

1 Routine Urinary Biochemistry Does Not Accurately Predict Stone Type Nor Recurrence in
2 Kidney Stone Formers: A Multi-Centre, Multi-Model, Externally Validated Machine-Learning
3 Study
4

5 Robert M Geraghty, Ian Wilson, Eric Olinger, Paul Cook, Susan Troup, David Kennedy, Alistair
6 Rogers, Bhaskar K Somani, Nasser A Dhayat, Daniel G Fuster, John A Sayer
7

8 Mr Robert Geraghty - Corresponding Author

9 Academic Clinical Fellow in Urology

10 Department of Urology

11 Freeman Hospital

12 Newcastle-upon-Tyne

13 Rob.geraghty@newcastle.ac.uk
14

15 Dr Eric Olinger

16 Post-Doctoral Researcher

17 Translational and Clinical Institute, Faculty of Medical Sciences, Newcastle University,

18 Central Parkway, Newcastle upon Tyne, England, United Kingdom.
19

20 Dr Ian Wilson

21 Lecturer in Biomedical Statistics

22 Biosciences Institute

23 Newcastle University

24 International Centre for Life

25 Central Parkway

26 Newcastle-upon-Tyne
27

28 Dr Paul Cook

29 Consultant Chemical Pathologist

30 Department of Biochemistry

31 University Hospital Southampton

32 Southampton
33

34 Ms Susan Troup

35 Principal Clinical Biochemist

36 Department of Biochemistry

37 Queen Elizabeth Hospital

38 Gateshead
39

40 Dr David Kennedy

41 Consultant Clinical Biochemist

42 Department of Biochemistry

43 Queen Elizabeth Hospital

44 Gateshead
45

46 Mr Alistair Rogers

47 Consultant Urologist

48 Department of Urology
 49 Freeman Hospital
 50 Newcastle-upon-Tyne
 51
 52 Professor Bhaskar Somani
 53 Consultant Urologist and Professor of Urology
 54 Department of Urology
 55 University Hospital Southampton
 56 Southampton
 57
 58 Dr Nasser Dhayat
 59 Consultant Nephrologist
 60 Department of Nephrology and Hypertension, Inselspital, Bern University Hospital,
 61 University of Bern
 62 Switzerland
 63
 64 Professor Daniel Fuster
 65 Consultant Nephrologist and Associate Professor of Nephrology
 66 Department of Nephrology and Hypertension, Inselspital, Bern University Hospital,
 67 Department for Biomedical Research, University of Bern
 68 Switzerland
 69
 70 Professor John Sayer
 71 Professor of Nephrology and Honorary Consultant Renal Physician
 72 Translational and Clinical Institute, Faculty of Medical Sciences, Newcastle University,
 73 Central Parkway, Newcastle upon Tyne, England, United Kingdom.
 74 The Newcastle upon Tyne Hospitals NHS Foundation Trust, Freeman Road, Newcastle upon
 75 Tyne, England, United Kingdom.
 76 National Institute for Health Research Newcastle Biomedical Research Centre, Newcastle
 77 upon Tyne, England, United Kingdom.
 78
 79 **Keywords:** Kidney stones, Machine learning, Prediction, Urinary Biochemistry
 80
 81 **Word Count:**
 82 Abstract: 271/300
 83 Main body: 3574/4000
 84 References:
 85 Figures: 3
 86 Tables: 3
 87 Appendices: 3

88 Abstract

89 **Objectives:** Urinary biochemistry is used to detect and monitor conditions associated with
90 recurrent kidney stones. There are no predictive machine learning (ML) tools for kidney
91 stone type or recurrence. We therefore aimed to build and validate ML models for these
92 outcomes using age, gender, 24-hour urine biochemistry and stone composition.

93 **Materials and Methods:** Data from 3 cohorts were used, Southampton, UK (n=3013),
94 Newcastle, UK (n=5984) and Bern, Switzerland (n=794). Of these 3130 had available 24-hour
95 urine biochemistry measurements (calcium, oxalate, urate, pH, volume), and 1684 had
96 clinical data on kidney stone recurrence. Two predictive models were constructed (UK and
97 Swiss) using two ML techniques (Partitioning and Random Forests [RF]) and validated
98 internally with a subset of the same dataset (e.g UK model/UK test set), and externally with
99 the other dataset (UK model/Swiss test set).

100 **Results and Limitations:** For kidney stone type, on external validation accuracy of UK RF
101 model=0.79 (95% CI: 0.73-0.84), sensitivity: calcium oxalate=0.99 and calcium
102 phosphate/urate=0.00. Specificity: calcium oxalate=0.00 and calcium
103 phosphate/urate=0.99. For the Swiss RF model accuracy=0.87 (95% CI: 0.83-0.89),
104 sensitivity: calcium oxalate=0.99 and calcium phosphate/urate=0.00. Specificity: calcium
105 oxalate=0.00, calcium phosphate=0.00 and urate=1.00.

106 For stone recurrence, on external validation accuracy of UK RF model=0.22 (95% CI:
107 0.19-0.25), sensitivity=0.93 and specificity=0.09. Swiss RF model accuracy=0.42 (95% CI:
108 0.39-0.47), sensitivity=0.03 and specificity=0.97.

109 **Conclusions:** Neither kidney stone type nor kidney stone recurrence can be accurately
110 predicted using modelling tools built using specific 24-hour urinary biochemistry values
111 alone. Further studies to delineate accurate predictive tools should be undertaken using
112 both known and novel risk factors.

Introduction

Kidney stones are prevalent¹, costly² and debilitating, with a recurrence rate of around 40-50% at 10 years³. Although there are rare well described forms of kidney stones such as cystinuria leading to cystine stones, the pathophysiology underlying the most common types of kidney stones (calcium stones and urate) have not been satisfactorily described. Given the rising prevalence and cost of surgical treatment, tools are needed to accurately predict kidney stone type and more importantly, risk of recurrence so that preventative measures can be more focused and beneficial.

Kidney stone composition may not always be investigated, but remains vital for the identification of rarer stone types that require further management. There has previously been only one machine learning (ML) study that looked at prediction of kidney stone type⁴. Several types of ML models were constructed for kidney stone type prediction, achieving high accuracy (area under the curve (AUC)=0.99), although it was neither internally nor externally validated. They did not use urinary biochemistry, instead utilising demographics and basic medical history. Another study applied logistic regression to 24-hour urinary biochemistries to predict stone type, but this demonstrated poor predictive value⁵. Regression models are not as powerful as more modern ML techniques⁶.

The Recurrence of Kidney Stone (ROKS) nomogram was created to predict kidney stone recurrence and is largely based on non-biochemical risk factors and radiologic appearances, rather than biochemistry⁷. It had a moderate predictive value (AUC=0.65) for kidney stone recurrence at 2-years on external validation⁸. Until accuracy improves, this predictive model is not useful for influencing clinical practice.

Twenty four hour urine samples are often taken from high-risk stone formers⁹ to identify stone forming risk factors such as hyperoxaluria, hypercalciuria, hyperuricosuria etc. Although the benefit of urinary biochemistry for managing kidney stone patients has not been fully established¹⁰, there is evidence from randomized controlled trials that medications used in specific, limited circumstances reduce the number and increase the time to, recurrence¹¹. A systematic review by Hsi et al. delineates the limitations of 24-hour urine collection, none of the included studies demonstrate any predictive value for recurrence¹⁰.

To date there have been no predictive ML models for kidney stone type or recurrence based on 24-hour urinary biochemistry. We therefore aimed to construct predictive ML models for these outcomes using clinical and biochemical data from kidney stone patients to assess the predictive value of 24-hour urinary biochemistry.

Methods

Data Collection across all Cohorts

Age, sex, kidney stone recurrence, stone type and 24-hour urinary biochemistry (free-choice diet).

Urinary biochemistry obtained: calcium (mmol/24 hour), oxalate (mmol/24 hour), urate (mmol/24 hour), urine volume (L) and pH. Citrate (mmol/24 hour) was obtained from the Newcastle and Bern cohorts.

Stone type was ascertained using Fourier Transform IR spectroscopy in all cohorts. Differentiation between calcium oxalate types (monohydrate/dihydrate) was not available in the Southampton/Newcastle cohorts and therefore not included in analysis.

160

161 Southampton, UK

162 *Study Population*

163 The cohort consisted of patients with kidney stone disease presenting to University
164 Hospital Southampton, UK (tertiary referral hospital) referred for metabolic assessment
165 between 1990-2007. The initial cross-sectional study (n=2801)¹² and subsequent cohort
166 subset² have been described. During this period, stone formers were routinely referred
167 from across Hampshire, Isle of Wight and Dorset. The subset includes 1000 patients with
168 further information ascertained retrospectively using hospital/general practice electronic
169 records.

170 Patients who had no documentation (i.e no evidence of subsequent follow-
171 up/consultation, lived outside/have left Hampshire or no documentation on CHIE) were
172 excluded [see fig. 1].

173

174 *Ethical Approval*

175 Ethical approval for this arm of the study was granted by the NHS Bristol Research
176 Ethics Committee (Rec ref: 18/SW/0185; IRAS ID: 240061).

177

178 Newcastle upon Tyne, UK

179 *Study Population*

180 The cohort consisted of patients with kidney stone disease presenting to hospitals in
181 the north of England between 2013 and 2018 [Appendix 1]. These include: The Newcastle
182 upon Tyne Hospitals NHS Foundation Trust (Newcastle-upon-Tyne), James Cook University
183 Hospital (Middlesbrough), Sunderland Royal Hospital, Durham University Hospital, Queen
184 Elizabeth Hospital (Gateshead), Darlington Memorial Hospital, Cumberland Infirmary
185 (Carlisle) and Northumbria Hospitals Trust (Wansbeck, Hexham, Cramlington).

186 Data ascertained retrospectively using electronic hospital records. Urinary biochemistry
187 and stone type were obtained from the regional renal calculi processing unit (Queen Elizabeth
188 Hospital, Gateshead).

189

190 *Ethical Approval*

191 This arm was registered as an audit (current patients of authors) with Newcastle-
192 upon-Tyne Hospitals NHS Foundation Trust (audit number: 10337).

193

194 Bern, Switzerland

195 *Study Population*

196 Patients enrolled in the Bern Kidney Stone Registry (BKSR) were referred to the
197 Department of Nephrology and Hypertension, Inselspital, Bern University Hospital and
198 included kidney stone formers suffering at least one stone episode. This population has
199 previously been described in the literature¹³. Those <18 years were excluded.

200

201 *Ethical Approval*

202 The BKSR adheres to the Declaration of Helsinki, was approved by the Ethical
203 Committee of the Kanton Bern (#95/06) and all patients provided written informed consent.

204

205 Definitions

Stone recurrence was defined as subsequent stone surgery or acute presentation following previous treatment/spontaneous passage. Presence of asymptomatic recurrence on imaging was not available.

Low urine volume was defined as <1.5L/24 hour. Acid urine was defined as pH <5.5.

Biochemical abnormalities were defined as follows, as per local laboratory values: hypercalciuria in women >6.2 mmol/24 hour and men >7.5 mmol/24 hour; hyperoxaluria in women >0.32 mmol/24 hour, men >0.49 mmol/24 hour; hyperuricosuria as >4.46 mmol/24 hour and hypocitraturia as <0.6 mmol/24 hour.

Primary kidney stone type was defined as >50% of particular type.

Statistical Analysis

Analysis was performed in R (version no.4.0.2). For robustness, two machine learning models were used: Partitioning (rpart[19]) and Random Forests (randomForest[20] and missForest[21]) for both Stone Type and Recurrence. Missing data was accounted for with values imputed from existing data (missForest). Variables included in both analyses: age, sex, raw 24-hour urinary values (calcium, oxalate, urate, volume), pH and number of biochemical abnormalities. Urinary citrate was excluded due to low numbers with documented citrate. To investigate whether unbalanced data was driving discrimination, up-sampling of training data was used to create balanced training data sets using caret[22], which were run through subsequent analyses.

Recurrence rate was excluded in stone type models as this would be unknown at point of presentation. Stone type was included in the recurrence models. Those with unclear stone type were excluded from the Stone Type models. Data was partitioned into training (70% of total) and test (30%) datasets. Diagnostic accuracy statistics are presented in the results section. Sample sizes were not calculated as an effect size was not being calculated. For robustness, two models were used for each measure, and models were constructed from two cohorts. Sensitivity analysis is represented by the second model (random forests), and by the second dataset (Bern).

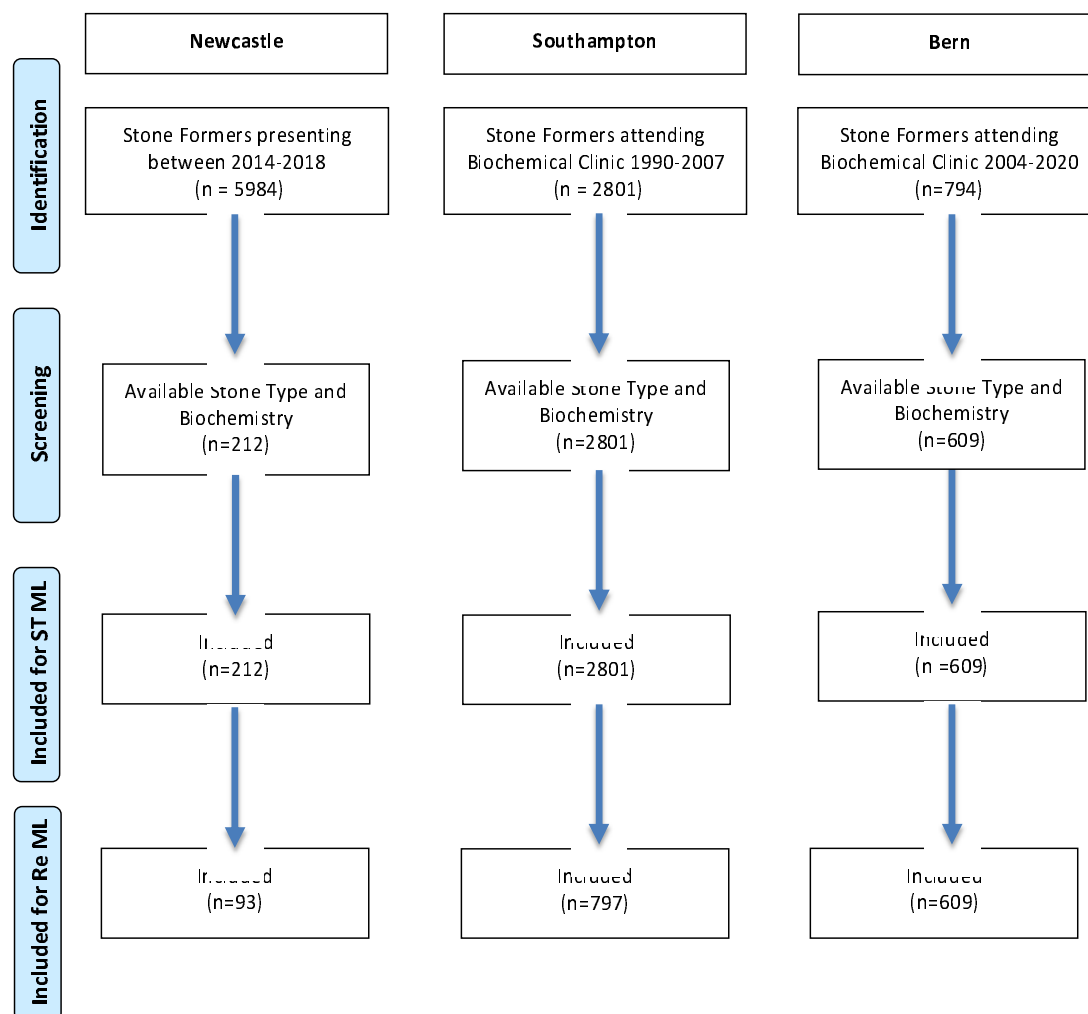


Figure 1. Flow diagram of patient selection/Inclusion. ML=Machine Learning, ST=Stone type, Re=Recurrence.

Results

Patient and Biochemistry Demographics

There were 3013 kidney stone patients in the Southampton/Newcastle combined dataset [see table 1]. There were 2074 men (mean age 49 ± 15 years) and 939 women (mean age 49 ± 17 years). Stone types were as follows: calcium oxalate ($n=1476$, 49%), calcium phosphate ($n=119$, 4%), urate ($n=89$, 3%), cystine ($n=5$, 0.2%), struvite (magnesium ammonium phosphate) ($n=3$, 0.1%) and no data ($n=1321$, 44%). Those with no data are likely to have passed their stone/post treatment fragments (e.g. shockwave lithotripsy) spontaneously. The stone was then not available for analysis. Of these 3013, 890 had information on whether or not the patient had a kidney stone recurrence. Of these, 374 had a recurrence (39%). The low numbers of patients with follow-up information are likely due to the patient not accessing medical services following the initial stone episode, moving away or dying.

There were 794 kidney stone patients in the Bern dataset, with 204 women (mean age 45 ± 15 years) and 590 men (mean age 48 ± 13 years). Stone types were as follows: calcium oxalate ($n=598$, 76%), calcium phosphate ($n=81$, 10%), urate ($n=47$, 6%), cystine ($n=16$, 2%), struvite ($n=6$, 0.8%) and no data ($n=40$, 5.1%) [see fig 2]. All had follow-up data available for stone recurrence. There were 676 recurrences (85%).

Across all 3 datasets, 3130 patients had full data on urinary biochemistry. There were 631 (20.1%) patients who had no biochemical diagnoses, 1331 (42.5%) patients with a single diagnosis and 1668 (37.3%) with a combination of biochemical abnormalities.

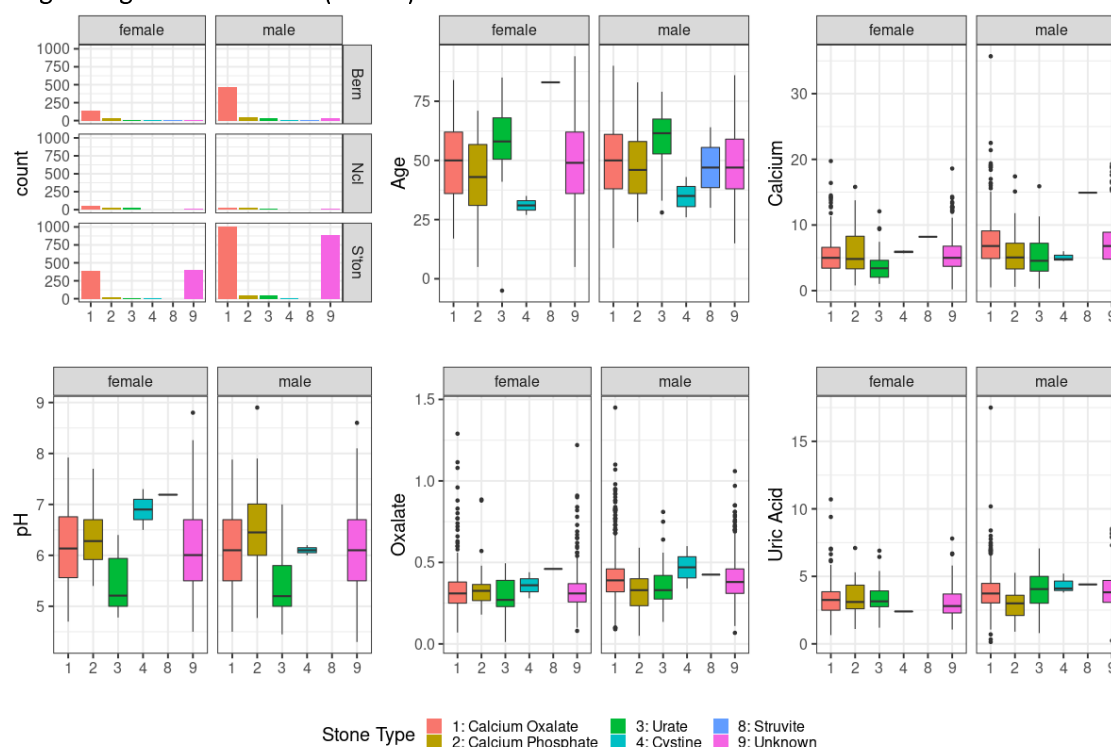


Figure 2. Graphical representation of differences between cohorts and sexes. On top row, from left to right: bar chart of number of patients in each sex, in each cohort by stone type; boxplot of age for each sex per stone type; boxplot of 24-hour urinary calcium (mmol/24 hour) for each sex per stone type. Bottom row, from left to right: boxplot of urinary pH for each sex per stone type; boxplot of 24-hour urinary oxalate (mmol/24 hour) for each sex per stone type; boxplot of 24 hour urinary uric acid (mmol/24 hour) for each sex per stone type.

Of those with full data on urinary biochemistry, 1002 patients had data on kidney stone recurrence. In total 688 (68.7%) had a recurrence. There were 272 (74.7%) recurrences in those with no biochemical diagnosis, 258 (67.0%) with a single biochemical risk factor and 163 (64.4%) in those with a combination [see appendices 2 & 3].

The Southampton dataset did not contain 24-hour urinary citrate measurements. In the Newcastle & Bern datasets there were 45 and 121 patients, respectively, with hypocitraturia. Recurrence data was available in 144 patients, with recurrence in 11/23 (Newcastle, 47.8%) and 109/121 (Bern, 90.0%), respectively.

Discordant primary kidney stone type and 24-hour urine biochemistry was seen in 103 patients (5% of known stone type). Cystine stones (n=21) and struvite stones (n=9) were not included in the appendices given their low numbers and alternative aetiologies.

Data on subsequent primary stone type was available in the Bern dataset for a subset of those with a stone recurrence (n=188). Change in primary stone type from initial episode was seen in 42 patients (22%): calcium oxalate (n=26, 18%), calcium phosphate (n=10, 37%), urate (n=5, 38%), cystine (n=0, 0%), struvite (n=1, 100%) [Figure 3].

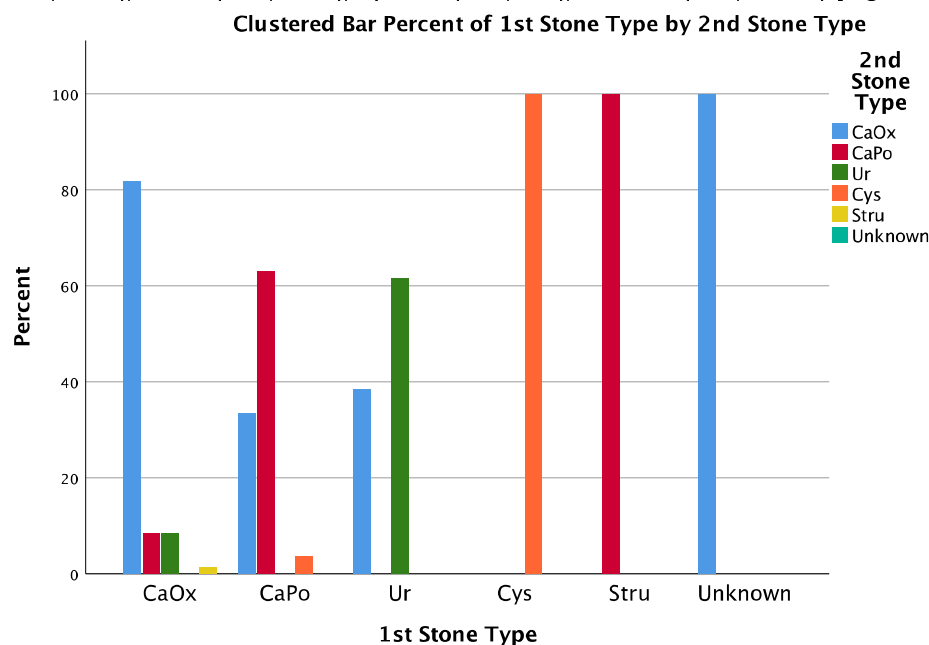


Figure 3. Bar chart showing percentage of subsequent stone type for each first stone type, data from the Bern Cohort. CaOx=Calcium Oxalate, CaPo=Calcium Phosphate, Ur=Urate, Cys=Cystine, Stru=Struvite

285

		Southampton	Newcastle	Combined Southampton & Newcastle	Bern
Total, n		2801	212	3013	794
Sex	Female, n	818 (29%)	91 (43%)	939 (31%)	204 (26%)
	Female, Age±SD	49±16	50±19	49±17	45±15
	Male, n	1984 (71%)	121 (57%)	2073 (69%)	590 (74%)
	Male, Age±SD	49±14	48±16	49±15	48±13
Stone Type	Calcium Oxalate, n (%)	1393 (50%)	83 (39%)	1476 (49%)	598 (75.9%)
	Calcium Phosphate, n (%)	59 (2%)	60 (28%)	119 (4.0%)	81 (10.3%)
	Urate, n (%)	55 (2%)	34 (16%)	89 (3.0%)	47 (6.0%)
	Cystine, n (%)	5 (0.2%)	0	5 (0.2%)	16 (2.0%)
	Struvite, n (%)	1 (0.04%)	2 (1%)	3 (0.1%)	6 (0.8%)
	No data, n (%)	1288 (46%)	33 (16%)	1321 (44%)	40 (5.1%)
Recurrence, n/total with data (%)		325/855 (38%)	49/111 (44%)	374/890 (39%)	676/794 (85.1%)

286 *Table 1. Patient and Kidney Stone Demographics. Age in years.*

287 Machine Learning Models:

288 Stone Type

289 The training group had n=3092 patients, n=2360 of which had immediate clearance on
290 fluoroscopy. On internal validation (test set: n with outcome/total; n=1015/1326), the
291 diagnostic accuracy statistics were: Random forests (RF) AUC=0.73, accuracy= 0.77 (95%
292 CI:0.74-0.79), sensitivity=0.27, specificity=0.94; Partitioning AUC=0.63, accuracy=0.71(95%
293 CI:0.69-0.74), sensitivity=0.22, specificity=0.89; Extreme Gradient Boosting (XGBoost)
294 AUC=0.75, accuracy=0.77 (95% CI:0.75-0.80), sensitivity=0.27, specificity=0.94; Logistic
295 Regression (LR) AUC=0.63, accuracy=0.76 (95% CI:0.73-0.78), sensitivity=0.12,
296 specificity=0.97; Classical Neural Network (NN) AUC=0.74, accuracy=0.75 (95% CI:0.73-0.78),
297 sensitivity=0.23, specificity=0.94; Bayesian Generalised Linear Model (BGLM)
298 AUC=0.75, accuracy=0.77 (95% CI:0.75-0.80), sensitivity=0.29, specificity=0.94; Deep Neural
299 Network (DNN) Single-outcome model AUC=0.59, accuracy=0.77 (95% CI:0.75-0.79),
300 sensitivity=0.57, specificity=0.79; DNN Multiple-outcome model AUC=0.53,
301 accuracy=0.77 (95% CI:0.75-0.79), sensitivity=0.60, specificity=0.78 [see figure 2
302 and supplementary table 2].

Two models were built to predict kidney stone type, partitioning and Random Forests (RF). Both models used all available 24 hour urinary biochemical data (volume, calcium, oxalate, urate and pH), age and sex. Citrate was excluded as the Southampton data did not have this. Due to missing values (as detailed above), some biochemical values were imputed. Up-sampling of rare stone data, while marginally improving classification into rare stone types, overall gave inferior classification, therefore simple training sets are reported.

For the internally validated Southampton/Newcastle models, the partitioning model accuracy was 0.89 (95% CI: 0.85-0.91) and RF model accuracy was 0.90 (95% CI: 0.87-0.92) [see table 3].

For the internally validated Bern models, the partitioning model accuracy was 0.75 (95% CI: 0.69-0.81) and RF model was 0.81 (95% CI: 0.78-0.84).

None of the models were able to predict cystine/struvite stones.

Each RF model was externally validated using the other dataset. For the Southampton/Newcastle model, validated with Bern data, the accuracy was 0.79 (95% CI: 0.73-0.84). The sensitivity for: calcium oxalate=0.99, calcium phosphate/urate=0.00. The specificity for: calcium oxalate=0.00 and calcium phosphate/urate=0.99.

For the Bern model, validated with Southampton/Newcastle data, the accuracy was 0.87 (95% CI: 0.83-0.89). The sensitivity for: calcium oxalate=0.99, calcium phosphate=0.00 and urate=0.04. The specificity for: calcium oxalate=0.02, calcium phosphate=0.99 and urate=1.00.

Addition of citrate into the models with the Newcastle and Bern datasets did not demonstrate any difference in accuracy, sensitivity or specificity for stone type. Increasing the predominant stone type threshold for calcium phosphate stones to 90% did not increase accuracy, sensitivity or specificity.

Recurrence

Two models were used to predict recurrence (Partitioning and RF). Both models used all biochemistry previously described, age and sex. Biochemical value imputation as above.

For the internally validated Southampton/Newcastle models, the partitioning model accuracy was 0.52 (95% CI: 0.45-0.60), correctly predicting 25/72 recurrences and 70/109 without. The RF model accuracy was 0.55 (95% CI: 0.48-0.63), correctly predicting 25/72 recurrences and 75/109 without [see Table 4].

For the internally validated Bern models, the partitioning model accuracy was 0.85 (95% CI: 0.80-0.89), correctly predicting 201/212 recurrences, and 2/27 without. The RF model accuracy was 0.89 (95% CI: 0.84-0.93), correctly predicting 163/165 recurrences, and 1/19 without.

Recurrence	Southampton/Newcastle Model		Bern Model	
	Partitioning	Random Forests	Partitioning	Random Forests
Accuracy (95% CI)	0.52 (0.45-0.60)	0.55 (0.48-0.63)	0.85 (0.80-0.89)	0.89 (0.84-0.93)
Sensitivity	0.64	0.69	0.07	0.05
Specificity	0.35	0.35	0.95	0.99
PPV	0.60	0.61	0.15	0.33
NPV	0.40	0.42	0.89	0.90
Detection Rate	0.39	0.41	0.008	0.005

Table 2. Descriptive statistics for Partitioning and Random Forests models predicting recurrence. PPV=positive predictive value, NPV=negative predictive value. Test set=30% of same cohort

Each RF model was externally validated as above. The Southampton/Newcastle model, validated with Bern data, accuracy was 0.22 (95% CI: 0.19-0.25), sensitivity=0.93, specificity=0.09, correctly predicting 55/632 recurrences and 108/116 without.

The Bern model, validated with Southampton/Newcastle data, accuracy was 0.42 (95% CI: 0.39-0.47), sensitivity=0.03, specificity=0.97, correctly predicting 235/242 recurrences and 11/332 without.

Addition of citrate into the models with the Newcastle and Bern datasets did not demonstrate any difference in accuracy, sensitivity or specificity for stone type. Increasing the predominant stone type threshold for calcium phosphate stones to 90% did not increase accuracy, sensitivity or specificity.

Discussion

This study utilised 24-hour urinary biochemistry data from kidney stone patients attending three European tertiary referral centres to make two cohorts (Southampton/Newcastle and Bern), used to construct internally validated machine learning (ML) models for both kidney stone type and kidney stone recurrence. These models were then externally validated on the other cohort. The recurrence models had very poor accuracy, whilst the kidney stone type models demonstrated high overall accuracy. However this is likely due to the high number of calcium oxalate kidney stones in our datasets

(reflected in the literature [23]), the diagnostic statistics demonstrate very poor predictive value for other stone types. In essence the models predict every stone as calcium oxalate.

The poor predictive value for both outcomes are important negative findings. The composition of most kidney stones is unknown (36% in this study and $\leq 85\%$ in the literature[24]), and 24-hour urine collection is used to identify recurrent/high-risk stone formers who could be started on prophylaxis [11,12].

The main strengths of this study are the large cohort sizes, internal & external validation[19,20], multiple ML models (superior to regression analyses[8]) showing similar outcomes and consistent outcomes across both datasets.

There are several weaknesses with this study. Few of the Newcastle kidney stone patient cohort had 24-hour urinary biochemical analysis available, and fewer (in both datasets) had data on recurrence. This was addressed with imputation of missing data[21]. Unfortunately, 24-hour urinary citrate was not available in the Southampton dataset, which may bias our main results as hypocitraturia is an important risk factor for calcium-containing kidney stones[25]. To investigate this both models were repeated with citrate included for the Bern/Newcastle datasets, this did not demonstrate any improvement in diagnostic statistics. The datasets did not include 24-hour urinary phosphate or sodium, both of which influence the risk of stone formation[26]. Therefore, predictive value may be improved with the addition of these variables. The stone type models could not predict struvite or cystine stones. This is likely due to minimal numbers of these stones, thus underpowering the model, and their alternative pathophysiology[27,28]. Although it is accepted that those with cystinuria, are at high risk of recurrence[28].

The datasets differ in terms of stone types proportions, notably uric acid and calcium phosphate stones. For instance, in the Southampton and Bern cohorts uric acid stones represent 3.6% and 6% respectively, whilst in the Newcastle dataset 19% have these (10% of total 5984 patients). Uric acid stones are associated with obesity/metabolic syndrome[29], and this observation may represent rising levels in Newcastle compared to an historic population (Southampton), or a contemporary European population (Bern). This could also represent referral bias, e.g. in the Bern cohort referral rates for urate stones were historically lower than present. As information on BMI or diabetes mellitus was not available this limits the predictive value of the models to predict stone type.

Proportions of calcium phosphate stones are also different between datasets (2% in Southampton, whilst 28% in Newcastle). This may be due to selection bias, as large numbers of patients in the Newcastle set did not have urinary biochemistry available. This may also be due to our definition of predominant stone type (i.e $>50\%$). Pure calcium phosphate stones ($>90\%$) are associated with a specific phenotype (alkaline urine and hypocitraturia) [30]. Increasing the cut-off for calcium phosphate stones to $>90\%$ did not demonstrate an increase in predictive power for either stone type or recurrence. This may be due to the low numbers with this stone type (UK: from 4% to 2.8%, Bern: 10.3% to 2.6%). This touches on the wider issue of stone type interpretation. There are no guidelines on how to interpret stone composition, nor are there consistent methods within the literature. It is well known that stones can take many forms/compositions, but translation of these observations has yet to reach everyday practice. Our use of a $>50\%$ cut-off was an attempt to use standard laboratory outputs to improve model accessibility.

As well as differences in stone type, recurrence data also differed between datasets, Bern had a far higher rate than the UK cohorts (which are similar to the literature[5]). This is likely due to the selection of patients going to biochemical clinic/tertiary care and will likely

skew the results as these patients are not representative of the general stone forming population. This is reflected in the recurrence models, the Bern model predict most as recurrence, whilst the Southampton/Newcastle model predict most as no recurrence.

24-hour urinary biochemistry is touted as a good measure to identify urinary risk factors for stone formation[11,12]. In certain circumstances this is true[15], however there are a significant minority of recurrent kidney stone formers who have no biochemical abnormalities (20% in these cohorts and similar to rates seen in other studies[14,31]). This may be due to dietary influence on urinary biochemistry, with a random diet unpredictably influencing the results[32]. Further, a small proportion of patients have discordant urinary biochemistry and stone type (3.3% across these cohorts), which makes prediction of unknown stone type unreliable. In summary, the main challenge to model accuracy was not unbalanced data but a lack of factors that could provide effective discrimination.

There is a growing body of evidence that kidney stone disease is the common end point of a heterogenous group of disorders that include environmental[33], dietary[34,35], monogenic[36] and polygenic risk factors[37,38], whether in isolation or combination. Therefore reliance on urinary biochemistry alone as a predictive tool, as this study demonstrates, is inadvisable. This study, along with a recent study demonstrating the futility of serum biochemistry (except calcium)[39], demonstrate that current practices[11,12] to delineate risk of kidney stone recurrence do not work well. At least 15% of kidney stones have a monogenic cause[36], with a significant proportion forming calcium stones[40]. It is evident that a precision medicine approach is needed to differentiate the differing causes of KSD, which may then be treated effectively[41].

Future studies are needed to delineate an effective kidney stone screen based on traditional risk factors in combination with kidney stone analysis, 24-hour urinary biochemistry, monogenic screen and polygenic risk scoring.

Conclusion

Neither kidney stone type nor kidney stone recurrence can be accurately predicted by 24-hour urinary biochemistry. Further studies to delineate an effective stone screen should be undertaken.

Acknowledgements

Robert Geraghty is a National Institute for Health Research Academic Clinical Fellow. Eric Olinger is supported by an Early Postdoc Mobility Stipendium of the Swiss National Science Foundation and Kidney Research UK. John Sayer is funded by Kidney Research UK and the Northern Counties Kidney Research Fund.

We would like to acknowledge Dr Valerie Walker for initiating the Southampton stone database.

References

- [1] Antonelli JA, Maalouf NM, Pearle MS, Lotan Y. Use of the National Health and Nutrition Examination Survey to Calculate the Impact of Obesity and Diabetes on Cost and Prevalence of Urolithiasis in 2030. *European Urology* 2014;66:724–9. doi:10.1016/j.eururo.2014.06.036.

- 452 [2] Geraghty RM, Cook P, Walker V, Somani BK. Evaluation of the economic burden of
453 kidney stone disease in the UK: a retrospective cohort study with a mean follow-up
454 of 19 years. *BJU Int* 2020;125:586–94. doi:10.1111/bju.14991.
- 455 [3] Lien C-S, Huang C-P, Chung C-J, Lin C-L, Chang C-H. Increased risk of anxiety among
456 patients with urolithiasis: A nationwide population-based cohort study. *Int J Urol*
457 2015;22:937–42. doi:10.1111/iju.12865.
- 458 [4] Canales BK, Sharma N, Yuzhakov SV, Bozorgmehri S, Otto BJ, Bird VG. Long-term
459 Recurrence Rates in Uric Acid Stone Formers With or Without Medical
460 Management. *Urology* 2019;131:46–52. doi:10.1016/j.urology.2019.05.023.
- 461 [5] Ljunghall S, Danielson BG. A Prospective Study of Renal Stone Recurrences. *British*
462 *Journal of Urology* 1984;56:122–4. doi:10.1111/j.1464-410X.1984.tb05346.x.
- 463 [6] Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using
464 ensemble learning. *Artif Intell Med* 2018;84:117–26.
465 doi:10.1016/j.artmed.2017.12.001.
- 466 [7] Moreira DM, Friedlander JI, Hartman C, Elsamra SE, Smith AD, Okeke Z. Using 24-
467 hour urinalysis to predict stone type. *J Urol* 2013;190:2106–11.
468 doi:10.1016/j.juro.2013.05.115.
- 469 [8] Boulesteix AL, Schmid M. Machine learning versus statistical modeling. *Biometrical*
470 *Journal* 2014;56:588–93. doi:10.1002/bimj.201300226.
- 471 [9] Rule AD, Lieske JC, Li X, Melton LJ, Krambeck AE, Bergstralh EJ. The ROKS nomogram
472 for predicting a second symptomatic stone episode. *J Am Soc Nephrol*
473 2014;25:2878–86. doi:10.1681/ASN.2013091011.
- 474 [10] Iremashvili V, Li S, Penniston KL, Best SL, Hedican SP, Nakada SY. External Validation
475 of the Recurrence of Kidney Stone Nomogram in a Surgical Cohort. *J Endourol*
476 2019;33:475–9. doi:10.1089/end.2018.0893.
- 477 [11] Türk C, Petřík A, Sarica K, Seitz C, Skolarikos A, Straub M, et al. EAU Guidelines on
478 Interventional Treatment for Urolithiasis. *European Urology* 2016;69:475–82.
479 doi:10.1016/j.eururo.2015.07.041.
- 480 [12] Pearle MS, Goldfarb DS, Assimos DG, Curhan G, Denu-Ciocca CJ, Matlaga BR, et al.
481 AUA Guidelines Medical Management of Kidney Stones: AUA Guideline. *Juro*
482 2014;192:316–24. doi:10.1016/j.juro.2014.05.006.
- 483 [13] Eisner BH, Goldfarb DS. A Nomogram for the Prediction of Kidney Stone Recurrence.
484 *Journal of the American Society of Nephrology* 2014:1–3.
485 doi:10.1681/ASN.2013.111180.
- 486 [14] Hsi RS, Sanford T, Goldfarb DS, Stoller ML. The Role of the 24-Hour Urine Collection
487 in the Prevention of Kidney Stone Recurrence. *J Urol* 2017;197:1084–9.
488 doi:10.1016/j.juro.2016.10.052.
- 489 [15] Fink HA, Wilt TJ, Eidman KE, Garimella PS, MacDonald R, Rutks IR, et al. Medical
490 management to prevent recurrent nephrolithiasis in adults: a systematic review for
491 an American College of Physicians Clinical Guideline. *Ann Intern Med* 2013;158:535–
492 43. doi:10.7326/0003-4819-158-7-201304020-00005.
- 493 [16] Walker V, Stansbridge EM, Griffin DG. Demography and biochemistry of 2800
494 patients from a renal stones clinic. *Annals of Clinical Biochemistry* 2013;50:127–39.
495 doi:10.1258/acb.2012.012122.
- 496 [17] Dhayat NA, Lüthi D, Schneider L, Mattmann C, Vogt B, Fuster DG. Distinct
497 phenotype of kidney stone formers with renal phosphate leak. *Nephrol Dial*
498 *Transplant* 2019;34:129–37. doi:10.1093/ndt/gfy170.

- 499 [18] Dhayat NA, Schaller A, Albano G, Poindexter J, Griffith C, Pasch A, et al. The
500 Vacuolar H⁺-ATPase B1 Subunit Polymorphism p.E161K Associates with Impaired
501 Urinary Acidification in Recurrent Stone Formers. *J Am Soc Nephrol* 2016;27:1544–
502 54. doi:10.1681/ASN.2015040367.
- 503 [19] Therneau TM, Atkinson B, Ripley MB. The rpart package 2010.
- 504 [20] Segal MR. Machine Learning Benchmarks and Random Forest Regression 2004.
- 505 [21] Stekhoven DJ. missForest: Nonparametric missing value imputation using random
506 forest. *Astrophysics Source Code Library* n.d.:ascl:1505.011.
- 507 [22] Kuhn M. caret: Classification and Regression Training. *Astrophysics Source Code*
508 *Library* n.d.:ascl:1505.003.
- 509 [23] Daudon M, Lacour B, Jungers P. Influence of body size on urinary stone composition
510 in men and women. *Urological Research* 2006;34:193–9. doi:10.1007/s00240-006-
511 0042-8.
- 512 [24] Yasui T, Iguchi M, Suzuki S, Kohri K. Prevalence and Epidemiological Characteristics
513 of Urolithiasis in Japan: National Trends Between 1965 and 2005. *Urology*
514 2008;71:209–13. doi:10.1016/j.urology.2007.09.034.
- 515 [25] Eisner BH, Sheth S, Dretler SP, Herrick B, Pais VM. Abnormalities of 24-hour urine
516 composition in first-time and recurrent stone-formers. *Urology* 2012;80:776–9.
517 doi:10.1016/j.urology.2012.06.034.
- 518 [26] Curhan GC, Taylor EN. 24-h uric acid excretion and the risk of kidney stones. *Kidney*
519 *International* 2008;73:489–96. doi:10.1038/sj.ki.5002708.
- 520 [27] Flannigan R, Choy WH, Chew B, Lange D. Renal struvite stones--pathogenesis,
521 microbiology, and management strategies. *Nat Rev Urol* 2014;11:333–41.
522 doi:10.1038/nrurol.2014.99.
- 523 [28] Chillarón J, Font-Llitjós M, Fort J, Zorzano A, Goldfarb DS, Nunes V, et al.
524 Pathophysiology and treatment of cystinuria. *Nat Rev Nephrol* 2010;6:424–34.
525 doi:10.1038/nrneph.2010.69.
- 526 [29] Geraghty R, Abdi A, Somani B, Cook P, Roderick P. Does chronic hyperglycaemia
527 increase the risk of kidney stone disease? results from a systematic review and
528 meta-analysis. *BMJ Open* 2020;10:e032094. doi:10.1136/bmjopen-2019-032094.
- 529 [30] Grases F, Costa-Bauzá A, Ramis M, Montesinos V, Conte A. Simple classification of
530 renal calculi closely related to their micromorphology and etiology. *Clin Chim Acta*
531 2002;322:29–36. doi:10.1016/s0009-8981(02)00063-3.
- 532 [31] Gentle DL, Stoller ML, Bruce JE, Leslie SW. Geriatric urolithiasis. *Juro*
533 1997;158:2221–4. doi:10.1016/s0022-5347(01)68203-x.
- 534 [32] Fuster DG, Morard GA, Schneider L, Mattmann C, Lüthi D, Vogt B, et al. Association
535 of urinary sex steroid hormones with urinary calcium, oxalate and citrate excretion
536 in kidney stone formers. *Nephrol Dial Transplant* 2020. doi:10.1093/ndt/gfaa360.
- 537 [33] Yamout H, Goldberg S. Genetic and Environmental Risk Factors for Kidney Stones.
538 *Nutritional and Medical Management of Kidney Stones*, vol. 62, Cham: Humana,
539 Cham; 2019, pp. 43–52. doi:10.1007/978-3-030-15534-6_3.
- 540 [34] Hiatt RA, Ettinger B, Caan B, Quesenberry CP, Duncan D, Citron JT. Randomized
541 controlled trial of a low animal protein, high fiber diet in the prevention of recurrent
542 calcium oxalate kidney stones. *Am J Epidemiol* 1996;144:25–33.
- 543 [35] Fink HA, Akornor JW, Garimella PS, MacDonald R, Cutting A, Rutks IR, et al. Diet,
544 fluid, or supplements for secondary prevention of nephrolithiasis: a systematic

545 review and meta-analysis of randomized trials. *European Urology* 2009;56:72–80.
546 doi:10.1016/j.eururo.2009.03.031.

547 [36] Halbritter J, Baum M, Hynes AM, Rice SJ, Thwaites DT, Gucev ZS, et al. Fourteen
548 monogenic genes account for 15% of nephrolithiasis/nephrocalcinosis. *J Am Soc*
549 *Nephrol* 2015;26:543–51. doi:10.1681/ASN.2014040388.

550 [37] Paranjpe I, Tsao N, Judy R, Paranjpe M, Chaudhary K, Klein D, et al. Derivation and
551 Validation of Genome Wide Polygenic Score for Urinary Tract Stone Diagnosis.
552 *Kidney International* 2020. doi:10.1016/j.kint.2020.04.055.

553 [38] Howles SA, Wiberg A, Goldsworthy M, Bayliss AL, Gluck AK, Ng M, et al. Genetic
554 variants of calcium and vitamin D metabolism in kidney stone disease. *Nat Commun*
555 2019;10:5175–10. doi:10.1038/s41467-019-13145-x.

556 [39] Eyre KS, Lewis F, Cui H, Grout E, Mihai R, Turney BW, et al. Utility of blood tests in
557 screening for metabolic disorders in kidney stone disease. *BJU Int* 2020.
558 doi:10.1111/bju.15250.

559 [40] Hill F, Sayer JA. Precision medicine in renal stone-formers. *Urolithiasis* 2019;47:99–
560 105. doi:10.1007/s00240-018-1091-5.

561 [41] Sayer JA, Moolhala SH, Thomas DJ. The Medical Management of Urolithiasis:.
562 *British Journal of Medical and Surgical Urology* 2010;3:87–95.
563 doi:10.1016/j.bjmsu.2010.02.004.

564



565

566

Appendix 1. Study centre locations.

Southampton + Newcastle 24-hour urine finding	Total, n (%) (total n=2521)	Stone Type, n (% of Biochemical Diagnosis)				Recurrence, n/total n (total with data on Recurrence, n=391)
		Calcium Oxalate	Calcium Phosphate	Urate	Unknown	
No abnormality	520 (20.6%)	256 (49.2%)	29 (5.6%)	7 (1.3%)	227 (43.7%)	40/95
Low Urine Volume	344 (13.6%)	166 (48.3%)	11 (3.2%)	3 (0.9%)	164 (47.7%)	21/53
Hypercalciuria	333 (13.2%)	169 (50.8%)	7 (2.1%)	3 (0.9%)	154 (46.2%)	19/39
Hyperoxaluria	172 (6.8%)	84 (48.8%)	4 (2.3%)	1 (0.6%)	81 (47.1%)	6/22
Acid urine	167 (6.6%)	72 (43.1%)	2 (1.2%)	9 (5.4%)	84 (50.3%)	11/32
Hypercalciuria + Hyperoxaluria	134 (5.3%)	66 (49.3%)	5 (3.7%)	0	63 (47.0%)	6/17
Hypercalciuria + Low Urine Volume	128 (5.1%)	70 (54.7%)	2 (1.6%)	1 (0.8%)	55 (43.0%)	9/20
Acid urine + Low Urine Volume	121 (4.8%)	51 (42.1%)	0	11 (9.1%)	59 (48.8%)	6/16
Hypercalciuria + Acid Urine	81 (3.2%)	35 (43.2%)	2 (2.5%)	1 (1.2%)	43 (53.1%)	1/4
Hyperoxaluria + Low Urine Volume	66 (2.6%)	38 (57.6%)	1 (1.5%)	0	27 (40.9%)	2/5
Hyperoxaluria + Acid Urine	47 (1.9%)	25 (53.2%)	1 (2.1%)	1 (2.1%)	20 (42.6%)	0/0
Hyperuricosuria	46 (1.8%)	25 (54.3%)	1 (2.2%)	2 (4.3%)	18 (39.1%)	3/10
Hypercalciuria + Hyperuricosuria	46 (1.8%)	32 (69.6%)	0	0	14 (30.4%)	5/10
Hyperoxaluria + Acid Urine + Low Urine Volume	33 (1.3%)	19 (57.6%)	0	2 (6.1%)	12 (36.4%)	2/7
Hypercalciuria + Hyperoxaluria + Hyperuricosuria	30 (1.2%)	12 (40.0%)	0	1 (3.3%)	17 (56.7%)	4/9
Hypercalciuria + Hyperoxaluria + Acid Urine	29 (1.2%)	15 (51.7%)	0	1 (3.4%)	13 (44.8%)	4/5
Hypercalciuria + Hyperoxaluria + Low Urine Volume	27 (1.1%)	17 (63.0%)	0	0	9 (33.3%)	1/1
Hypercalciuria + Acid Urine + Low Urine Volume	25 (1.0%)	13 (52.0%)	0	0	12 (48.0%)	2/2
Hyperuricosuria + Low Urine Volume	23 (0.9%)	8 (34.8%)	1 (4.3%)	2 (8.7%)	11 (47.8%)	1/2
Hyperuricosuria + Acid Urine	22 (0.9%)	9 (40.9%)	0	3 (13.6%)	10 (45.5%)	4/7
Hyperoxaluria + Hyperuricosuria	21 (0.8%)	11 (52.4%)	0	2 (9.5%)	8 (38.1%)	4/9
Hypercalciuria + Hyperuricosuria + Low Urine Volume	19 (0.8%)	8 (42.1%)	0	0	11 (57.9%)	2/5
Hypercalciuria + Hyperuricosuria	18 (0.7%)	10 (55.6%)	0	1 (5.6%)	7 (38.9%)	2/5

+ Acid Urine						
Hyperuricosuria + Acid Urine + Low Urine Volume	13 (0.4%)	6 (46.2%)	0	4 (30.8%)	3 (23.1%)	2/2
Hypercalciuria + Hyperoxaluria + Hyperuricosuria + Acid Urine	13 (0.4%)	6 (46.2%)	0	3 (23.1%)	4 (30.8%)	6/8
Hyperoxaluria + Hyperuricosuria + Low Urine Volume	10 (0.3%)	3 (30.0%)	1 (10.0%)	0	6 (60.0%)	1/1
Hyperoxaluria + Hyperuricosuria + Acid Urine	8 (0.3%)	4 (50.0%)	0	0	4 (50.0%)	0/1
Hypercalciuria + Hyperoxaluria + Acid Urine + Low Urine Volume	7 (0.2%)	3 (42.9%)	0	0	4 (57.1%)	0/2
Hypercalciuria + Hyperuricosuria + Acid Urine + Low Urine Volume	7 (0.2%)	3 (42.9%)	0	1 (14.3%)	3 (42.9%)	2/2
Hypercalciuria + Hyperoxaluria + Hyperuricosuria + Low Urine Volume	6 (0.2%)	4 (66.7%)	0	0	2 (33.3%)	0/1
Hyperoxaluria + Hyperuricosuria + Acid Urine + Low Urine Volume	4 (0.1%)	3 (75.0%)	0	0	1 (25.0%)	0/1
Hypercalciuria + Hyperoxaluria + Hyperuricosuria + Acid Urine + Low Urine Volume	1 (0.0%)	1 (100.0%)	0	0	0	0/0

Appendix 2. Biochemistry diagnoses and recurrences for Newcastle/Southampton dataset.

Bern 24-hour urine finding	Total, n (%) (total n=609)	Stone Type, n (% of Biochemical Diagnosis)				Recurrence, n/total n (total with data on Recurrence, n=609)
		Calcium Oxalate	Calcium Phosphate	Urate	Unknown	
No abnormality	111 (18.2%)	85	13	3	3	97/111
Low Urine Volume	80 (13.1%)	65	5	3	4	66/80
Hypercalciuria	53 (8.7%)	43	7	0	2	46/53
Hyperoxaluria	78 (12.8%)	50	13	6	4	71/78
Acid urine	50 (8.2%)	36	2	6	4	41/50
Hypercalciuria + Hyperoxaluria	24 (3.9%)	19	4	0	1	23/24
Hypercalciuria + Low Urine Volume	12 (2.0%)	11	0	0	1	11/12
Acid urine + Low Urine Volume	31 (7.1%)	24	0	5	2	26/31
Hypercalciuria + Acid Urine	5 (0.8%)	5	0	0	0	4/5
Hyperoxaluria + Low Urine Volume	15 (2.5%)	12	0	1	2	11/15
Hyperoxaluria + Acid Urine	32 (5.3%)	26	3	1	2	25/32
Hyperuricosuria	8 (1.3%)	6	1	1	0	8/8
Hypercalciuria + Hyperuricosuria	15 (2.5%)	12	2	0	1	13/15
Hyperoxaluria + Acid Urine + Low Urine Volume	11 (1.8%)	8	0	3	0	11/11
Hypercalciuria + Hyperoxaluria + Hyperuricosuria	15 (2.5%)	11	2	0	2	14/15
Hypercalciuria + Hyperoxaluria + Acid Urine	9 (1.5%)	7	0	1	1	8/9
Hypercalciuria + Hyperoxaluria + Low Urine Volume	2 (0.3%)	2	0	0	0	2/2
Hypercalciuria + Acid Urine + Low Urine Volume	6 (1.0%)	5	0	1	0	5/6
Hyperuricosuria + Low Urine Volume	4 (0.7%)	4	0	0	0	3/4
Hyperuricosuria + Acid Urine	7 (1.1%)	5	2	0	0	5/7
Hyperoxaluria + Hyperuricosuria	11 (1.8%)	8	0	0	1	10/11
Hypercalciuria + Hyperuricosuria + Low Urine Volume	2 (0.3%)	1	1	0	0	2/2
Hypercalciuria + Hyperuricosuria + Acid Urine	7 (1.1%)	5	2	0	0	7/7
Hyperuricosuria + Acid Urine + Low Urine Volume	2 (0.3%)	2	0	0	0	2/2
Hypercalciuria + Hyperoxaluria + Hyperuricosuria + Acid Urine	5 (0.8%)	4	1	0	0	4/5
Hyperoxaluria + Hyperuricosuria + Low Urine Volume	1 (0.2%)	1 (100%)	0	0	0	1/1

Hyperoxaluria + Hyperuricosuria + Acid Urine	4 (0.7%)	3 (75%)	0	1 (25%)	0	4/4
Hypercalciuria + Hyperoxaluria + Acid Urine + Low Urine Volume	4 (0.7%)	3 (75%)	0	1 (25%)	0	4/4
Hypercalciuria + Hyperuricosuria + Acid Urine + Low Urine Volume	2 (0.3%)	1 (50%)	0	1 (50%)	0	1/2
Hypercalciuria + Hyperoxaluria + Hyperuricosuria + Low Urine Volume	1 (0.2%)	1 (100%)	0	0	0	1/1
Hyperoxaluria + Hyperuricosuria + Acid Urine + Low Urine Volume	1 (0.2%)	1 (100%)	0	0	0	1/1
Hypercalciuria + Hyperoxaluria + Hyperuricosuria + Acid Urine + Low Urine Volume	1 (0.2%)	1 (100%)	0	0	0	1/1

570

571

Appendix 3. Biochemistry diagnoses and recurrences for Bern database.

572

573