonstrated an excellent ability to discriminate (AUC = 0.90), which was stable when tested in the validation set (AUC = 0.90).
- The PPV in the top 1% and top 2% of patients identified by the model was approximately 72% and 59%, respectively.

Hypothetic Use Case: 1% Threshold

If the model screened **10,000** patients

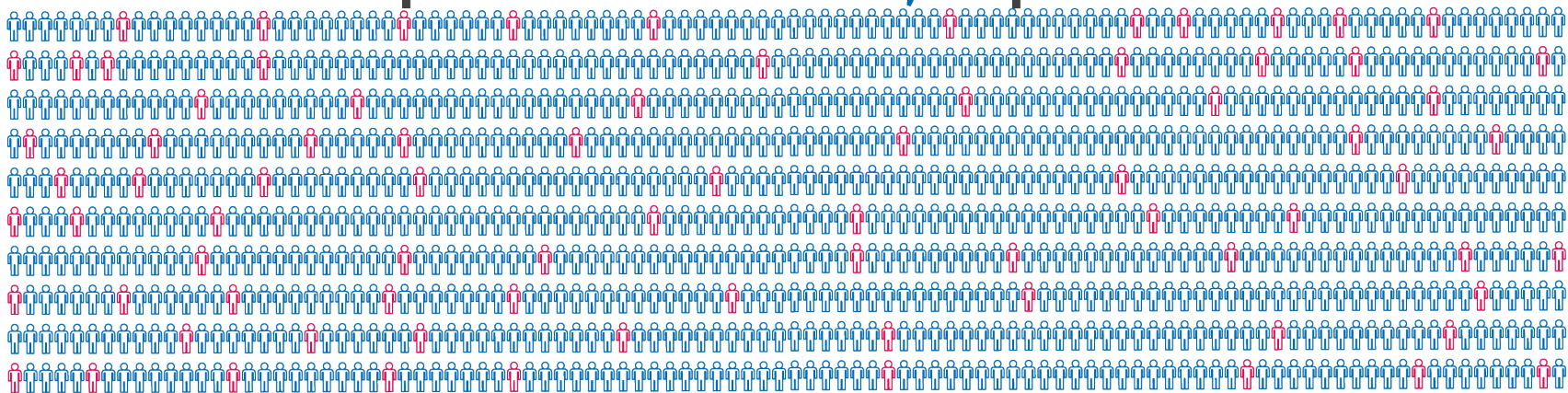


100 would be called out, **72** of whom would have **CKD**

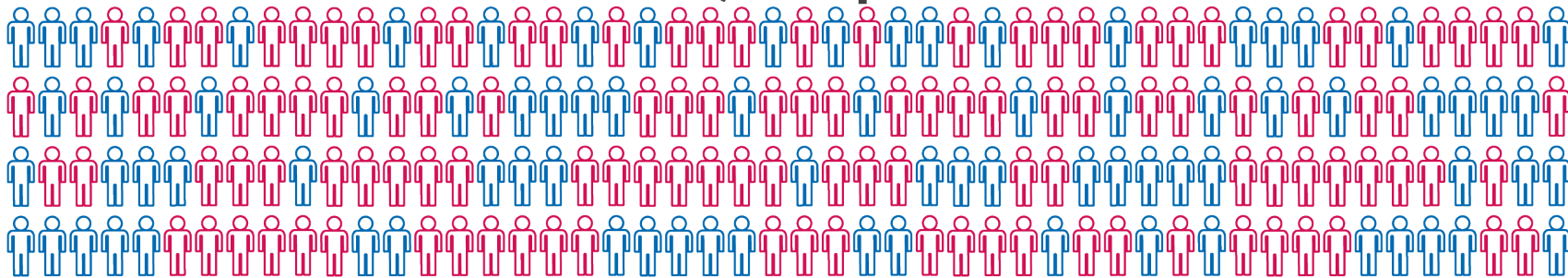


Hypothetic Use Case: 2% Threshold

If the model screened **10,000** patients



200 would be called out, **118** of whom would have **CKD**



Conclusion

- We developed an algorithm that can use administrative data that could be applied on a health system basis that may identify patients with undiagnosed CKD
- The model is highly discriminant (AUC = 0.90) with PPV 59-72% in typical use cases