Kidney360 Publish Ahead of Print, published on September 16, 2021 as doi:10.34067/KID.0000852021

Kidney360

American Society of Nephrology
1401 H St NW, Suite 900
Washington, DC 20005
Phone: 202-640-4660 | Fax 202-637-9793
vramsey@kidney360.org

Article Type: Original Investigation

A Computable Phenotype for Autosomal Dominant Polycystic Kidney Disease

DOI: 10.34067/KID.0000852021

Mohamad Kalot, Abdallah Alayli, Mohammed Al Khatib, Nedaa Husainat, Kerri McGreal, Diana Jalal, Alan Yu, and Reem Mustafa

Key Points:
*The ADPKD computable phenotype based on ICD 9/10 is an excellent screening tool to identify patients with ADPKD.

*Patients who were followed in nephrology clinics had a higher sensitivity, specificity, positive and negative predictive values.

*Specificity of the ADPKD computable phenotype is comparable to other medical conditions.

Abstract:
Background: A computable phenotype is an algorithm used to identify a group of patients within an electronic medical record system. Developing a computable phenotype that can accurately identify Autosomal Dominant Polycystic kidney disease (ADPKD) patients will assist researchers in defining patients eligible to participate clinical trials and other studies. Our objective was to assess the accuracy of a computable phenotype using ICD (International Classification of Diseases) 9 and 10 codes (ICD-9/10) to identify patients with ADPKD. Methods: We reviewed four random samples of approximately 250 patients based on ICD-9/10 codes from the EHR from the Kansas University Medical Center database: patients followed in nephrology clinics who had ICD-9/10 codes for ADPKD (Neph+), patients seen in nephrology clinics without ICD codes of ADPKD (Neph-), patients who were not followed in nephrology clinics with ICD codes for ADPKD (No Neph+), and patients not seen in nephrology clinics without ICD codes for ADPKD (No Neph-). We reviewed charts and determined ADPKD status based on internationally accepted diagnostic criteria for ADPKD. Results: The computable phenotype to identify patients with ADPKD who attended nephrology clinics has a sensitivity of 98.7% (95% confidence interval (95% CI); 96.4-99.7), and a specificity of 84.1% (95% CI; 79.5-88.1). For those who did not attend nephrology clinics the sensitivity was 97.1% (95% CI; 93.3-99.0), and a specificity was 82.0% (95% CI; 77.4-86.1). Conclusion: A computable phenotype using the ICD-9/10 codes can correctly identify most patients with ADPKD and can be utilized by researchers to screen healthcare records for ADPKD patient cohorts with acceptable accuracy.

Disclosures: K. McGreal reports the following: Medical Center Research Funding: Reata- PI on FAlcon study, Sanofi - Staged PKD trial -sub-I. D. Jalal reports the following: Research Funding: AstraZenica, Corvidia; Honoraria: K-INBRE, Reata; Scientific Advisor or Membership: Reata. A. Yu reports the following: Consultancy Agreements: Regulus Therapeutics, Calico, Otsuka, Navitor; Ownership Interest: Amgen Corp., Gilead Sciences, Prothena; Honoraria: Elsevier, Wolters Kluwer; Scientific Advisor or Membership: Otsuka Advisory Board; Other Interests/Relationships: The University of Kansas Medical Center and the Jared Grantham Kidney Institute receives royalties from Otsuka for tolvaptan. R. Mustafa reports the following: Scientific Advisor or Membership: The GRADE guidance group, The American College of Physicians Clinical Guideline Committee, The Canadian Society of Nephrology Clinical Practice Guidelines Committee; Other Interests/Relationships: Advisory board for the renal round table, NKF Midewest. The remaining authors have nothing to disclose.

Funding: Kansas PKD Research and Translation Core Center:, P30 DK106912; CTSA Award:, UL1TR002366; HERON

Author Contributions: Mohamad Kalot: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing - original draft; Writing - review and editing Abdallah Alayli: Data curation; Formal analysis; Investigation; Writing - original draft; Writing - review and editing Mohammed Al Khatib: Data curation; Formal analysis; Investigation; Writing - original draft Nedaa Husainat: Data curation; Formal analysis; Writing - original draft Kerri McGreal: Data curation; Investigation; Methodology; Writing - original draft; Writing - review and editing Diana Jalal: Conceptualization; Formal analysis; Investigation; Methodology; Writing - review and editing Alan Yu: Formal analysis; Investigation; Methodology; Writing - review and editing Reem Mustafa: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Supervision; Validation; Writing - original draft; Writing - review and editing

Clinical Trials Registration: No

Registration Number:

Registration Date:

How to Cite this article: Mohamad Kalot, Abdallah Alayli, Mohammed Al Khatib, Nedaa Husainat, Kerri McGreal, Diana Jalal, Alan Yu, and Reem Mustafa, A Computable Phenotype for Autosomal Dominant Polycystic Kidney Disease, Kidney360, Publish Ahead of Print, 10.34067/KID.0000852021

A Computable Phenotype for Autosomal Dominant Polycystic Kidney Disease

Mohamad A. Kalot[1], Abdallah El Alayli[2], Mohammad Al Khatib[3], Nedaa Husainat[4], Kerri McGreal[2], Diana I. Jalal[5], Alan S. L. Yu[2], Reem A. Mustafa[2]

[1] Department of Internal Medicine, State University of New York at Buffalo, Buffalo, New York, USA

[2] University of Kansas Medical Center; Department of Internal Medicine; Division of Nephrology and Hypertension and the Jared Grantham Kidney Institute, Kansas City, Kansas, USA

[3] Hamad Medical Corporation; Department of Internal Medicine; Doha, Qatar

[4] St. Mary's Hospital, Department of Internal Medicine, 6420 Clayton Rd Saint Louis, Missouri, USA

[5] University of Iowa Health Care; Department of Internal Medicine; Division of Nephrology, Iowa City, Iowa, USA

Corresponding author:
Reem A. Mustafa MD, MPH, PhD
University of Kansas Medical Center, Division of Nephrology
Kansas City, KS 66160, USA
Phone: 913 588 6048
Fax: 913 588 3867
Email: rmustafa@kumc.edu

KEY POINTS

- The ADPKD computable phenotype based on ICD 9/10 is an excellent screening tool to identify patients with ADPKD.
- Patients who were followed in nephrology clinics had a higher sensitivity, specificity, positive and negative predictive values.
- Specificity of the ADPKD computable phenotype is comparable to other medical conditions.

ABSTRACT

Background: A computable phenotype is an algorithm used to identify a group of patients within an electronic medical record system. Developing a computable phenotype that can accurately identify Autosomal Dominant Polycystic kidney disease (ADPKD) patients will assist researchers in defining patients eligible to participate clinical trials and other studies. Our objective was to assess the accuracy of a computable phenotype using ICD (International Classification of Diseases) 9 and 10 codes (ICD-9/10) to identify patients with ADPKD.

Methods: We reviewed four random samples of approximately 250 patients based on ICD-9/10 codes from the EHR from the Kansas University Medical Center database: patients followed in nephrology clinics who had ICD-9/10 codes for ADPKD (Neph+), patients seen in nephrology clinics without ICD codes of ADPKD (Neph-), patients who were not followed in nephrology clinics with ICD codes for ADPKD (No Neph+), and patients not seen in nephrology clinics

without ICD codes for ADPKD (No Neph-). We reviewed charts and determined ADPKD status based on internationally accepted diagnostic criteria for ADPKD.

Results: The computable phenotype to identify patients with ADPKD who attended nephrology clinics has a sensitivity of 98.7% (95% confidence interval (95% CI); 96.4-99.7), and a specificity of 84.1% (95% CI; 79.5-88.1). For those who did not attend nephrology clinics the sensitivity was 97.1% (95% CI; 93.3-99.0), and a specificity was 82.0% (95% CI; 77.4-86.1).

Conclusion: A computable phenotype using the ICD-9/10 codes can correctly identify most patients with ADPKD and can be utilized by researchers to screen healthcare records for ADPKD patient cohorts with acceptable accuracy.

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is a genetic condition that causes bilateral renal cyst formation.[1] This disease is the most common hereditary kidney disease and it affects 1 in 400-1,000 people worldwide.[2] Half of these patients will require end-stage kidney disease (ESKD) management.[3] Although genetic testing is emerging in the detection of ADPKD,[4] genetic testing is limited by the number of missense mutations that need further confirmation and the diagnosis is routinely confirmed based on imaging studies such as CT scan, MRI and abdominal ultrasound (US), which carry the highest sensitivity in detecting ADPKD. Among these three techniques, abdominal ultrasound is the most widely used because of its low cost, availability, high sensitivity with advanced disease and its safety.[5]

In the presence of healthcare systems that utilize electronic health records (EHR), a large amount of stored data could be easily accessed. This setting is ideal to using a computable phenotype to identify different conditions/disease including ADPKD. A computable phenotype is a clinical

phenotype, a characteristic or a group of several clinical features that can be automatically extracted from an EHR without healthcare providers interpretation or intervention [6]. In this way, computable phenotype may provide a reliable, and easy way to identify patients with the condition of interest in a timely manner.[7]

The World Health Organization (WHO) created the ICD (International Classification of Diseases) in the 1948.[8] The coding professionals transformed medical terms, procedures, and diagnoses into universal alphanumeric codes. ICD-10 is the most recent update and was adopted in the United States in 2013.[9] ICD codes were implemented to promote international comparability for collection, classification, processing, and presentation of health statistics.[10] Since ADPKD is a relatively rare disease, using the ICD-9/10 codes to find all patients with the disease in an EHR would be very time and cost effective during recruitment for ADPKD studies.[11] However, ICD codes are often inaccurate and there is a considerable controversy regarding their value in identifying patients with specific clinical conditions. For this reason, we sought to assess test accuracy (sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV)), and Likelihood ratio positive and negative of a computable phenotype using ICD-9/10 codes in identifying patients with ADPKD.

We developed a computable phenotype using ICD-9/10 to identify patients with ADPKD. In this study, we present the test accuracy results of the computable phenotype (sensitivity, specificity, PPV, and NPV) in identifying patients with ADPKD, for patients who follow up in nephrology clinics and those who do not. Additionally, we estimate the prevalence of ADPKD using the University of Kansas Medical Center (KUMC) database.

METHODS

Study design and participants

We conducted a cross sectional test accuracy study by reviewing four random samples of approximately 250 patients based on ICD-9/10 codes and nephrology clinic visit, from the EHR from the KUMC database. The samples were stratified into four groups: Group Neph+ and Neph- included patients followed in nephrology clinics who had ICD-9/10 codes for ADPKD and those who did not have ICD-9/10 codes of ADPKD respectively. Group No-neph + and No-neph- included patients who were not followed in nephrology clinics, with and without ICD-9/10 codes for ADPKD respectively.

Test methods

We used de-identified patient information in HERON data repository, an i2b2 data access platform[12,13] to identify patients with ADPKD using ICD-9 codes 753.12 and 753.13, and ICD-10 codes Q61.2 and Q61.3. We used the ICD-9 code 593.2 and the ICD-10 code N28.1 to label patients with renal cysts that are not ADPKD and to enrich the sample of patients without ADPKD.

We evaluated four random samples from the de-identified dataset based on the eligibility criteria for each group. At least two reviewers reviewed the medical records of each patient, in duplicate. Although we reviewed the charts after we generated the computable phenotype, the reviewers were blinded to the strata based on the phenotype results. For every chart, each of the reviewers had to make a determination of whether the patient has or does not have ADPKD. In patients with family history of the disease, we used the unified imaging diagnosis criteria and in patients with no family history of ADPKD, the diagnosis was made if the patient has at least 10 cysts in

each of the 2 kidneys, with kidneys measuring more than 13 cm in length (Table 1) [14]. We used data from the last imaging available. When ADPKD status was still ambiguous, an experienced nephrologist reviewed all medical records to decide whether ADPKD is present or not according to clinical criteria. If there was insufficient information to decide whether ADPKD was present or not, we excluded patients from the analysis. We did not consider the timing of insertion of the ICD 9/10 codes into the system.

Analysis

After reviewing the charts, we collected the data and developed two separate 2X2 contingency tables after classifying patients into those with and without ADPKD and those with and without the computable phenotype. We split the results for those who attended and those who did not attend nephrology clinic. We calculated the sensitivity, specificity, negative and positive predictive values with 95% confidence intervals (95% CI). Confidence intervals for sensitivity and specificity are "exact" Clopper-Pearson CI.[15] Confidence intervals for the likelihood ratios are calculated using the "Log method" as described by Altman et al.[16] We conducted a sensitivity analyses to determine whether classifying the excluded patients as having ADPKD, or as not having ADPKD, meaningfully changed the results. We calculated the prevalence of ADPKD in the available electronic medical records of the healthcare system. We have reported the results based on the STARD guidelines for reporting diagnostic accuracy studies.[17]

RESULTS

Our random sample included a total of 1071 patients, of which 536 were followed in nephrology clinic and 535 did not follow in nephrology clinic. The average age of the patients was 63 years,

53% were males, 76% were white, and 15% were black or African American. The prevalence of ADPKD based on positive ICD-9/10 codes in the deidentified dataset is 6/10000. Descriptive analysis of age, gender, and race were provided in Tables 1 and 2.

In the group Neph+, we found that 236 truly had ADPKD out of the 283 patients who had an ICD-9/10 codes for ADPKD. In the group Neph-, 249 patients truly did not have ADPKD out of 253 patients who did not have an ICD-9/10 codes for ADPKD. One patient had a “likely no” diagnosis (Figure 1). The specificity did not change when considering the patient with “likely no” diagnosis as a true negative (84.1%).

For group No-neph+, we found that 165 patients were correctly diagnosed with ADPKD out of the 223 patients that were diagnosed with ADPKD using the ICD-9/10 codes, 2 patients were diagnosed as “likely yes”, 2 patients as “likely no”, and 34 patients were classified as unknown due to absence of family history and renal imaging. As for group No-neph-, we found that 265 didn’t have the disease out of the 275 patients who did not have the ICD-9/10 codes for ADPKD, 2 patients were diagnosed as “likely yes”, and 2 patients were diagnosed as “likely no” (Figure 2). The sensitivity did not change when considering the patients with “likely yes” diagnosis as a true positives (97.1%). The specificity did not change when considering the patients with “likely no” diagnosis as true negatives (82.3%). Tables 4 and 5 summarize the 2 X 2 contingency tables and Table 6 summarizes the test accuracy for those followed and those who were not followed in nephrology clinic.

DISCUSSION

Computable phenotypes are efficient to screen patients for a condition of interest, in this study being ADPKD. Computable phenotypes like any other diagnostic technique should be accurate and easy to use. The accuracy could be evaluated using sensitivity, specificity, PPV, and NPV [18]. In this study, we calculated the test accuracy values for a computable phenotype comprised of ICD-9 (753.12 and 753.13) and the ICD-10 codes (Q61.2 and Q61.3) in identifying patients with ADPKD in the KUMC EHR. Overall, we found that the computable phenotype had an excellent sensitivity and NPV and an acceptable specificity and PPV. Not surprisingly, we found that patients who were followed in nephrology clinics had a higher sensitivity, specificity, positive and negative predictive values compared to those who were not seen in nephrology clinics. As one would expect, these results support that nephrologists are more likely to accurately label patients as ADPKD when they actually have the disease compared to other specialists. However, this could be partially explained by some providers using the ICD 9/10 codes when they are referring patients to the nephrology clinic to rule out ADPKD.

Regardless of nephrology follow up, the ADPKD computable phenotype based on ICD-9/10 codes has a relatively high sensitivity 97.1% to 99.2%, compared to other computable phenotypes of other medical conditions, such as acetaminophen toxicity 94%,[19] myocardial infarction 93.5%, cerebrovascular disease 83%, and dementia 92.7%.[8] Additionally, the specificity of the ADPKD computable phenotype of 82.0% to 84.1% was comparable to other medical conditions, such as acetaminophen toxicity 83%,[19] but lower than the specificity for myocardial infarction 94.6%, cerebrovascular disease 95.4%, and dementia 98.9%.[19,20] These results confirm that the ADPKD computable phenotype using ICD-9/10 codes is a practical tool to identify potential patients with ADPKD and to rule out ADPKD, but less accurate for confirming the ADPKD diagnosis.

This study is the first to comprehensively assess all aspects of test accuracy of an ADPKD computable phenotype. Blanchette et al, reviewed records of 132 patients having ICD-9 code for ADPKD (753.12) with a reported PPV of 95%.[21] Kalatharan et al. reviewed records of 201 patients using ICD-10 code for ADPKD (Q61.2 or Q61.3) with a reported PPV of 85%.[22] These two studies didn't assess the sensitivity and specificity of the computable phenotype in identifying patients with ADPKD. Our findings of a PPV of 73.4 - 83.4 are more comparable to the findings by Kalatharan et al which also utilized data from a large healthcare system.

Our study has multiple strengths. First, this is the largest study ever done to assess the test accuracy of a computable phenotype. Additionally, we have assessed all aspects of test accuracy results (Table 4), and we have reported the results based on the STARD guidelines for reporting diagnostic accuracy studies. Finally, we have included both ICD-9 and ICD-10 codes, and have compared those who follow in the nephrology clinic and those who do not.

We note a few limitations in our study. First, our results likely underestimate the specificity estimate because we enriched our sample for patients with renal cysts. This reflects the worst-case scenario for specificity, but was important to consider because that is the group that most likely gets confused with ADPKD. Another limitation was the 34 patients that had inconclusive results and were excluded from the analysis due to lack of information in the EHR to allow for categorization as having or not having ADPKD. Additionally, we did not consider the timing of insertion of the ICD 9/10 codes into the system; however, we think this is reflective of what providers will find in the EMR and is consistent with our attempt to keep the computable phenotype simple and make it practical. Furthermore, there might be an overestimation of the sensitivity since our study did not evaluate the whole healthcare system and accounted for those with ADPKD who are missing a diagnosis code. However, when we evaluated a random sample

of patients without ADPKD ICD 9/10 codes, and despite enriching this group with patients who have ICD codes for renal cysts, we only identified eight additional ADPKD patients out of 522 records evaluated. Finally, this computable phenotype was only tested in one healthcare system, and KUMC is considered to be a PKD referral center, so there may be bias in that there is significantly more institutional knowledge about PKD than in non-referral centers. Future efforts should focus on testing the computable phenotype in other healthcare systems.

Our results show that a computable phenotype of ADPKD ICD-9/10 codes is a good tool to screen for patients and assess the feasibility of participating in ADPKD trials. This is especially important in an era with an approved treatment and many more in the pipeline that will need to be tested in trials. The computable phenotype is not accurate enough to confirm the diagnosis of ADPKD. However, it is accurate to rule-out an ADPKD diagnosis. This will support a strategy focusing on patients who have ICD-9/10 for ADPKD when screening for trials and not waisting time and resources looking up patients without these codes. The final confirmation of ADPKD diagnosis relies on the patients' radiology reports such as ultrasound, CT scan, and MRI, and on the accurate characterization of patients' family history. One of the main challenges to automating a final ADPKD diagnosis is that information in radiology reports are summarized as open texts, which does not directly translate into ICD-9 or ICD-10 codes or searchable elements in EHR. This could be a reason for the limited specificity of ICD-9 and ICD-10 codes in diagnosing ADPKD. Another reason may be that providers are either not knowledgeable of the disease and how to differentiate it from simple renal cysts, or that they are not familiar with the ADPKD ICD codes. Developing and evaluating algorithms that enhance the accurate detection using ICD-9/10 is an important next step to further improving the specificity of this computable

phenotype. Natural language processing algorithm of radiology reports and notes documenting family history could be considered and studied.

Conclusion

The ADPKD computable phenotype based on ICD 9/10 is an excellent screening tool to identify patients with ADPKD. Assessing the accuracy of the ADPKD computable phenotype is an important step in defining best strategies to identify and recruit patients with ADPKD for trials at a time when many innovative interventions are being developed and will need to be tested in trials. Additional searches including specific medications and procedures could enhance the accuracy of the computable phenotype.

Disclosures: K. McGreal reports the following: Medical Center Research Funding: Reata- PI on FAlcon study, Sanofi - Staged PKD trial -sub-I. D. Jalal reports the following: Research Funding: AstraZenica, Corvidia; Honoraria: K-INBRE, Reata; Scientific Advisor or Membership: Reata. A. Yu reports the following: Consultancy Agreements: Regulus Therapeutics, Calico, Otsuka, Navitor; Ownership Interest: Amgen Corp., Gilead Sciences, Prothena; Honoraria: Elsevier, Wolters Kluwer; Scientific Advisor or Membership: Otsuka Advisory Board; Other Interests/Relationships: The University of Kansas Medical Center and the Jared Grantham Kidney Institute receives royalties from Otsuka for tolvaptan. R. Mustafa reports the following: Scientific Advisor or Membership: The GRADE guidance group, The American College of Physicians Clinical Guideline Committee, The Canadian Society of Nephrology Clinical Practice Guidelines Committee; Other Interests/Relationships: Advisory board for the renal round table, NKF Midewest. The remaining authors have nothing to disclose.

Funding: This study was supported by the Kansas PKD Research and Translation Core Center (P30 DK106912), as well as CTSA Award # UL1TR002366 and HERON.

Author contribution: Mohamad A. Kalot: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing - original draft; Writing - review and editing
Abdallah Alayli: Data curation; Formal analysis; Investigation; Writing - original draft; Writing - review and editing
Mohammed Alkhatib: Data curation; Formal analysis; Investigation; Writing - original draft
Nedaa Husainat: Data curation; Formal analysis; Writing - original draft
Kerri McGreal: Data curation; Investigation; Methodology; Writing - original draft; Writing - review and editing
Diana Jalal: Conceptualization; Formal analysis; Investigation; Methodology; Writing - review and editing
Alan Yu: Formal analysis; Investigation; Methodology; Writing - review and editing
Reem Mustafa: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Supervision; Validation; Writing - original draft; Writing - review and editing

Supplemental material:

- Suppl Fig 1. STARD flow diagram for a study of 536 patients from the nephrology clinic for ADPKD diagnosis status.
- Suppl Fig 2. STARD flow diagram for a study of 535 patients who don't have nephrology visits for ADPKD diagnosis status.

References

1. Harris PC, Torres VE. Polycystic kidney disease. *Annu Rev Med.* 2009;60:321-337.
2. Vlasschaert ME, Bejaimal SA, Hackam DG, Quinn R, Cuerden MS, Oliver MJ, et al. Validity of administrative database coding for kidney disease: a systematic review. *Am J Kidney Dis.* 2011;57(1):29-43.
3. Torres VE, Harris PC. Autosomal dominant polycystic kidney disease: the last 3 years. *Kidney Int.* 2009;76(2):149-168.
4. Gradzik M, Niemczyk M, Gołębiowski M, Pączek L. Diagnostic Imaging of Autosomal Dominant Polycystic Kidney Disease. *Polish journal of radiology.* 2016;81:441-453.
5. Gradzik M, Niemczyk M, Golebiowski M, Paczek L. Diagnostic Imaging of Autosomal Dominant Polycystic Kidney Disease. *Pol J Radiol.* 2016;81:441-453.
6. Richesson RL SM. Electronic health records-based phenotypin. *NIH Health Care Systems Research Collaboratory Rethinking Clinical Trials: a living textbook of pragmatic clinical trials.*
7. Tasker RC. Why Everyone Should Care About "Computable Phenotypes". *Pediatr Crit Care Med.* 2017;18(5):489-490.
8. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res.* 2008;43(4):1424-1441.
9. Topaz M, Shafran-Topaz L, Bowles KH. ICD-9 to ICD-10: evolution, revolution, and current debates in the United States. *Perspect Health Inf Manag.* 2013;10:1d.
10. Prevention CfDCa. International Classification of Diseases,Tenth Revision (ICD-10). https://www.cdc.gov/nchs/icd/icd10.htm. Published April 15 2016. Accessed 04/22/2019.
11. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. *Obstet Gynecol.* 2010;116(5):1018-1019.
12. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc.* 2010;17(2):124-130.
13. Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. *AMIA Annu Symp Proc.* 2011;2011:1454-1463.
14. Soroka S, Alam A, Bevilacqua M, Girard LP, Komenda P, Loertscher R, et al. Updated Canadian Expert Consensus on Assessing Risk of Disease Progression and Pharmacological Management of Autosomal Dominant Polycystic Kidney Disease. *Can J Kidney Health Dis.* 2018;5:2054358118801589.
15. Wallis S. Binomial Confidence Intervals and Contingency Tests: Mathematical Fundamentals and the Evaluation of Alternative Methods. *Journal of Quantitative Linguistics.* 2013;20(3):178-208.
16. Altman DG, Machin, David, Bryant, Trevor N., Gardner, Martin J. Statistics with Confidence: Confidence Intervals and Statistical Guidelines, 2nd Edition. In:2000.
17. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6(11):e012799.
18. Simundic AM. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC.* 2009;19(4):203-211.

19. Myers RP, Leung Y, Shaheen AAM, Li B. Validation of ICD-9-CM/ICD-10 coding algorithms for the identification of patients with acetaminophen overdose and hepatotoxicity using administrative data. *BMC Health Services Research.* 2007;7(1):159.
20. So L, Evans D, Quan H. ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Serv Res.* 2006;6:161.
21. Blanchette CM, Liang C, Lubeck DP, Newsome B, Rossetti S, Gu X, et al. Progression of autosomal dominant kidney disease: measurement of the stage transitions of chronic kidney disease. *Drugs Context.* 2015;4:212275.
22. Kalatharan V, Pei Y, Clemens KK, McTavish RK, Dixon SN, Rochon M, et al. Positive Predictive Values of International Classification of Diseases, 10th Revision Coding Algorithms to Identify Patients With Autosomal Dominant Polycystic Kidney Disease. *Can J Kidney Health Dis.* 2016;3:2054358116679130.

Table 1: Diagnostic criteria used for ADPKD diagnosis

	Age (years)	Criteria
Positive family history (The unified imaging diagnosis criteria published in the Canadian Journal of Kidney Health and Disease.)	15-40	At least three unilateral or bilateral kidney cysts
	40-59	At least two cysts in each kidney
	>60	At least four cysts in each kidney
Negative family history	Any age	- Innumerable cysts in both kidneys (as at least 10 cysts in each of the 2 kidneys) - Each kidney greater than 13 cm in length

Table 2 – Descriptive analysis for patients who follow in the nephrology clinic

Status	Baseline characteristics		**ADPKD**	**No ADPKD**	**Totals**
Positive computable phenotype	**Age (mean** ± SD)		51±15	58±17	52±15
	Gender	Female	130	19	149
		Male	106	28	134
	Race	Caucasian	208	34	242
		Black or AA	16	7	23
		Native American	0	1	1
		Asian	3	2	5
		Other	9	3	12
Negative computable phenotype	**Age (mean** ± SD)		51±28	68±14	68±14
	Gender	Female	1	101	102
		Male	2	148	150
	Race	Caucasian	1	151	152
		Black or AA	2	75	77
		Native American	0	1	1
		Asian	0	3	3
		Other	0	19	19

Table 3 – Descriptive analysis for patients who do not follow in the nephrology clinic					
Status	Baseline characteristics		**ADPKD**	**No ADPKD**	**Totals**
Positive computable phenotype	**Age (mean** ± SD)		59±13	52±24	57±17
	Gender	Female	57	16	73
		Male	89	26	115
	Race	Caucasian	142	39	181
		Black or AA	12	10	22
		Native American	0	0	0
		Asian	1	2	3
		Other	1	7	17
Negative computable phenotype	**Age (mean** ± SD)		70±7	66±15	66±15
	Gender	Female	2	117	119
		Male	0	6	6
	Race	Caucasian	4	207	211
		Black or AA	1	32	33
		Native American	0	0	0
		Asian	0	7	7
		Other	0	19	19

Table 4 – Contingency table displaying frequency distribution of patients who follow in the nephrology clinic			
	ADPKD	**No ADPKD**	**Totals**
Positive computable phenotype	236	47	283
Negative computable phenotype	3	249	252
Totals	239	296	535

Table 5 – Contingency table displaying frequency distribution of patients who do not follow in the nephrology clinic			
	ADPKD	**No ADPKD**	**Totals**
Positive computable phenotype	165	58	223
Negative computable phenotype	5	265	270
Totals	170	323	493

Table 6 - ADPKD diagnostic test accuracy results		
Test Accuracy results	**Patients who follow in the nephrology clinics – effect estimate (95% CI)**	**Patients who do not follow in the nephrology clinics – effect estimate (95% CI)**
Sensitivity	99.2% (97-99)	97.1% (93- 99)
Specificity	84.1% (79-88)	82.1% (77-86)
Positive Predictive Value	83.4% (79-87)	73.9% (69-78)
Negative Predictive Value	99.2% (97-100)	98.2% (96-99)
Likelihood Ratio Positive	6.22 (4.79-8.09)	5.44 (4.27-6.83)
Likelihood Ratio Negative	0.01 (0.00-0.04)	0.04 (0.02-0.09)