



Optimization of urinary dipstick pH: Are multiple dipstick pH readings reliably comparable to commercial 24-hour urinary pH?

Joel E. Abbott¹, Daniel L. Miller¹, William Shi¹, David Wenzler¹, Fuad F. Elkhoury², Nishant D. Patel¹, Roger L. Sur¹

¹Division of Urology, Department of Surgery, University of California San Diego, San Diego, CA, ²Department of Urology, University of California Los Angeles, Los Angeles, CA, USA

Purpose: Accurate measurement of pH is necessary to guide medical management of nephrolithiasis. Urinary dipsticks offer a convenient method to measure pH, but prior studies have only assessed the accuracy of a single, spot dipstick. Given the known diurnal variation in pH, a single dipstick pH is unlikely to reflect the average daily urinary pH. Our goal was to determine whether multiple dipstick pH readings would be reliably comparable to pH from a 24-hour urine analysis.

Materials and Methods: Kidney stone patients undergoing a 24-hour urine collection were enrolled and took images of dipsticks from their first 3 voids concurrently with the 24-hour collection. Images were sent to and read by a study investigator. The individual and mean pH from the dipsticks were compared to the 24-hour urine pH and considered to be accurate if the dipstick readings were within 0.5 of the 24-hour urine pH. The Bland-Altman test of agreement was used to further compare dipstick pH relative to 24-hour urine pH.

Results: Fifty-nine percent of patients had mean urinary pH values within 0.5 pH units of their 24-hour urine pH. Bland-Altman analysis showed a mean difference between dipstick pH and 24-hour urine pH of -0.22, with an upper limit of agreement of 1.02 (95% confidence interval [CI], 0.45–1.59) and a lower limit of agreement of -1.47 (95% CI, -2.04 to -0.90).

Conclusions: We concluded that urinary dipstick based pH measurement lacks the precision required to guide medical management of nephrolithiasis and physicians should use 24-hour urine analysis to base their metabolic therapy.

Keywords: Kidney calculi; Nephrolithiasis; Prevention and control; Urine specimen collection

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 unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Nephrolithiasis recurs frequently, with contemporary recurrence rates for first-time stone formers ranging from 11%–39% [1,2]. Control of urinary pH can reduce or even prevent recurrence in many patients. Hence, accurate mea-

surement of urinary pH is the first step in the prevention of future episodes of nephrolithiasis.

American Urological Association guidelines recommend a 24-hour urine collection for recurrent or high-risk stone formers and urinary alkalization for those with uric acid and cystine stones [3]. However, monitoring response

Received: 23 January, 2017 • **Accepted:** 6 April, 2017

Corresponding Author: Joel E. Abbott

Division of Urology, Department of Surgery, University of California San Diego, 200 W. Arbor Dr. unit 8897, San Diego, CA 92103, USA
TEL: +1-702-715-4155, FAX: +1-410-787-1911, E-mail: jabbott@cua.md

to medical therapy with repeated 24-hour urine collections is cumbersome. Urinary dipstick monitoring offers a cheap and convenient method of measuring urinary pH, but single dipstick pH measurements have been shown to produce an unacceptable rate of clinically significant deviation [4]. Furthermore, due to the known diurnal variation in urinary pH, prior studies may not be useful given they only examined single, spot dipstick pH [5-7]. Multiple urinary dipstick pH measurements over the course of a day may more accurately reflect the 24-hour urine pH at a fraction of the cost of a formal 24-hour urine collection.

The objective of this study was therefore to determine whether the average of multiple urinary dipstick pH measurements taken the same day has sufficient accuracy and precision to guide kidney stone management in patients who would otherwise require 24-hour urine collections for pH measurement.

MATERIALS AND METHODS

The University of California San Diego Human Research Protection Program Institutional Review Board approval was obtained prior to initiation of the study (approval number: 130606). Informed consent was obtained from all individual participants included in the study. We prospectively enrolled stone-formers who already had been instructed to perform a 24-hour urine collection. Patients were excluded

for the inability to collect voided urine, a history of struvite calculi, ureteral obstruction, neurogenic bladder, the presence of a foreign body in the urinary tract (including Foley catheter, nephrostomy tube, ureteral stent), and for bacteruria on urine culture (>100,000 colony-forming units).

Study participants were provided with 3 sterile specimen cups, and an insulated container with 3 Chemstrip 10SG Dipsticks (Roche Pharmaceutical, Basel, Switzerland). Patients were instructed to perform urinary dipstick testing of their first 3 voids of the day concurrently with the 24-hour urine collection. A digitally time and date-stamped image of each dipstick was taken by the subject within 1 minute of urine exposure. Dipstick images were emailed to a trained study investigator, (WS), who determined the pH of each dipstick. The pH of the 24-hour urine collection was determined by a commercially available company (Litholink Corp, Chicago, IL, USA) using a pH meter.

The individual and mean dipstick pH values were defined as accurate if they were within 0.5 pH units with the 24-hour urine pH. The Bland-Altman test of agreement was used to compare the mean of the 3 dipstick readings with the 24-hour urine pH, and 95% confidence intervals (CIs) were determined for the upper and lower limits of agreement. The Bland-Altman plot assesses the agreement between two methods of clinical measurement by plotting the difference of dipstick pH and 24-hour pH relative to the mean of both measured pHs. The bias of the newer method

Table 1. Dipstick and 24-hour urine pH measurements pH is displayed in pH units by measured by colorimetric dipstick for dipstick pH and electrochemical pH meter for 24-hour pH

Patient	1st void dipstick pH (time)	2nd void dipstick pH (time)	3rd void dipstick pH (time)	Mean dipstick pH	24-Hour urine pH	Deviation from 24-hour pH
1	8 (8:30 AM)	6 (2:00 PM)	6 (NR)	6.7	6.965	-0.265
2	6 (9:00 AM)	6 (11:00 AM)	6 (3:00 PM)	6.0	6.700	-0.700
3	7 (9:00 AM)	8 (1:00 PM)	9 (2:30 PM)	8.0	6.654	1.346
4	5 (10:00 AM)	5 (1:00 PM)	5 (7:00 PM)	5.0	6.446	-1.446
5	8 (8:00 AM)	5 (10:15 AM)	7 (11:15 AM)	6.7	6.892	-0.192
6	7 (1:30 PM)	7 (8:30 PM)	5 (11:10 PM)	6.3	6.349	-0.049
7	5 (9:00 AM)	7 (12:30 PM)	5 (5:30 PM)	5.7	5.970	-0.270
8	8 (12:50 PM)	8 (4:35 PM)	8 (8:50 PM)	8.0	7.492	0.508
9	6 (6:18 AM)	5 (12:52 PM)	6 (6:48 PM)	5.7	6.177	-0.477
10	7 (11:30 AM)	7 (2:30 PM)	8 (5:30 PM)	8.0	7.405	-0.105
11	6 (4:00 PM)	6 (4:58 PM)	7 (6:13 PM)	6.3	6.562	-0.262
12	5 (1:25 PM)	5 (5:00 PM)	5 (9:00 PM)	5.0	5.761	-0.761
13	6 (2:40 PM)	5 (8:00 PM)	6 (9:20 PM)	5.7	5.524	0.176
14	7 (1:30 PM)	5 (6:00 PM)	5 (11:30 PM)	5.7	5.910	-0.210
15	5 (11:42 AM)	5 (9:05 PM)	5 (6:12 AM)	5.0	5.865	-0.865
16	6 (12:30 PM)	5 (4:00 PM)	6 (8:30 PM)	5.7	5.213	0.087
17	5 (2:13 PM)	8 (11:27 PM)	5 (4:55 AM)	6.0	6.732	-0.732

Time is displayed in a 12-hour clock system.

Table 2. Patient demographics, medications, comorbidities

Patient	Age (y)	Sex	Thiazide	Citrate medication	Phenazopyridine	CKD \geq stage 3	Diabetes mellitus	Stone composition	BMI (kg/m ²)
1	78	Male	No	No	No	No	No	Not recorded	23.1
2	57	Female	No	No	No	Yes	No	Calcium phosphate	16.4
3	61	Female	No	No	No	No	No	Calcium phosphate	21.4
4	41	Male	No	No	No	No	No	Not recorded	29.8
5	62	Female	No	No	Yes	No	No	Calcium oxalate	18.9
6	44	Male	No	Yes	No	No	No	Not recorded	34.4
7	51	Male	No	No	No	Yes	No	Not recorded	28.4
8	46	Female	Yes	Yes	No	No	No	Not recorded	24.2
9	39	Male	No	No	No	No	No	Not recorded	30.5
10	44	Male	Yes	Yes	No	No	Yes	Mixed calcium	36.5
11	59	Male	No	No	No	No	No	Not recorded	23.0
12	65	Male	No	No	No	No	No	Calcium oxalate	27.6
13	65	Female	No	No	No	No	No	Calcium oxalate	23.4
14	61	Male	No	No	No	No	No	Mixed calcium	28.0
15	69	Male	Yes	No	No	No	No	Not recorded	25.1
16	69	Male	No	No	No	No	No	Not recorded	24.2
17	51	Male	No	No	No	No	No	Calcium oxalate	24.8

CKD, chronic kidney disease; BMI, body mass index.

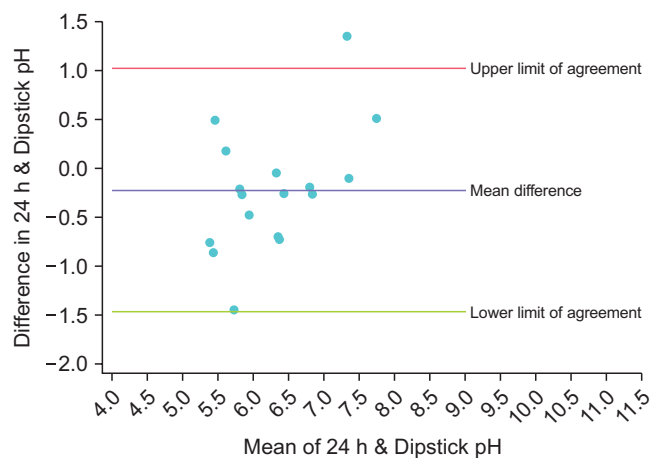


Fig. 1. Bland-Altman plot comparing the mean of 3 dipstick pH readings with the concurrent 24-hour urine pH.

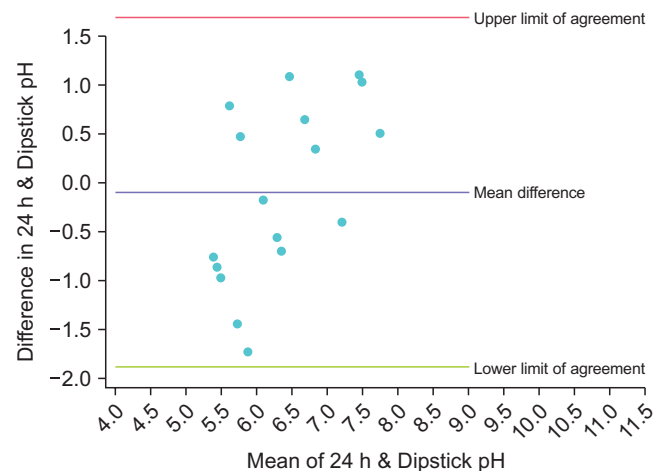


Fig. 2. Bland-Altman plot comparing the dipstick pH from the first void of the day with the concurrent 24-hour urine pH.

relative to the standard (dipstick and 24-hour pH, respectively) is given by the mean difference between the 2 methods. The precision is demonstrated by the width of the limits of agreement. Statistical analysis was performed using Stata 13.1 (StataCorp LP, College Station, TX, USA).

RESULTS

We consented 43 subjects; 17 completed both their spot urinary dipsticks and concurrent 24-hour urine collection. Patient characteristics are displayed in Table 1 and pH data is displayed in Table 2. Fifty-nine percent (59%, 10 of 17) had

mean urinary pH values within 0.5 pH units of their 24-hour urine pH. Mean dipstick accuracy was also evaluated for subgroups: pH<6.0, 67% (4 of 6); 6.0≤pH<7.0, 56% (5 of 9); and pH≥7.0, 50% (1 of 2).

The Bland-Altman plot for the mean of the 3 dipstick pH readings is displayed in Fig. 1. The mean difference between the dipstick pH and the 24-hour pH is -0.22, representing a small negative bias for the dipsticks compared to the 24-hour urine. The upper limit of agreement was 1.02 (95% CI, 0.45–1.59) and the lower limit of agreement was -1.47 (95% CI -2.04 to -0.90); assuming our study participants are a representative sample of the population of stone formers, 95%

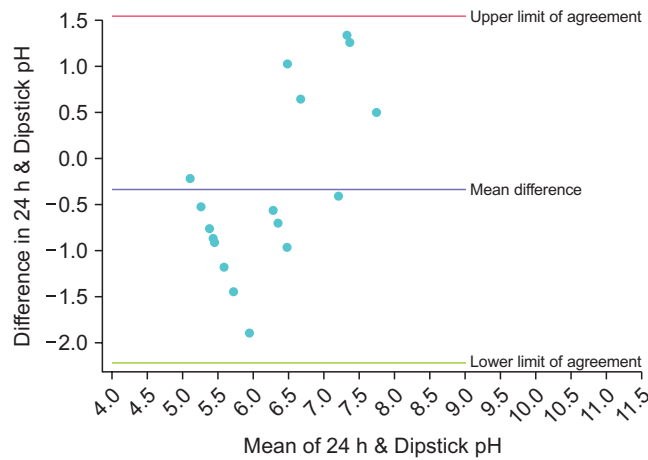


Fig. 3. Bland-Altman plot comparing the dipstick pH from the second void of the day with the concurrent 24-hour urine pH.

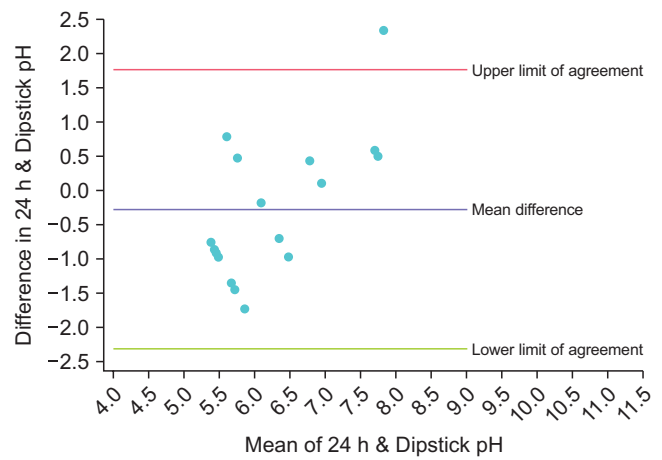


Fig. 4. Bland-Altman plot comparing the dipstick pH from the third void of the day with the concurrent 24-hour urine pH.

of patients would have a mean dipstick pH within ~1 pH unit above and ~1.5 pH units below the 24-hour urine pH. To illustrate the increasing precision of combining multiple dipstick readings, Figs. 2–4 show the Bland-Altman plots for the pH of the 1st, 2nd, and 3rd voids, respectively. There is minimal change in the mean difference but the limits of agreement widen considerably when comparing any of the single voided pH readings to the 24-hour urine pH.

DISCUSSION

Urinary pH control is important for preventing stone recurrence in certain populations of stone formers and dipstick pH monitoring offers an inexpensive, convenient mechanism for patients to monitor their pH at home. In our study we examined whether multiple urinary dipstick readings can provide clinically equivalent results to 24-hour urine pH. We found significant variation for the dipstick pH; only 59% of the mean urinary dipstick readings were within 0.5 units of the 24-hour urine pH. The Bland-Altman analysis for the mean urinary dipstick readings showed a small bias (-0.22 pH units) but wide limits of agreement, with no apparent difference in bias nor precision across the range of physiologic pH. This indicates that even after dipstick pH has been optimized by taking the average of multiple readings, dipstick pH is insufficiently precise to guide urinary pH monitoring.

We did take measures to optimize the accuracy and reliability of the urinary dipstick readings. We used Chemstrip urinary dipsticks, given they have been shown to be more accurate than other urinary dipsticks and also had a trained study investigator read the dipsticks [8]. Given the diurnal variation of urinary pH, we used the average of multiple

voided urine specimens in an effort to reflect the true average daily pH [9].

Prior studies examining the accuracy and reliability of urinary pH monitoring have focused primarily on spot urinary analysis. Kwong et al. [4] tested 390 urine samples with both urinary dipstick and electrochemical pH meter and concluded that dipsticks had an unacceptable frequency of clinically relevant error. Similarly, Ilyas et al. [10] compared a clinic-based portable pH meter, urinary dipsticks, and litmus paper to laboratory-based pH meter and found that dipsticks lacked the precision needed to base a clinical decision. However, to our knowledge, our study is the first to look beyond single dipstick pH readings in an attempt to account for diurnal variation in urinary pH.

The shortcomings of urinary dipstick pH monitoring are likely due to the imprecision of the colorimetric dipstick pH scale. Most dipsticks measure pH based on qualitative color changes in increments of 1 or 0.5 pH units whereas an electrochemical pH meter, whether assessing spot or 24-hour urine pH, measures pH quantitatively down to 0.001 of a pH unit. Our hypothesis had been that we could overcome the imprecision of an individual dipstick pH reading by averaging multiple readings over the course of a day. We did observe an increase in precision when comparing the average to any individual pH measurement, represented by narrower limits of agreement (Figs. 1–4). However, the precision of our averaged pH dipstick reading is essentially the same as that found by Ilyas et al. [10] when they compared 200 individual dipstick pH readings to laboratory pH meter. This indicates that we likely have reached the limits of precision achievable by dipstick pH monitoring, but dipsticks are not sufficiently precise to guide clinical management.

Limitations to our study include a relatively small over-

all number of patients. However, as previously noted, we found similar deviation from our standard (24-hour urine pH) and similar precision to studies that evaluated hundreds of urine samples, indicating that more patients would be unlikely to change our conclusions. We had study participants take 3 dipstick readings; theoretically a higher number of dipstick readings would more closely approximate 24-hour urine pH. However, this may have negatively impacted the compliance of patients and likely would not have increased our precision. We used Chemstrip 10SG Dipsticks (Roche Pharmaceutical) which measure pH in 1-unit increments of pH; a brand that measures in 0.5-unit increments could offer higher precision. However, when evaluated, Chemstrip brand actually was shown to be more accurate than a brand with 0.5-unit increments (Multistix, Siemens Medical Solutions USA Inc, Malvern, PA, USA) [8]. Urinary pH is known to be affected by various factors such as diet and hydration, however these patient-specific characteristics were not collected in our data set. It may be of importance to identify if these factors affect variation between dipstick testing and 24-hour urine collection.

CONCLUSIONS

In conclusion, we demonstrated that urinary pH monitoring by both individual dipsticks and the average of three dipsticks lacks the precision needed to guide medical management of nephrolithiasis. Our data suggests that clinicians should use the pH of a 24-hour urine collection for the evaluation of nephrolithiasis patients over urinary dipsticks.

CONFLICTS OF INTEREST

The authors have nothing to disclose.

ACKNOWLEDGMENTS

We thank Angela Wang for assistance with data collec-

tion and data organization.

REFERENCES

1. Rule AD, Lieske JC, Li X, Melton LJ 3rd, Krambeck AE, Bergstralh EJ. The ROKS nomogram for predicting a second symptomatic stone episode. *J Am Soc Nephrol* 2014;25:2878-86.
2. Kang HW, Seo SP, Kwon WA, Woo SH, Kim WT, Kim YJ, et al. Distinct metabolic characteristics and risk of stone recurrence in patients with multiple stones at the first-time presentation. *Urology* 2014;84:274-8.
3. Pearle MS, Goldfarb DS, Assimos DG, Curhan G, Denucciocca CJ, Matlaga BR, et al. Medical management of kidney stones: AUA guideline. *J Urol* 2014;192:316-24.
4. Kwong T, Robinson C, Spencer D, Wiseman OJ, Karet Frankl FE. Accuracy of urine pH testing in a regional metabolic renal clinic: is the dipstick accurate enough? *Urolithiasis* 2013;41:129-32.
5. Strohmaier WL, Hoelz KJ, Bichler KH. Spot urine samples for the metabolic evaluation of urolithiasis patients. *Eur Urol* 1997;32:294-300.
6. Murayama T, Sakai N, Yamada T, Takano T. Role of the diurnal variation of urinary pH and urinary calcium in urolithiasis: a study in outpatients. *Int J Urol* 2001;8:525-31.
7. Murayama T, Taguchi H. The role of the diurnal variation of urinary pH in determining stone compositions. *J Urol* 1993;150(5 Pt 1):1437-9.
8. Desai RA, Assimos DG. Accuracy of urinary dipstick testing for pH manipulation therapy. *J Endourol* 2008;22:1367-70.
9. Cameron M, Maalouf NM, Poindexter J, Adams-Huet B, Sakhaee K, Moe OW. The diurnal variation in urine acidification differs between normal individuals and uric acid stone formers. *Kidney Int* 2012;81:1123-30.
10. Ilyas R, Chow K, Young JG. What is the best method to evaluate urine pH? A trial of three urinary pH measurement methods in a stone clinic. *J Endourol* 2015;29:70-4.