Derivation and validation of text search algorithms for renal and adrenal lesion identification in radiology text reports

Luke T. Lavallée, MDCM MSc[1,2,3]; Christopher Knee, ND MSc[2]; James Ross, MD[1]; Johnathan L. Lau, HBSc[4]; Nikhile Mookerji, BSc[4]; Carl van Walraven, MD MSc[2,3,4,5]

[1]Division of Urology, Department of Surgery, The Ottawa Hospital and University of Ottawa, Ottawa, ON, Canada; [2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada; [3]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, ON, Canada; [4]School of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; [5]Institute for Clinical Evaluative Sciences, Ottawa, ON, Canada

***

Abstract

Introduction: Most cohort studies are limited by sampling and accrual bias. The capability to detect specific lesions identified in radiological text reports could eliminate these biases and benefit patient care, clinical research, and trial recruitment. This study derived and internally validated text search algorithms to identify four common urological lesions (solid renal masses, complex renal cysts, adrenal masses, and simple renal cysts) using radiology text reports.

Methods: A simple random sample of 10 000 abdominal ultrasound (US) and computed tomography (CT) reports was drawn from our hospital's data warehouse. Reports were manually reviewed to determine the true status of the four lesions. Using commonly available software, we created logistic regression models having as predictors the status of a priori selected text terms in the report. We used bootstrap sampling with 95th percentile thresholds to select variables for the final models, which were modified into point systems. A second independent, random sample of 2855 reports, stratified by the number of points for each abnormality, was reviewed in a blinded fashion to measure the accuracy of each lesion's point system.

Results: The prevalence of solid renal mass, complex renal cyst, adrenal mass and simple renal cyst, was 2.0%, 1.7%, 3.2%, and 20.0%, respectively. Each model contained between one and

five text terms with c-statistics ranging between 0.66 and 0.90. In the independent validation, the scoring systems accurately predicted the probability that a text report cited the four lesions.
Conclusions: Textual radiology reports can be analyzed using common statistical software to accurately determine the probability that important abnormalities of the kidneys or adrenal glands exist. These methods can be used for case identification or epidemiological studies.

Introduction

It would be very helpful for physicians and researchers to be able to identify all cases of a condition in a patient population. This would allow researchers to conduct studies that are less susceptible to sampling bias (incorrect conclusions due to small or unrepresentative study samples). Recruitment into clinical studies would also be simplified since candidates for recruitment could be easily identified, thereby avoiding a screen of all patients. Finally, it could improve processes of care since it could automatically direct patients with specific conditions to regional experts or specialized clinics.

Currently, it is very difficult to accurately identify disease conditions within typical hospital administrative datasets. Numeric laboratory data are easily processed and analyzed but are capable – by themselves – of identifying only a limited number of specific conditions. For example, identifying patients with elevated prostate specific antigen (PSA) levels after prostatectomy may help identify a cohort of men with recurrent prostate cancer, but cannot differentiate between locally recurrent and metastatic disease. However, textual reports such as radiology and pathology reports are more likely to contain disease-defining information, such as the location of a cancer recurrence, although their analysis is much more difficult. Recently, natural language processing (NLP) has emerged as a promising tool for efficient and automated data extraction; however, this requires advanced software and considerable computer science expertise [1,2].

In this study, we used commonly available statistical software (SAS) to derive and validate text search algorithms capable of identifying four lesions of the kidneys and adrenal glands (solid renal masses, complex renal cysts, adrenal masses, and simple renal cysts) contained within ultrasound (US) and computed tomography (CT) reports. Our goal was to determine whether our simple methodology could accurately predict the probability a lesion exists within a sample of radiology imaging reports, thereby permitting broad use of our methods for cohort identification.

CUAJ – Original Research
Lavallée et al
Renal/adrenal mass text algorithms

Methods

Study setting and administrative datasets used for the study
The study took place at The Ottawa Hospital (TOH) in Ottawa, Canada. This is a multi-institutional, three-campus tertiary care hospital with an average of 60,000 hospitalizations and over 1,000,000 medical encounters annually. The study used data from our hospital's administrative data warehouse, which stores the text reports of all abdominal US and CT studies. The study was approved by our research ethics board.

Lesions studied
We chose to study four lesions of the kidneys and adrenal glands: solid renal masses, complex renal cysts, adrenal masses, and simple renal cysts. Definitions for each lesion are provided in Appendix A. These lesions were chosen because they are relatively common, may have quality of care issues, and were aligned with our clinical interests.

Determining lesion status in radiology reports
We first created a simple random sample of 10,000 reports from abdominal US and CT studies conducted between 1 January 1996 and 15 May 2012. Sampling was stratified by imaging type (US and CT) so that we reviewed 5,000 reports of both. *A priori,* we identified (based on clinical experience) possible text terms/strings (keywords) that radiologists might use to identify each of the four lesions (Appendix B). Each report was then manually reviewed by a team member (urologist, resident, medical student, or research assistant) to determine the true status (present or absent) of each of the four lesions. To facilitate manual review of the reports, the report text was presented is lower case font except for these keywords, which were placed in capital letters. Reports for which a lesion status was rated as 'uncertain' by the initial reviewer were reviewed as a group to determine the final status and achieve consensus.

Creation of lesion scores
For each lesion type, we used multivariable logistic regression to determine the association between lesion status and the presence or absence of each of the text strings identified *a priori*. For each of these models, the dependent variable was lesion status (i.e. present or absent) as determined by manual review. Independent variables included the presence or absence of the text terms/strings in the report (Appendix B). Reports were screened for these strings using text functions in SAS (Cary, North Carolina) with the condition that the text "no" was absent prior to the keyword within that sentence. Certain scoring criteria (e.g. cancer term present, cyst with special characteristics, complex renal cyst indicator, or adrenal metastases term present) may have been met by one of several text strings as defined in Table 1.

We used bootstrap variable selection methods described by Austin in which: 1) 1,000 bootstrap samples of the original dataset were created (having the same sample size as the original, with case replacement); 2) a logistic regression model was created in each bootstrap

sample using forward variable selection (with candidate variables that did *not* enter the model given a parameter estimate of 0); and 3) the median, 5th and 95th percentiles of the parameter estimates of each variable in the 1,000 bootstrap samples were determined [3]. Variables were kept in the final model if they had 5th and 95th percentile values distinct from 0 (and having the same sign) with the median value used as the final parameter estimate. We tried other modelling strategies as well, including bootstrap variable selection with a more liberal variable inclusion criteria (namely 25th and 75th percentiles), as well as chi-square recursive partitioning, but found that resulting models were very similar. Therefore, only the bootstrap variable selection with 95% confidence intervals is presented here.

Lesion scores were created for each lesion of interest (solid renal mass, complex renal cyst, adrenal mass, and simple renal cyst) by modifying the final logistic models into point systems using steps described by Sullivan *et al* [4]. Expected probabilities of lesions for each lesion score value was calculated using another 1,000 bootstrap samples of the original dataset (with each sample the same size as the original cohort) with the proportion of patients in each point truly having the relevant lesion total being determined.

Validation of lesion scores

Finally, we created a new independent random sample of 2,855 abdominal US and CT reports (that were not included in the original sample of 10,000 reports used for model derivation) from a total of 40,600 reports in the hospital data warehouse conducted from 1 January 2016 to 31 December 2017. This sample was stratified by each point system value with a goal of including 200 reports randomly selected within each unique lesion score value (if available). These reports were independently reviewed by two reviewers (resident and research assistant) to determine the true status of all four lesions. Reviewers were blinded to each report's lesion scores. If disagreement occurred between the two reviewers regarding lesion status, the report was reviewed by the group for consensus and final determination.

Results

Between 1 January 1996 and 15 May 2012, there were 217,870 ultrasound and 85,941 computerized tomography reports in the hospital's data warehouse. Of the 10,000 reports randomly selected (5,000 from each imaging modality), the prevalence of solid renal mass, complex renal cyst, adrenal mass, and simple renal cyst was 2.0%, 1.7%, 3.2%, and 20.0%, respectively.

The lesion score systems contained between one and five key terms each (Table 1). Many of these terms were very strongly associated with the lesion; for example, two variables ("AML" and "Renal Neoplasm") in the solid renal mass model had adjusted odds ratios (adj-OR) exceeding 100. Imaging modality (i.e. US vs. CT) was significantly associated with lesion status for adrenal mass (in which CT was associated with an increased likelihood) and simple renal cyst (in which US was associated with an increased likelihood). The discrimination of the models

(measured using the c-statistic) varied notably between the lesion types: solid renal mass: 0.715; complex renal cyst: 0.658; adrenal mass; 0.817; and simple renal cyst: 0.902.

As the number of points increased for each lesion score, there was an increased probability that a lesion truly existed in the derivation sample (Table 2). For both solid renal mass and complex renal cyst, more than 98% of the sample population had lesion point totals of 0, with the associated expected probabilities of lesions being 1.1% and 1.2%, respectively. For adrenal mass and simple renal cyst, low-risk reports were divided into two groups having either a low or a very low expected probability of a lesion. Except for complex renal cyst, patients with the highest point score category had an expected lesion probability that exceeded 90%. 95% confidence intervals for the lesion probability at each point total were, for the most part, relatively narrow.

Application of the point system to the independent validation cohort showed that each point system was useful in determining the probability that a lesion truly existed (Table 2). Each point system was very strongly associated with lesion status (the p-value of the Chi-square statistic of each point system's association with their lesion was <0.0001). As in the derivation cohort, lesion score distribution was highly skewed for solid renal mass and complex renal cyst, with 98% of reports in the validation cohort having a lesion score of 0. For all lesion types, the observed lesion probability increased as scores increased. Overall, observed lesion probability in the validation cohort was similar to expected probability of a lesion from the derivation cohort with two exceptions: 1) the observed probability of a solid renal mass was lower than expected with lesion scores of 2 or 3+; and 2) the observed probability of a complex renal cyst with a lesion score of 0 (3.7%) was slightly higher than predicted (1.2%).

If all reports with a lesion score above 0 were manually reviewed, the completeness of lesion capture varied between models. For example, this approach would expect to identify 99.1% of simple renal cysts since only 0.9% of reports with a simple renal cyst score of 0 truly had the lesion. Lesion capture was also very good for adrenal masses, with only 14.6% of reports with a lesion score of 0 truly having an adrenal mass. Conversely, a high proportion of solid renal masses (59.1%) and complex renal cysts (83.7%) had radiology reports with lesion scores of 0.

Discussion

Complete and accurate identification of specific clinical findings (e.g. adrenal masses) would be very useful for clinicians, researchers, and administrators. Structural diseases are frequently identified by radiographic procedures; however, the textual reports of these studies can be difficult to analyze with commonly available software used by researchers. In this study, we derived and validated models to identify four urological lesions of the kidneys and adrenal glands (solid renal mass, complex renal cyst, adrenal mass, and simple renal cyst) using text functions present in commonly used statistical software.

The models developed in our study could be used to screen radiological reports with one of two objectives: 1) to identify patients having a very high probability of disease (e.g. identifying all reports with solid renal mass score of 2 or more, having an expected probability of a solid renal mass of approximately 80%); or 2) to identify essentially *all* lesions in the population (e.g. retrieving all reports except those with a simple renal cyst of 0 to identify all but approximately 1% of these lesions identified in our hospital). These capabilities provide powerful tools that could be used for patient recruitment, clinical research, and quality improvement. This approach for cohort creation has several advantages over traditional research methods. Most published cohorts of kidney or adrenal masses are derived from one or a few large academic institutions [5–7]. These cohorts are usually created by individual case identification in clinics or after surgery and often require patient consent. As a result, they are prone to selection-, accrual-, and treatment-bias, and may have questionable external validity (i.e. such cohorts may be more likely to include healthy, younger patients living in urban areas near the study centre) [8,9]. Other factors, such as clinical barriers (e.g. specific or limited eligibility criteria), physician selection (e.g. excluding patients with advanced disease or those treated palliatively after diagnosis), patient attitudes (e.g. fear of randomization or required follow up), and socioeconomic status, could also lead to patient exclusion [10–14]. As a result, the findings generated by such cohorts may not be generalizable to the entire population. In addition, the creation of study cohorts in this fashion is both time-consuming and costly. Conversely, our automated method of case identification using radiology reports could, in the case of adrenal mass and simple renal cysts, identify the vast majority or essentially all patients with a particular radiologic finding quickly. While manual review of the medical record would still be required to confirm cases and retrieve granular data for analyses, it is much less likely that cases will be missed and the potential for bias will be reduced. These methods will produce more generalizable study results and will allow for identification of unique patient populations that are frequently excluded for most studies.

The models we created using simple statistical software can be used to create models for other clinically significant lesions. Accurate and complete identification of radiologic and pathologic findings is important. At this time, many new structured reporting systems are being developed in radiology, such as the Prostate Imaging – Reporting and Data System (PI-RADS[TM]) for prostate cancer, now used worldwide [15]. Cancer Care Ontario (CCO), the organization that oversees cancer care in Canada's most populous province, has laid out a vision for province-wide synoptic reporting in radiology and pathology. One of the principle goals of synoptic reporting is to facilitate automated data searches of reports for population health research [16]. It is likely that the accuracy of our models will improve when radiology reports are synoptic as completeness and consistency will be enhanced.

Our study demonstrates several important points. First, it shows that the reports of radiological studies can be analyzed using text functions found in commonly available software

(i.e. without the need for advanced software or experience with natural language processing) to accurately return the probability that a particular lesion is present. This facilitates the retrospective identification of large numbers of these lesions by manually reviewing only those reports that have a high probability of the lesion. Without this process, one would either have to review randomly selected radiological reports or prospectively review studies for the lesion of interest, thereby taking much more time (especially if the lesion is uncommon). In addition, the ability to accurately determine the probability that a radiological report cites a particular lesion makes it possible to conduct analyses that correct for misclassification bias [17–19]. Second, we show here that a highly discriminative lesion model (such as our simple renal cyst model) makes it possible to study an almost entire population of lesions. This capacity avoids potential biases from studying traditional sources of case series (e.g. all patients referred to a particular physician or hospital). Third, we can be confident we identified *all* imaging reports at our hospital for a specific time period, as we used administrative hospital data as the data source. As more hospitals in our health care region move toward a common electronic medical record, even more comprehensive coverage of reports will be possible, resulting in fewer cases missed.

Several issues should be kept in mind when interpreting our results. First, it is unknown how generalizable our models are to other centres. It is possible that different radiologists might use different terminology to identify different conditions. Therefore, our models should be tested before being applied to other settings. Second, our models are not accurate enough to avoid manual review of reports to confirm the presence of these lesions. Even reports classified with the highest risk of having a particular lesion need to be manually reviewed to confirm the lesion since this probability is not high enough (with the potential exception of simple renal cyst) to consider all such reports as truly having the lesion. Third, it is possible that these models could be improved with different analytical methods or using true natural language processing and more elaborate data-mining techniques. Alternatively, if the textual content of radiology reports becomes more homogenous over time our score could become more accurate as individual terms become more predictive; for example, through the use of structured reporting systems such a PI-RADS™ for prostate MRI. Lastly, although our scoring models were not able to identify every positive report (with the exception of simple renal cysts), it is possible that further examination of missed cases, such as the 15% of expected adrenal masses that were assigned a score of zero, would allow us to create another model that could be applied to these cases alone to improve our overall capture.

Conclusions

In summary, we derived and validated models to identify four common kidney and adrenal lesions by analyzing text of abdominal ultrasound and computed tomography studies using commonly available analytical software. These models could be tested and used in other hospitals to identify these lesions and reduce common epidemiologic study biases occurring from

patient recruitment. The processes we used in this study could be used to identify other structural lesions in radiological text reports.

References

1. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural Language Processing in Radiology: A Systematic Review. *Radiology*. 2016;279(2):329-343. doi:10.1148/radiol.16142770

2. Cai T, Giannopoulos AA, Yu S, et al. Natural Language Processing Technologies in Radiology Research and Clinical Applications. *Radiographics*. 2016;36(1):176-191. doi:10.1148/rg.2016150080

3. Austin PC. Using the bootstrap to improve estimation and confidence intervals for regression coefficients selected using backwards variable elimination. *Stat Med*. 2008;27(17):3286-3300. doi:10.1002/sim.3104

4. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631-1660. doi:10.1002/sim.1742

5. Bhindi B, Thompson RH, Lohse CM, et al. The Probability of Aggressive Versus Indolent Histology Based on Renal Tumor Size: Implications for Surveillance and Treatment. *Eur Urol*. 2018;74(4):489-497. doi:10.1016/j.eururo.2018.06.003

6. European Network for the Study of Adrenal Tumors (ens@t). ens@t databases. http://www.ensat.org/page-1317325. Published 2018. Accessed September 12, 2018.

7. Lavallée LT, Tanguay S, Jewett MA, et al. Surgical management of stage T1 renal tumours at Canadian academic centres. *Can Urol Assoc J*. 2015;9(3-4):99-106. doi:10.5489/cuaj.2598

8. Song JW, Chung KC. Observational studies: cohort and case-control studies. *Plast Reconstr Surg*. 2010;126(6):2234-2242. doi:10.1097/PRS.0b013e3181f44abc

9. Sheth SA, Saver JL, Starkman S, et al. Enrollment bias: frequency and impact on patient selection in endovascular stroke trials. *J Neurointerv Surg*. 2016;8(4):353-359. doi:10.1136/neurintsurg-2014-011628

10. Unger JM, Hershman DL, Albain KS, et al. Patient income level and cancer clinical trial participation. *J Clin Oncol*. 2013;31(5):536-542. doi:10.1200/JCO.2012.45.4553

11. Sharrocks K, Spicer J, Camidge DR, Papa S. The impact of socioeconomic status on access to cancer clinical trials. *Br J Cancer*. 2014;111(9):1684-1687. doi:10.1038/bjc.2014.108

12. Unger JM, Cook E, Tai E, Bleyer A. The Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies. *Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meet*. 2016;35:185-198. doi:10.14694/EDBK_156686

13. Ford JG, Howerton MW, Lai GY, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: A systematic review. *Cancer*. 2008;112(2):228-242. doi:10.1002/cncr.23157

14. Lara PN, Higdon R, Lim N, et al. Prospective Evaluation of Cancer Clinical Trial Accrual Patterns: Identifying Potential Barriers to Enrollment. *J Clin Oncol*. 2001;19(6):1728-1733. doi:10.1200/JCO.2001.19.6.1728

15. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging –

Reporting and Data System: 2015, Version 2. *Eur Urol*. 2016;69(1):16-40. doi:10.1016/j.eururo.2015.08.052

16. Cancer Care Ontario. Synoptic Radiology Reporting. https://archive.cancercare.on.ca/ocs/clinicalprogs/imaging/synoptic_radiology_reporting/. Published 2016. Accessed September 12, 2018.

17. van Walraven C. Bootstrap imputation minimized misclassification bias when measuring Colles' fracture prevalence and its associations using health administrative data. *J Clin Epidemiol*. 2018;96:93-100. doi:10.1016/j.jclinepi.2017.12.012

18. van Walraven C. A comparison of methods to correct for misclassification bias from administrative database diagnostic codes. *Int J Epidemiol*. 2018;47(2):605-616. doi:10.1093/ije/dyx253

19. van Walraven C. Improved Correction of Misclassification Bias With Bootstrap Imputation. *Med Care*. 2018;56(7):e39-e45. doi:10.1097/MLR.0000000000000787

Figures and Tables

Table 1. Lesion score systems for solid renal mass, complex renal cyst, adrenal mass, and simple renal cyst			
	Parameter estimate (95% CI)	**Adjusted odds ratio (95% CI)**	**Points**
Solid renal mass			
Intercept	-4.45 (-4.66, -4.27)		-
Cancer term present[†]	3.55 (0.00, 19.2)	34.8 (1, >100)	1
"RENAL MASS" present[*]	4.27 (3.64, 4.88)	71.5 (38.1, 131.6)	1
"AML" present[*]	4.84 (3.88, 5.91)	>100 (48.4, >100)	1
"RENAL NEOPLASM" present[*]	14.5 (0.00, 18.7)	>100 (1, >100)	4
Complex renal cyst			
Intercept	-4.51 (-4.90, -4.3)		-
Any CRC indicator present[‡‡]	4.25 (3.59, 6.39)	70.1 (36.2, >100)	1
Adrenal mass			
Intercept	-5.75 (-6.36, -5.27)		-
CT (instead of US)	2.52 (2.01, 3.12)	12.4 (7.46, 22.6)	1
Adrenal metastases term present[†††]	5.92 (3.67, 36.3)	>100 (39.2, >100)	2
"ADRENAL ADENOMA" present[*]	6.96 (5.85, 34.9)	>100 (>100, >100)	3
Simple renal cyst			
Intercept	-4.10 (-4.29, -3.93)		-
"CYST" with special characteristics[‡]	-0.81 (-1.66, 0.00)	0.44 (0.19, 1)	-2
CT (instead of US)	0.45 (0.32, 0.58)	1.57 (1.38, 1.79)	1
"SIMPLE CYST" present*	0.54 (0.33, 0.78)	1.72 (1.39, 2.18)	1
"CYST" present*	3.99 (3.82, 4.17)	54.0 (45.6, 64.7)	9
"PARAPELVIC CYST" present*	3.86 (3.16, 5.32)	47.5 (23.6, >100)	9

*Without a preceding "NO" in the sentence. †Includes: chromophobe RCC, chromophobe renal cell carcinoma, clear cell RCC, clear cell renal cell carcinoma, collecting duct carcinoma, cystic rcc, cystic Wilms tumor, kidney cancer, kidney metastases, kidney metastasis, kidney neoplasm, malignant kidney mass, malignant renal mass, metanephric adenoma, metastasis to kidney, oncocytoma, papillary RCC, RCC, renal adenocarcinoma, renal adenoma, renal cell carcinoma, renal cell neoplasm, renal medullary carcinoma, renal metastases, renal metastasis, renal neoplasm, and reninoma. ‡Includes: hemorrhagic, hyperdense, or infected. ‡‡ Includes: benign complex cyst, Birt-Hogg Dube, Bosniak, complex kidney cyst, complex renal cyst, cystic RCC, cystic Wilms tumor, hemorrhagic cyst, high density cyst, hyperdense cyst. †††Includes: adrenal metastases, adrenal metastasis, metastasis to adrenal.

Table 2. Prevalence of lesion scores and probability of lesion in derivation and validation cohorts

	Points			
Solid renal mass	**0**	**1**	**2**	**3+**
Derivation cohort				
Prevalence	98.1%	1.7%	0.16%	0.05%
Expected probability of lesion** (%; 95% CI)	1.1% (0.94–1.4)	39.8% (32.8–47.1)	94.4% (80–100)	100% (100–100)
Validation cohort				
n (%)	39 560 (97.4%)	907 (2.2%)	92 (0.2%)	41 (0.1%)
Observed probability of lesion**	47/2452 (1.9%)	129/270 (45.7%)	73/92 (79.4%)	33/41 (80.5%)
Expected # lesions (% of total)	751 (59.1%)	414 (32.6%)	73 (3.2%)	33 (0.3%)
Complex renal cyst				
Derivation cohort				
Prevalence	98.8%	1.2%	-	-
Expected probability of lesion* (%; 95% CI)	1.2% (0.96–1.4)	46.2% (38.3–55.9)	-	-
Validation cohort				
n (%)	39 957 (98.4%)	643 (1.6%)		
Observed probability of lesion	95/2536 (3.7%)	143/319 (44.8%)		
Expected # lesions (% of total)	1478 (83.7%)	288 (16.3%)		
Adrenal mass				
Derivation cohort				
Prevalence	50.0%	48.9%	-	1.2%
Expected probability of lesion* (%; 95% CI)	0.34% (0.18–0.51)	3.9% (3.3–4.4)	-	92.9% (80.7–100)
Validation cohort				
n (%)	36 372 (90.0%)	3975 (9.8%)	6 (0.01%)	247 (0.6%)
Observed probability of lesion	4/1976 (0.2%)	33/626 (5.3%)	3/6 (50.0%)	214/247 (86.6%)
Expected # lesions (% of total)	73 (14.6%)	211	3 (0.6%)	214

		(42.3%)		(42.7%)
Simple renal cyst	**0**	**1–7**	**8–11**	**12+**
Derivation cohort				
Prevalence	35.3%	32.4%	30.3%	2.0%
Expected probability of lesion* (%; 95% CI)	1.1% (0.76–1.4)	3.2% (2.6–3.8)	55.1% (53.4–57.0)	98.0% (95.9–99.5)
Validation cohort				
n (%)	24 222 (60.0%)	2651 (6.5%)	1270 (3.1%)	12 457 (30.7%)
Observed probability of lesion	5/952 (0.5%)	25/586 (4.3%)	620/1029 (60.3%)	274/288 (95.1%)
Expected # lesions (% of total)	121 (0.9%)	114 (0.8%)	800 (6.2%)	11 847 (91.7%)

For each of the four conditions (Column 1), the prevalence of each lesion score and its associated probability of truly containing a lesion is presented (Columns 2–5) for both the derivation cohort (simple random sample of 10 000 abdominal US and CT reports) and the validation cohort (all abdominal US and CT reports, 2016–17; n = 40,600). Lesion scores are presented in Table 1.
*From bootstrap sampling of derivation cohort, median (2.5th, 97.5th percentile). CT: computed tomography; US: ultrasound.

APPENDIX A. Definitions of lesions for which text search algorithms were created.

1. **Solid renal mass:**
    –   Described as solid/mostly solid growth or mass, possibly enhancing, and usually suspicious for malignancy.
    –   Frequently the report will say renal cell carcinoma (RCC)/malignant/neoplasm until proven otherwise.
    –   The report may favor a solid benign tumor (e.g. oncocytoma, angiomyolipoma (AML)) and these should be included as the imaging is not definitive and these lesions may be clinically significant.
    –   These tumors may be associated with metastasis or tumor thrombus.
    –   EXCLUDES: Tumors in the ureter/renal pelvis

2. **Complex renal cyst:**
    –   Complex cysts are usually described as having septations/calcifications/solid component.
    –   They may or may not be given Bosniak classification. If a Bosniak classification is provided, Bosniak 2F, 3, and 4 cysts should be included here.
    –   Usually has some suspicion or implied risk of malignancy, although this may not be explicitly stated in report.
    –   EXCLUDES: Renal abscess, pyelonephritis

3. **Adrenal mass:**
    –   Any adrenal lesion/mass/nodule
    –   May be solid or cystic
    –   May be classified/suggested as benign by radiologist (ex: adenoma) and should still include as imaging is not definitive and these may be clinically significant even if benign.

4. **Simple renal cyst:**
    –   Simple cyst with no internal septation/calcification
    –   There should be no mention of risk of malignancy for these (i.e. sometimes cysts are complex and there is a risk of malignancy, these are NOT to be included here).
    –   A patient may have both simple cysts and separate solid/suspicious renal mass, in this case both categories would be included.
    –   EXCLUDES: Hydronephrosis, renal abscess

APPENDIX B. List of a priori identified key words or phrases that might indicate a particular lesion.

1. **Solid renal mass:**
   – Renal cell carcinoma
   – RCC
   – Renal neoplasm
   – Kidney neoplasm
   – Kidney cancer
   – Kidney neoplasm
   – Renal cell neoplasm
   – Renal mass
   – Renal adenocarcinoma
   – Enhancing renal mass
   – Malignant renal mass
   – Malignant kidney mass
   – Kidney metastasis
   – Kidney metastases
   – Renal metastasis
   – Renal metastases
   – Metastasis to kidney
   – Clear cell renal cell carcinoma
   – Clear cell RCC
   – Chromophobe renal cell carcinoma
   – Chromophobe RCC
   – Papillary renal cell carcinoma
   – Papillary RCC
   – Collecting duct carcinoma
   – Renal medullary carcinoma
   – Tuberous sclerosis
   – Von hippel lindau
   – VHL
   – Birt hogg dube
   – Angiomyolipoma
   – AML
   – Fat poor angiomyolipoma
   – Fat poor AML
   – Oncocytoma

–   Renal adenoma
–   Metanephric adenoma
–   Cystic nephroma
–   Reninoma
–   Juxtaglomerular cell tumor

2. **Complex renal cyst:**
    –   Bosniak 2
    –   Bosniak 2F
    –   Bosniak 3
    –   Bosniak 4
    –   Complex renal cyst
    –   Complex kidney cyst
    –   Cystic RCC
    –   Hemorrhagic cyst
    –   Hyperdense cyst
    –   High density cyst
    –   Infected cyst
    –   Benign complex cyst
    –   Cystic nephroma
    –   Mixed epithelial-stromal tumor
    –   Cystic Wilms tumor
    –   Wilms tumor

3. **Adrenal mass:**
    –   Adrenocortical carcinoma
    –   ACC
    –   Adrenal neoplasm
    –   Adrenal adenoma
    –   Adrenal cortical adenoma
    –   Pheochromocytoma
    –   Conn's syndrome
    –   Functional adrenal mass
    –   Functional adrenal nodule
    –   Adrenal metastasis
    –   Adrenal metastases
    –   Metastasis to adrenal
    –   Aldosteronoma

- Hyperaldosteronism
- Li-Fraumeni syndrome
- Beckwith-Wiedemann syndrome
- Cushing syndrome

4. Simple renal cyst:
- Simple cyst
- Parapelvic cyst
- Uncomplicated renal cyst
- Bosniak 1