How Much Information is Lost When You Only Collect One 24-Hour Urine Sample during the Initial Metabolic Evaluation?

Abdulrahman F. Alruwaily, Casey A. Dauw, Maggie J. Bierlein, Phyllis Yan, John R. Asplin, Khurshid R. Ghani, J. Stuart Wolf, Jr.* and John M. Hollingsworth†

From the Divisions of Endourology and Health Services Research, Department of Urology, University of Michigan, Ann Arbor, Michigan (AFA, CAD, MJB, PY, KRG, JSW, JMH), Department of Surgery, College of Medicine, Al-Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia (AFA), and Litholink Corporation, Laboratory Corporation of America Holdings, Chicago, Illinois (JRA)

Purpose: During the initial metabolic evaluation the need for 1 vs 2, 24-hour urine collections is debated. While data suggest that mean urine chemistry measures are similar on consecutive samples, it remains unclear how much, if any, information is lost when only 1 sample is collected.

Materials and Methods: Using analytical files from Litholink Corporation® (1995 to 2013) we identified adults with kidney stones who underwent initial metabolic testing. Next we determined the subset of patients who collected 2, 24-hour urine samples with urine creatinine varying by 10% or less during a 7-day time window. We then examined the degree of variability in urine chemistry profiles. Specifically we calculated the mean absolute value of the difference between samples as well as the percent difference for individual urine parameters.

Results: We identified 70,192 patients meeting our eligibility criteria. While the overall means for individual urine parameters did not vary between samples, the percent difference between the samples varied widely. For example, nearly 1 in 3 patients had a 30% or greater difference in urine calcium and volume between 2 consecutive samples. We noted that inconsistencies between samples often involved multiple parameters. For instance, 29% and 25% of patients had a 20% difference in 2 and 3 or more parameters, respectively.

Conclusions: We observed substantial differences between consecutive 24-hour urine samples that could affect clinical decision making. In light of these findings clinicians must weigh the information lost from only 1 collection vs the burden to the patient of collecting 2.

Key Words: nephrolithiasis, urinalysis, quality of health care, secondary prevention

Accepted for publication April 8, 2016.

No direct or indirect commercial incentive associated with publishing this article.

The corresponding author certifies that, when applicable, a statement(s) has been included in the manuscript documenting institutional review board, ethics committee or ethical review board study approval; principles of Helsinki Declaration were followed in lieu of formal ethics committee approval; institutional animal care and use committee approval; all human subjects provided written informed consent with guarantees of confidentiality; IRB approved protocol number; animal approved project number.

Supported in part by the Agency for Healthcare Research and Quality, Grant 1K08HS020927-01A1 (JMH).

* Financial interest and/or other relationship with Urology Times®.

† Correspondence: 2800 Plymouth Rd., Building 16, 1st Floor, Room 112W, Ann Arbor, Michigan 48109 (telephone: 734-763-2797; FAX: 734-232-2400; e-mail: kinks@med.umich.edu).

Editor's Note: This article is the fourth of 5 published in this issue for which category 1 CME credits can be earned. Instructions for obtaining credits are given with the questions on pages 1326 and 1327.

With recurrence rates as high as 50%,[1,2] nephrolithiasis is considered a chronic condition for which secondary prevention efforts can have an important management role.[3,4] To guide these efforts under current practice guidelines it is recommended that clinicians perform metabolic testing in high risk patients and interested first-time stone formers.[5,6] One of the pillars of this testing is the 24-hour urine collection, whereby a patient's urine is assayed for various promoters and inhibitors of stone formation.

0022-5347/16/1964-1143/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2016.04.074
Vol. 196, 1143-1148, October 2016
Printed in U.S.A.

While there is consensus among experts about the value of 24-hour urine findings for guiding treatment decisions, the optimal number of collections that patients should complete during this initial evaluation remains the subject of debate. Although 2 collections are associated with a greater diagnostic yield,[7–10] some investigators have suggested that 1 may be sufficient, saving patients time and money.[11,12] However, when deciding how many collections to order, what a clinician needs to know is not whether the means of individual urine parameters from consecutive samples are similar,[11] but rather how much variability exists between samples for a given parameter.[10] This knowledge would help the clinician understand the potential information loss if only 1 sample was collected.

In this context we analyzed data from one of the largest national laboratories in the United States that provides services to patients with kidney stones. We determined the proportion of patients who performed 2 consecutive 24-hour urine collections as part of their initial evaluation. We then examined the degree of variation in urine chemistry measurements between the 2 samples. Findings from our study quantify for clinicians the amount of information that is lost when only 1 specimen is collected, which they must weigh against the burden to patients of 2 collections.

METHODS

Data Source and Study Population

For this study we used analytical files from the Litholink Corporation, which contain demographic data and 24-hour urine collection results from community dwelling individuals (1995 to 2013). From these files we identified adults 18 years or older who underwent metabolic testing for kidney stones. We determined which of these patients collected 1 vs 2 consecutive 24-hour urine samples (defined as those collected within 7 days of each other) as part of their initial evaluation. For patients who submitted 2 urines on multiple occasions we only included findings from the initial set in our study.

Statistical Analysis

For our initial analytic step we made bivariate comparisons between patients who collected 1 vs 2, 24-hour samples using t-tests and chi-square tests where appropriate. Specifically we compared patients with respect to age, gender, urban/rural status and region of residence (both based on billing ZIP Code™ data) as well as the specialty of the clinician who ordered the metabolic testing.

Among the patients who performed 2 consecutive 24-hour urine collections we isolated the subset with urine creatinine values varying by 10% or less between samples to ensure the consistency of the amount of urine collection between the 2 samples. We then examined the variability in urine volume and chemistries related to calcium stone formation (ie urine calcium, oxalate, citrate, uric acid and pH).

Using pairwise t-tests we compared the means of urine volumes and chemistries from the 2 samples. Our a priori hypothesis was that these means would be similar. We then calculated the mean absolute value of the differences in urine volume and chemistries between samples, as well as their percent difference. For each urine parameter the percent difference was calculated using the formula,

$$\%\,\text{Difference} = |\text{sample difference}|/(\text{sample 1} + \text{sample 2})/2.^{8}$$

By using the absolute difference between collections in the numerator, the calculation makes no assumptions about which collection is the correct one.

Since the urine pH is on a log scale, we only calculated the absolute difference between 2 samples, recognizing that a change in pH of 0.3 units represents a doubling (or halving) of the proton concentration. We also performed a series of sensitivity analyses to test the robustness of our findings. We assessed only urine collections performed on consecutive days (vs those with a gap between them). Then we looked at differences between collections based on the time of year during which the collections were performed (as defined by quarters). Finally, we assessed only urine chemistry measurements with clinically meaningful differences between consecutive collections. For instance, with urine calcium we included only those consecutive measurements that had at least one with a value of 100 mg or greater (cutoffs for citrate 150 mg or greater, oxalate 25 mg or greater, uric acid 0.2 gm or greater and volume 2 L or less).

We used multivariable regression to evaluate whether any patient characteristics were associated with variability between collections. We did this by fitting separate logistic regression models for each urine parameter, where our outcome was a 20% or greater difference in the parameter between 2 samples, and our predictors were patient age, gender, region of residence and treating provider specialty.

All analyses were done using SAS® software, version 9.4. We performed 2-sided significance testing and set a type I error rate at 0.05. The University of Michigan

Table 1. *Comparing patients with 1 vs 2 urine collections on initial evaluation*

	1 Collection	2 Collections
No. pts (%)	332,941 (82.6)	70,192 (17.4)
% Age:		
18–34	15.67	13.74
35–44	16.74	17.95
45–54	22.66	23.05
55–64	24.42	26.87
65–100	20.51	19.04
% Male	54.83	55.23
% Urban	85.63	82.61
% Region:		
Midwest	20.05	31.95
Northeast	30.40	29.58
South	36.02	29.50
West	13.53	8.68
% Specialty of ordering clinician:		
Urology	83.20	16.80
Nephrology	78.21	21.79
Endocrinology	83.70	16.30
Primary care	81.79	18.21
Other	86.67	13.33