



## ORIGINAL ARTICLE

# Diagnostic accuracy of administrative codes for autosomal dominant polycystic kidney disease in clinic patients with cystic kidney disease

Vinusha Kalatharan<sup>1</sup>, Eric McArthur<sup>2</sup>, Danielle M. Nash<sup>2</sup>, Blayne Welk<sup>1,2</sup>, Sisira Sarma<sup>1,2</sup>, Amit X. Garg<sup>1,2,3,\*</sup> and York Pei<sup>4,\*</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada, <sup>2</sup>ICES, London, Ontario, Canada, <sup>3</sup>Division of Nephrology, Department of Medicine, Western University, London, Ontario, Canada and <sup>4</sup>University Health Network, University of Toronto, Toronto, Ontario, Canada

Correspondence to: York Pei; E-mail: [york.pei@uhn.ca](mailto:york.pei@uhn.ca)

\*Co-senior authors.

## ABSTRACT

**Background.** The ability to identify patients with autosomal dominant polycystic kidney disease (ADPKD) and distinguish them from patients with similar conditions in healthcare administrative databases is uncertain. We aimed to measure the sensitivity and specificity of different ADPKD administrative coding algorithms in a clinic population with non-ADPKD and ADPKD kidney cystic disease.

**Methods.** We used a dataset of all patients who attended a hereditary kidney disease clinic in Toronto, Ontario, Canada between 1 January 2010 and 23 December 2014. This dataset included patients who met our reference standard definition of ADPKD or other cystic kidney disease. We linked this dataset to healthcare databases in Ontario. We developed eight algorithms to identify ADPKD using the International Classification of Diseases, 10th Revision (ICD-10) codes and provincial diagnostic billing codes. A patient was considered algorithm positive if any one of the codes in the algorithm appeared at least once between 1 April 2002 and 31 March 2015.

**Results.** The ICD-10 coding algorithm had a sensitivity of 33.7% [95% confidence interval (CI) 30.0–37.7] and a specificity of 86.2% (95% CI 75.7–92.5) for the identification of ADPKD. The provincial diagnostic billing code had a sensitivity of 91.1% (95% CI 88.5–93.1) and a specificity of 10.8% (95% CI 5.3–20.6).

**Conclusions.** ICD-10 coding may be useful to identify patients with a high chance of having ADPKD but fail to identify many patients with ADPKD. Provincial diagnosis billing codes identified most patients with ADPKD and also with other types of cystic kidney disease.

**Keywords:** administrative data, diagnostic accuracy, polycystic kidney disease, sensitivity, specificity

Received: 2.1.2019; Editorial decision: 28.11.2019

© The Author(s) 2020. Published by Oxford University Press on behalf of ERA-EDTA.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is characterized by focal cyst development leading to enlargement of both kidneys [1]. It is a relatively uncommon condition with a prevalence of 1 in 1000 to 1 in 400 (0.1–0.25%) [2]. For this reason, assembling a large cohort of patients with ADPKD for research poses a challenge. A possible way to overcome this challenge is to use the existing healthcare administrative databases and codes to assemble a group of patients with ADPKD.

Patients with ADPKD can be captured by International Classification of Diseases, 10th Revision (ICD-10) codes and Ontario Health Insurance Plan (OHIP) diagnosis codes. The International Classification of Diseases, 9th Revision (ICD-9) is an alphanumeric coding system developed by the World Health Organization in 1979 to allow comparability of mortality and morbidity data across countries. In 2002, ICD-9 codes were replaced by the more comprehensive set of ICD-10 codes in Ontario, Canada. In Canada, trained personnel reviewed the medical charts of each patient with a hospital encounter on an ongoing basis and assigned ICD-9 or ICD-10 codes to each hospital encounter according to the rules and guidelines provided by the Canadian Institute for Health Information (CIHI). These codes include descriptors for ADPKD. OHIP diagnosis codes are submitted by Ontario physicians to be reimbursed for the services they provide. The OHIP diagnosis code for other cystic kidney diseases (593) and congenital anomalies of the urinary system (753) may also capture patients with ADPKD, but also patients without ADPKD.

A primary goal of the current study was to determine if ADPKD administrative coding algorithms identify patients with ADPKD and distinguish them from patients with similar conditions. We must first ensure that ADPKD administrative coding algorithms can reliably identify patients with ADPKD and distinguish them from patients with similar conditions before using them. To date, two studies have assessed the performance of ICD-9 and ICD-10 codes related to ADPKD [3, 4]. Both studies showed that a high percentage of patients identified with a code for ADPKD truly had ADPKD (i.e. a high-positive predictive value) [3, 4]. However, these studies did not assess the sensitivity and specificity of these codes or the performance of the OHIP diagnosis codes. Calculating the sensitivity and specificity using a cohort of patients with ADPKD and patients with similar conditions would provide further insights into code performance. We conducted this study to understand the sensitivity and specificity of different coding algorithms containing ICD-10 and OHIP diagnosis codes for ADPKD in a clinic population with different types of cystic kidney disease to gain insight into what percentage of patients with ADPKD are captured by the codes and whether administrative codes in Ontario differentiate patients with ADPKD from patients with similar conditions. We also described and compared the characteristics of patients identified with the different coding algorithms at the time of code assignment.

## MATERIALS AND METHODS

### Study design

We conducted a validation study to assess the sensitivity and specificity of ADPKD coding algorithms using ICD-10 and OHIP diagnosis codes. We used prospectively collected data from a specialty Hereditary Kidney Disease Clinic linked to healthcare databases housed at ICES, a not-for-profit research institute.

We conducted and reported this study in accordance with the Standards for the Reporting of Diagnostic Accuracy studies checklist [5].

The institutional review board at the University Health Network, Toronto, Ontario, Canada approved this study. The research ethics board at the University Health Network waived the need to obtain consent to use data from all other patients who did not undergo genetic testing. We obtained patient consent for all individuals who underwent genetic testing to use their information for research in general. The use of data at ICES for this project was authorized under Section 45 of Ontario's Personal Health Information Protection Act, which does not require review by a research ethics board.

### Reference standard

Our study population included 674 adult patients ( $\geq 18$  years of age) with ADPKD and other types of cystic kidney disease who were seen at the Hereditary Kidney Disease Clinic at the Toronto General Hospital, Toronto, Ontario, Canada between 1 January 2010 and 23 December 2014. The case mix seen at the Hereditary Kidney Disease Clinic consists of patients referred to the specialty clinic by other community nephrologists. Patients underwent abdominal imaging prior to visiting the clinic. All patients included in the dataset underwent kidney function testing at their first visit. A subset of the patients (520 patients) also underwent a comprehensive mutation screen of the PKD1 or PKD2 genes [6, 7]. Details on the methods for comprehensively screening for mutations in the PKD1 and/or PKD2 gene are reported elsewhere [8]. A senior nephrologist with content expertise in ADPKD (Y.P.) adjudicated the ADPKD status of each patient based on if a pathogenic PKD1 or PKD2 mutation was detected after a comprehensive mutation screen and/or whether or not each patient met the current internationally accepted ultrasound diagnostic criteria for ADPKD. The current ultrasound diagnostic criteria for ADPKD are family history of ADPKD and an age-specific minimum number of cysts in the kidney(s) on a conventional kidney ultrasound: (i) three cysts in total when the number of cysts seen in both kidneys are combined for patients  $\leq 39$  years of age; (ii) at least two cysts in each kidney for patients between 40 and 59 years of age; (iii) at least three cysts in each kidney for patients  $\geq 60$  years of age [9]. This gave rise to the ADPKD status variable in the Hereditary Kidney Disease Clinic database, which served as the reference standard for this study. Those with an ADPKD status based on a comprehensive screen and/or ultrasound imaging were categorized as having ADPKD and those with autosomal recessive polycystic kidney disease or other cystic disease were categorized as not having ADPKD. We also collected and recorded demographic information, such as name, date of birth, postal code and gender, as well as the Ontario health card number and medical records number of each patient to allow for data linkage.

### Data sources, patient selection and data collection

We linked the patients in the Hereditary Kidney Disease Clinic database to five administrative databases held at ICES: (i) CIHI Discharge Abstract Database (DAD), which contains information on hospital discharges of patients admitted to hospitals in Ontario; (ii) National Ambulatory Care Reporting System (NACRS), which contains information on patients who visited the emergency department; (iii) the OHIP database, which contains physician billing and diagnosis information; (iv) the Registered Persons Database, which contains demographic and vital status

information for all Ontarians; and (v) Dynacare, which contains laboratory test values for a subset of Ontarians who visited a Dynacare laboratory. These datasets were linked using unique, encoded identifiers and analysed at ICES. We used the Hereditary Kidney Disease Clinic database to assemble our study population of patients with ADPKD and other cystic kidney diseases. As a data cleaning step, we excluded patients who were non-Ontario residents or with missing or invalid identifiers.

We looked in the period from 1 April 2002 through 31 March 2015 to determine if each patient had at least one ICD-10 code for ADPKD during a hospital encounter using CIHI-DAD and NACRS, or an OHIP diagnosis code for ADPKD billed by a physician. If a patient had more than one administrative code for ADPKD, then we selected the first code and used it as the date the patient was first recognized to have ADPKD using administrative data. We classified a patient as algorithm positive if any of the codes appeared at least once. The analyst was not blinded to ADPKD status. However, trained hospital medical staff conducted a standardized review of each hospital medical chart and assigned administrative database codes; these personnel were unaware of this study.

### Database algorithms

We identified three ICD-10 and two OHIP diagnosis codes that could be related to ADPKD (Table 1) and we evaluated the diagnostic performance of these codes singularly, as well as three combinations of these codes, to identify patients with ADPKD.

### Analysis

We calculated the sensitivity and specificity for each of the coding algorithms and calculated their respective 95% confidence intervals (CIs) using the Wilson score method [10]. An algorithm with high sensitivity classifies a large proportion of patients who truly have ADPKD as ADPKD positive. An algorithm with high specificity classifies a large proportion of patients without ADPKD as ADPKD negative. We did not calculate positive and negative predictive values because these measures are influenced by prevalence and the prevalence of ADPKD in our study population is higher than that in the general population.

For the patients we identified as having ADPKD with the algorithms we also described the mean [standard deviation (SD)] age, sex, percentage of patients who had end-stage renal disease, percentage of patients who were hypertensive and the distribution of Johns Hopkins' aggregated diagnosis group (ADG) score at the time of code assignment. The Johns Hopkins' ADG score is a comorbidity score based on healthcare utilization that ranges from 0 to 32, where a higher score indicates greater comorbidity (the Johns Hopkins ACG System version 10.0) [11, 12]. Among the patients with laboratory values, we also described the estimated glomerular filtration rate (eGFR) and chronic kidney disease stage. We conducted all analyses using SAS 9.3 (SAS Institute, Cary, NC, USA).

## RESULTS

### Data linkage and study population

The Hereditary Kidney Disease Clinic database initially contained 674 patients seen at the clinic between 1 January 2010 and 23 December 2014. We determined ADPKD status solely by imaging in 154 patients and both renal imaging and genetic screening in 520 patients. After linkage and exclusions, the final study cohort consisted of 646 patients with cystic kidney

Table 1. Administrative codes related to ADPKD

Database	Code	Description
CIHI-DAD and NACRS	Q611	Polycystic kidney disease, autosomal recessive
CIHI-DAD and NACRS	Q612	Polycystic kidney disease, autosomal dominant
CIHI-DAD and NACRS	Q613	Polycystic kidney disease, unspecified
OHIP	753	Congenital anomalies, urinary system
OHIP	593	Other disorders of the kidneys or ureter

disease (581 patients with ADPKD and 65 patients with other cystic kidney diseases; Figure 1). The mean age of patients with ADPKD was 35 (SD 16) years and 57% were female. The mean age of patients with other cystic diseases was 37 (SD 18) years and 46.2% were female.

### Sensitivity and specificity of the alternative different coding algorithms

The sensitivity and specificity of each of the five individual administrative codes and the three different coding algorithms are presented in Table 2. In general, the sensitivity was high for the OHIP diagnosis codes and relatively low for the ICD-10 codes. In contrast, the specificity was low for OHIP diagnosis codes and high for ICD-10 codes.

### Patient characteristics by coding algorithm

The characteristics of patients identified with the different coding algorithms at the time of code assignment are presented in Table 3. Patients identified with any one of the OHIP diagnosis codes tended to be younger on average compared with patients identified with the ICD-10 codes related to ADPKD. There were approximately equal percentages of females in all three groups. Patients identified with ICD-10 codes on average had a greater number of comorbidities than patients identified with OHIP diagnosis codes. Among those with available laboratory values, kidney function (defined by the most recent eGFR in the 1 year prior to code assignment) was also lower in the group identified with ICD-10 codes than the group with OHIP diagnosis codes. A greater percentage of patients identified with ICD-10 codes also had end-stage renal disease and were more likely to be diagnosed with hypertension compared with patients identified with the OHIP diagnosis codes.

## DISCUSSION

We estimated the sensitivity and specificity of administrative coding algorithms related to ADPKD by conducting a validation study within a hereditary kidney disease clinic where ADPKD was defined using accepted standards. Our study showed that ICD-10 codes for ADPKD have a high specificity but low sensitivity and OHIP diagnosis codes have a high sensitivity but low specificity. In other words, ICD-10 codes differentiated patients with ADPKD from patients with other cystic kidney diseases but failed to identify many patients with ADPKD from a group of patients with similar conditions. The OHIP diagnosis codes identified most patients with ADPKD and also identified patients with other cystic kidney diseases. The ICD-10 codes for

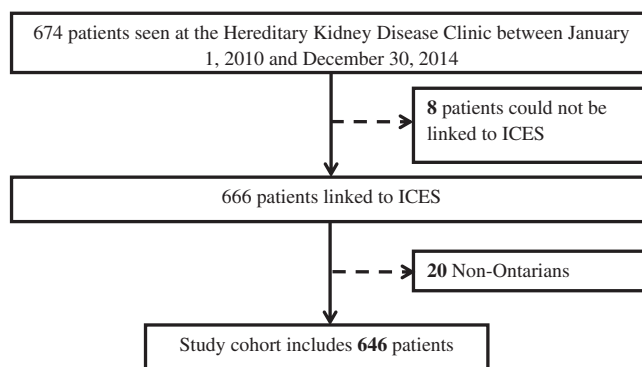


FIGURE 1: Study population.

**Table 2. Sensitivity and specificity of each administrative database coding algorithm**

Coding algorithm	Sensitivity (95% CI), %	Specificity (95% CI), %
ICD-10 Q611	≤0.9 (0.4–2.0) <sup>a</sup>	≥92.3 (83.2–96.7) <sup>a</sup>
ICD-10 Q612	11.5 (9.2–14.4)	≥92.3 (83.2–96.7) <sup>a</sup>
ICD-10 Q613	30.6 (27.0–34.5)	87.7 (77.6–93.6)
OHIP Dx 753	67.5 (63.6–71.2)	35.4 (24.9–47.5)
OHIP Dx 593	71.3 (67.5–74.8)	20.0 (12.08–31.3)
ICD-10 Q611, Q612 or Q613	33.7 (30.1–37.7)	86.2 (75.7–92.5)
OHIP Dx 753 or 593	91.0 (88.5–93.1)	10.8 (5.3–20.6)
ICD-10 Q611, Q612 or Q613 or OHIP Dx 753 or 593	92.3 (89.8–94.2)	10.8 (5.3–20.6)

<sup>a</sup>Sensitivity and specificities were calculated as if small cells were five to ensure individuals cannot be re-identified, as per ICES privacy policies.

ADPKD also tended to capture patients with more advanced ADPKD, as defined by lower eGFR and higher serum creatinine, on average compared with OHIP diagnosis codes. Overall, these findings highlight important considerations when using administrative data to identify patients with ADPKD.

One reason for the limited sensitivity of ICD-10 codes to identify patients with ADPKD is that they can only identify patients with hospital encounters in Ontario [13]. Physicians submit billing claims with accompanying diagnosis codes for the medical services they provide. The assigned OHIP diagnosis codes for these services may not be accurate. Based on the opinion of a nephrologist with expertise in ADPKD, the two most common diagnosis codes in Ontario billed for patients with ADPKD are 753 (congenital anomalies of the urinary system) and 593 (other disorders of the kidney and ureter). Since the database used to validate the codes consists of patients from a specialty clinic for hereditary kidney disease, almost all patients with ADPKD in the dataset would have at least one OHIP diagnosis code 753 or 593. This may explain our finding for high sensitivity of physician claims diagnosis codes. The code descriptions were not specific to ADPKD, so the codes may have also identified patients with other congenital anomalies of the urinary system (e.g. medullary cystic kidney disease) and other cystic kidney disease (e.g. simple cyst), resulting in low specificity for the detection of ADPKD.

Literature on the validity of administrative coding algorithms related to ADPKD is scarce. To date, only one study assessed the positive predictive value of ICD-9 code 75312 (polycystic kidney

disease, unspecified) and another study described the positive predictive value of ICD-10 coding algorithms related to ADPKD [3, 4]. Both studies showed that hospital encounter codes related to ADPKD have a high positive predictive value when applied in the general population, meaning that a high percentage of patients with ADPKD codes truly have ADPKD [3, 4]. To the best of our knowledge, our study is the first to estimate the sensitivity and specificity of administrative coding algorithms related to ADPKD in a specialized cystic kidney disease clinic population and provides insight into whether these algorithms differentiate patients with ADPKD from patients with other cystic conditions. Our study provides further assurance that ICD-10 codes can be used to assemble a robust cohort of patients with ADPKD since ICD-10 codes differentiate patients with ADPKD from patients with similar conditions. In contrast, OHIP diagnosis codes do not differentiate patients with ADPKD from patients with similar conditions well but capture the majority of patients with ADPKD. As suggested by the description of the OHIP codes, our study confirms that the codes are not specific to ADPKD. Therefore we cannot rely on OHIP diagnosis codes to assemble a robust cohort of patients with ADPKD. However, we can use OHIP diagnosis codes as exclusion codes or to flag patients who warrant a detailed review of medical records to determine whether they truly have ADPKD.

The fact that our study population is solely from Ontario limits the generalizability of our findings to other regions since healthcare coding practices vary across the world. In Ontario, OHIP diagnosis codes and fee-for-service codes are submitted by physicians for remuneration for the services they provide. ICD-10 codes are traditionally used for administrative purposes, such as assessing healthcare use and needs in hospital settings. In other regions, ICD-10 codes are used in outpatient settings as well. As a result, our findings should be generalized with caution to regions that use ICD-10 codes in outpatient settings.

There are strengths and limitations to using different administrative codes to identify patients with ADPKD. Using ICD-10 codes to assemble a study population will maximize the internal validity of future studies while limiting the generalizability of study findings to patients with a hospital encounter with ADPKD. In contrast, using OHIP diagnosis codes for cohort accrual will compromise the internal validity of the study. This validation study was done in a selected sample of specialized clinic patients. This study can be used in combination with others to understand the true utility of these codes [3, 4].

## ACKNOWLEDGEMENTS

ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC),



Table 3. Patient characteristics at the time of code assignment for the different coding algorithms

Patient characteristics	OHIP Dx 593 or 753 (n = 587)	ICD-10 Q611, Q612 or Q613 (n = 206)	ICD-10 Q611, Q612 or Q613 or OHIP Dx 753 or 593 (n = 594)
Age (years), mean ± SD	36 ± 16	41 ± 17	36 ± 16
Female, n (%)	330 (56.2)	121 (58.7)	333 (56.1)
ADG score, mean ± SD	10.4 ± 4.0	12.5 ± 3.9	10.2 ± 4.0
0–2, n (%)	19 (3.3)	10 (4.9) <sup>b</sup>	22 (3.7)
3–5, n (%)	48 (8.2)		50 (8.4)
>6, n (%)	520 (88.6)	196 (95.1)	522 (87.9)
End-stage renal disease, n (%) <sup>a</sup>	37 (6.3)	48 (23.3)	36 (6.1)
Hypertension, n (%)	234 (39.9)	118 (57.3)	232 (39.1)
Kidney function, n (%) with	133 (22.7)	32 (15.5)	133 (22.4)
Most recent serum creatinine (µmol/L), mean ± SD	104.6 ± 92.72	249.5 ± 241.4	105.8 ± 93.2
Most recent eGFR (mL/min/1.73 m <sup>2</sup> ), mean ± SD	84.9 ± 31.6	55.7 ± 48.2	84.0 ± 31.9
Chronic kidney disease stage, n (%)			
≥60 mL/min/1.73 m <sup>2</sup>	105 (78.9)	13 (40.6)	103 (77.4)
<60 mL/min/1.73 m <sup>2</sup>	28 (21.0)	19 (59.4)	30 (22.6)

<sup>a</sup>End-stage renal disease is defined as patients on chronic dialysis in the past 1 year or patients who received a kidney transplantation in the past 5 years.

<sup>b</sup>Cells are suppressed to ensure that individuals cannot be re-identified, as per ICES privacy policies.

supported this study. The opinions, results and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. Parts of this material are based on data and/or information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed in the material are those of the authors and not necessarily those of CIHI.

## FUNDING

The Polycystic Kidney Diseases Foundation (Kansas City, MO, USA) and the ICES Kidney, Dialysis and Transplantation Program provided funding for this study. V.K.'s training was supported by a Canadian Institutes of Health Research Doctoral Scholarship and a Doctoral Scholarship from the Kidney Research Scientist Core Education and National Training Program (KRESCENT) (a national kidney research training partnership of the Kidney Foundation of Canada, Canadian Society of Nephrology and Canadian Institutes of Health Research). A.X.G. was supported by the Dr Adam Linton Chair in Kidney Health Analytics and a Clinician Investigator Award from the Canadian Institutes of Health Research.

## CONFLICT OF INTEREST STATEMENT

Y.P. served as an expert consultant on drug development for Otsuka, Pfizer and Genzyme/Sanofi related to ADPKD. All other authors declare no competing interests. The results presented in this article have not been published previously in whole or part.

## REFERENCES

- Harris PC, Torres VE. Polycystic kidney disease. *Annu Rev Med* 2009; 60: 321–337
- Bergmann C, von Bothmer J, Bröchle NO et al. Mutations in multiple PKD genes may explain early and severe polycystic kidney disease. *J Am Soc Nephrol* 2011; 22: 2047–2056
- Kalatharan V, Pei Y, Clemens KK 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
- Blanchette CM, Liang C, Lubeck DP et al. Progression of autosomal dominant kidney disease: measurement of the stage transitions of chronic kidney disease. *Drugs Context* 2015; 4: 212275
- Bossuyt PM, Reitsma JB, Bruns DE et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1–W12
- European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 1994; 77: 881–894
- Mochizuki T, Wu G, Hayashi T et al. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* 1996; 272: 1339–1342
- Hwang Y-H, Conklin J, Chan W et al. Refining genotype–phenotype correlation in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 2016; 27: 1861–1868
- Pei Y, Obaji J, Dupuis A et al. Unified criteria for ultrasonographic diagnosis of ADPKD. *J Am Soc Nephrol* 2009; 20: 205–212
- Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998; 17: 857–872
- Weiner JP. Johns Hopkins ACG<sup>®</sup> Case-Mix System Version 10.0 Release Notes. Baltimore: Johns Hopkins University Bloomberg School of Public Health, Health Services Research and Development Center, Baltimore, Maryland, 2011
- Austin PC, Shah BR, Newman A et al. Using the Johns Hopkins' aggregated diagnosis groups (ADGs) to predict 1-year mortality in population-based cohorts of patients with diabetes in Ontario, Canada. *Diabetics Med* 2012; 29: 1134–1141
- Knight T, Schaefer C, Krasa H et al. Medical resource utilization and costs associated with autosomal dominant polycystic kidney disease in the USA: a retrospective matched cohort analysis of private insurer data. *Clinicoecon Outcomes Res* 2015; 7: 123–132