

Reliability of 3 Urine Specific Gravity Meters for Measuring Brix and Urine Solutions at Different Temperatures

Floris Wardenaar, PhD; Carmen P. Ortega-Santos, MS; Kaila Vento, MS; Stephanie Olzinski, MS; Jason Olig, MS; Stavros Kavouras, PhD; Carol Johnston, PhD

College of Health Solutions, Arizona State University, Phoenix

Context: Urine specific gravity (USG) should be measured at room temperature (20°C), but the temperature of the sample is not always considered.

Objective: To evaluate the effect of sample temperature on the measurement accuracy of a digital refractometer (DIG), manual optical refractometer (MAN), and hydrometer (HYD).

Design: Descriptive laboratory study.

Main Outcome Measure(s): Urine specific gravity.

Results: Experiment 1 (24 Brix (°Bx) samples) showed that measurements via the DIG and MAN did not differ from the reference, but HYD provided lower or inconsistent values compared with °Bx and was highly correlated with °Bx solutions ($r, > = 0.89$). The overall diagnostic ability of elevated USG cut-off values ($\geq 1.020, \geq 1.025, \geq 1.030$) was high for all tools (area under the curve > 0.92). Misclassification of samples increased from 0 to 2 at 1.020 to 1 to 3 samples at cutoffs of 1.025 and

1.030 USG. Bland-Altman analysis showed that the DIG 5°C underreported slightly without reporting bias ($r = -0.344, P = .13$); all other plots for the DIG, MAN, and HYD showed considerably larger underreporting at higher concentrations ($r = -0.21$ to -0.97 with $P > .02$) at all temperatures. The outcomes of experiment 2 (33 fresh urine samples) using DIG 20°C as the standard demonstrated only negligible differences between the DIG and MAN at all temperatures but larger differences using the HYD.

Conclusions: All tools showed reporting bias compared with the °Bx solutions, which can affect the classification of low and high urine concentration at higher USG cutoff values, especially with a sample temperature of 37°C.

Key Words: hydration status, validation, accuracy

Key Points

- Urine specific gravity is preferably measured at 20°C, but temperature of the sample is often not standardized in practical settings on or around the field.
- Refractometers showed reporting bias compared with the Brix solutions used as a reference standard for urine. In particular, values near a clinically relevant dehydration level (≥ 1.030) tended to be lower and were amplified by a 37°C sample temperature. The underreporting was in the range of 0.001 to 0.003 depending on the tool, temperature, and concentration.
- The tested digital and manual optical refractometer measures tended to be similar when Brix solutions and urine samples at 5°C and 20°C were analyzed. The hydrometer displayed a high degree of misreporting at all temperatures and should not be used to measure urine specific gravity.

Athletes should monitor their hydration status to prevent dehydration, heat-related illnesses, and fatigue.¹ However, no unambiguous standard for measuring hydration status in a field situation exists.² Therefore, combining at least 2 independent measurements to assess hydration status has been recommended.² The National Athletic Trainers' Association¹ advocated educating athletes to monitor acute differences in hydration status from body mass changes and additional assessments of chronic hydration status using urine concentration.

Although urine osmolality is seen as a more precise measurement, it may lack feasibility in a field setting.³ The simple low-cost measurement of urine specific gravity (USG)^{4,5} has consistently shown a good correlation with urine osmolality.⁶ In practice, USG is often measured as a

marker of hydration status and used for athlete education.⁷ As the results affect the hydration advice athletes receive based on their classification (low or high urine concentration), an accurate assessment of hydration status is important.

Multiple factors may influence the quality and outcome of USG measurements, as refractometers, hydrometers (HYDs), and dipsticks may under- or overreport values.^{8–10} Most of the comparisons among refractometers have been reported by researchers^{11–15} using canine and feline urine and showed no clear consensus on whether optical or manual versus digital refractometers (DIGs) could be used interchangeably. In humans, investigators have reported that manual refractometers and DIGs displayed good correlation^{8,16} but that manual refractometers and DIGs

may provide different values, especially when USG is high.¹⁶

Handling of the urine sample may influence the measurement. In clinical settings, the samples may be cooled and stored for multiple hours before they are measured,¹⁷ which is in contrast to athletic populations, whose samples are often measured directly after collection.^{18–21} The temperature differences of the samples (5°C to approximately 37°C) may affect measurement precision, as liquid density is influenced by temperature. However, manufacturers claim their refractometers automatically control for sample temperatures.¹⁶ The notion is that sample temperature has little or no bearing on the accuracy of the measurement reading, as the small amount of the sample needed immediately equilibrates to the temperature of the instrument. Yet, to date, no authors have provided detailed results that indicated sample temperature did not affect USG.

Our aim was to evaluate the reliable temperature range for a DIG, manual optical refractometer (MAN), and HYD by using reference Brix solutions and a representative sample range of urine samples at temperatures ranging from 5°C to 37°C against USG cutoff values that help to define stages of euhydration and underhydration in clinical sport nutrition practice.

METHODS

Design

We developed a quantitative study to evaluate the accuracy of different tools for measuring USG: DIG (model PEN-S.G.; Atago Co, Ltd), MAN (model RHC-200ATC; WMicro), and HYD (urinometer; C&A Scientific). All devices were operated according to the manufacturer's instructions, and we conducted the study in 2 parts. For experiment 1, we prepared aqueous sugar solutions across a range of degrees Brix (°Bx; 1°Brix = 1 g of sucrose in 100 g of solution) as an independent standard for specific gravity. For experiment 2, we used urine samples that were collected during a previous study that was approved by the Arizona State University Institutional Review Board (No. STUDY00007260).

Measurements of Specific Gravity

For experiment 1, in which we validated USG meters using Brix solutions, we prepared 24 validation solutions by mixing 500 mL of distilled water with laboratory-grade sucrose (crystal sucrose; Sargent Welch) and a precision scale with a resolution of 0.0001 (AL204; Mettler Toledo). We considered these solutions represented in °Bx as valid independent replacements for urine samples to objectively assess the USG, measured with a refractometer, ranging from 1.000 to 1.040 (Table 1).

For experiment 2, in which we compared the USG measurements of urine samples, we used 33 fresh urine samples (collected as part of another study) with USGs that ranged from 1.002 to 1.033 in a normal distribution (mean = 1.018 ± 0.009 based on PEN measurements). The fresh urine samples were measured at 20°C on the day of collection and then refrigerated at 5°C. Within 5 days of collection, the samples were measured again, first at 5°C and then after warming to 37°C on the same day.

Table 1. Sucrose Solutions Reflecting Urine Specific Gravity and Degrees of Brix

Urine Specific Gravity	Sucrose, g/L	Brix
1.000	0	0.05
1.010	23.0	2.30
1.015	35.5	3.55
1.016	38.0	3.80
1.017	40.5	4.05
1.018	43.0	4.30
1.019	45.5	4.55
1.020	48.0	4.80
1.021	51.0	5.10
1.022	53.5	5.35
1.023	56.0	5.60
1.024	59.0	5.90
1.025	61.5	6.15
1.026	64.0	6.40
1.027	67.0	6.70
1.028	70.0	7.00
1.029	72.5	7.25
1.030	75.0	7.50
1.031	78.0	7.80
1.032	81.0	8.10
1.033	84.0	8.40
1.034	87.0	8.70
1.035	89.5	8.95
1.040	104.0	10.40

For both experiments, the samples were prepared separately for measurement with the refractometers (2 × 15 mL in 45-mL transparent plastic freestanding centrifuge tubes; Evergreen Scientifics) and HYD (1 × 45 mL in glass Erlenmeyer flasks; Thermo Fisher Scientific). Measurements for both experiments were obtained in the following order: samples were cooled in the refrigerator (5°C) and then the closed samples were heated to 20°C and to 37°C. During experiment 1, we used an oven (model Heratherm OGS100; Thermo Fisher Scientific) and a Traceable Sentry Thermometer (Thermo Fisher Scientific) with a resolution of 1°C and accuracy of ±1°C to measure sample temperature. During experiment 2, we used a warm-water bath at 20 ± 1°C and 37 ± 1°C and a Traceable Long-Stem Digital ULTRA thermometer with a resolution of 0.1°C and accuracy of ±0.2°C. Each urine sample was measured at 3 temperatures on the same day within an 8-hour period. Before use, we calibrated both refractometers and the HYD using distilled water at 20°C. We either measured all samples in duplicate and used the average for analysis when the measurements were not more than 0.005 apart or we measured the samples in triplicate, used the median for analysis when the difference between the first 2 measurements >0.005, and used the average of the 2 measurements for analysis. Results were rounded to the thousandths.

Data Analysis

We performed separate analyses between all tools for the 2 experiments. As the HYD data were not distributed normally, we calculated the mean differences for all outcomes at different temperatures using Mann-Whitney *U* tests. Spearman correlations (including 95% CIs based on the Fisher *Z* transformation) were conducted to determine the relationships between reference measurements (°Bx) or the selected standard (DIG). The Bland-Altman analysis

Table 2. Single Measurements for Each Refractometer at Different Temperatures and Misclassification Based on Urine Specific Gravity Cutoff Values^a

Sucrose, g/500 mL	Urine Specific Gravity	Digital Refractometer			Manual Optical Refractometer			Hydrometer		
		5°C	20°C	37°C	5°C	20°C	37°C	5°C	20°C	37°C
0.5000	1.000	1.000	1.000	1.000	1.001	1.000	1.000	1.008	1.003	
11.5066	1.010	1.009	1.010	1.009	1.011	1.010	1.010	1.015	1.012	1.008
17.7518	1.015	1.014	1.015	1.013	1.016	1.015	1.015	1.021	1.015	1.011
19.0004	1.016	1.015	1.016	1.014	1.017	1.016	1.015	1.021	1.014	1.012
20.2514	1.017	1.015	1.016	1.015	1.017	1.017	1.016	1.023	1.015	1.012
21.5016	1.018	1.016	1.017	1.016	1.018	1.017	1.017	1.023	1.016	1.013
22.7515	1.019	1.018	1.018	1.017	1.019	1.018	1.018	1.024	1.016	1.013
24.0117	1.020	1.018	1.019	1.018	1.020	1.019	1.019	1.025	1.016	1.014
25.5000	1.021	1.020	1.020	1.018	1.020	1.020	1.020	1.017	1.016	1.015
26.7505	1.022	1.021	1.021	1.020	1.021	1.021	1.020	1.022	1.016	1.016
28.0048	1.023	1.021	1.022	1.021	1.022	1.022	1.021	1.022	1.017	1.016
29.5035	1.024	1.023	1.023	1.021	1.023	1.023	1.023	1.023	1.018	1.017
30.755	1.025	1.024	1.023	1.022	1.024	1.024	1.024	1.027	1.019	1.017
32.024	1.026	1.024	1.024	1.024	1.025	1.024	1.024	1.027	1.020	1.020
33.5044	1.027	1.026	1.025	1.024	1.026	1.025	1.025	1.026	1.021	1.019
35.0045	1.028	1.027	1.026	1.025	1.027	1.026	1.026	1.028	1.021	1.020
36.2508	1.029	1.027	1.027	1.026	1.028	1.027	1.027	1.028	1.025	1.021
37.5057	1.030	1.029	1.028	1.026	1.029	1.028	1.028	1.029	1.024	1.022
39.0138	1.031	1.029	1.029	1.028	1.030	1.029	1.029	1.035	1.025	1.022
40.5023	1.032	1.030	1.030	1.029	1.031	1.030	1.030	1.031	1.027	1.023
42.0000	1.033	1.031	1.031	1.030	1.032	1.031	1.031	1.031	1.027	1.024
43.5025	1.034	1.032	1.032	1.031	1.033	1.031	1.031	1.032	1.029	1.024
44.7513	1.035	1.033	1.033	1.032	1.034	1.032	1.033	1.032	1.030	1.025
52.0084	1.040	1.037	1.037	1.036	1.037	1.035	1.036	1.037	1.033	1.030

Abbreviation: USG, urine specific gravity.

^a Cutoff values are represented by horizontal lines at 1.020, 1.025, and 1.030. Colors indicate the measurements in comparison to the colors represented in the USG reference column on the left side of the table for the digital refractometer, manual optical refractometer, and hydrometer.

was used to determine mean estimation bias, direction, and 95% limits of agreement among refractometers with HYD as the standard. A level of $P \leq .05$ indicated significance for both analyses. Additionally, we calculated receiver operating characteristics to assess the diagnostic capability of the DIG, MAN, and HYD for identifying high urine concentration (USG ≥ 1.020) based on the area under the curve (AUC) and sensitivity and specificity. An AUC of ≥ 0.90 is considered *excellent*; 0.80 to 0.89, *good*; and 0.70 to 0.79, *fair* when sensitivity and specificity are preferably >0.80 .²² Finally, we evaluated the diagnostic validities of the equipment by comparing the refractometers with the accepted USG reference value of ≤ 1.020 as the upper limit classification of euhydration, and accuracy was depicted in a contingency table for each measurement temperature against the set reference.

We performed an a posteriori sample-size calculation using G*Power (version 3.0.10; Heinrich-Heine-Universität Düsseldorf) based on the mean difference of -0.0075 from the standard ($^{\circ}\text{Bx}$ reference solutions) and HYD 37°C for experiment 1 and the comparison of the mean difference ($+0.0049$) of the DIG 20°C and HYD 37°C for experiment 2.²³ To determine a sufficient sample size for experiment 1, we used $\alpha = .05$, power = 0.80, a calculated strong effect size ($d = 1.09$), and 2 tails that resulted in a minimum sample size of 9 for experiment 1. Repeating this calculation for experiment 2, we used $\alpha = .05$, power = 0.80, and a calculated medium effect size ($d = 0.71$) that resulted in a required sample size of 18 measurements. The

sample size of 33 for both experiments amply exceeded the desired power of 80%.

RESULTS

Experiment 1: Reporting the Accuracy of the USG Meters Against $^{\circ}\text{Bx}$ Reference Solutions

We compared the outcomes of the DIG, MAN, and HYD with the $^{\circ}\text{Bx}$ reference solutions. The measurements per reference solution are shown in Table 2. The USG cutoffs (1.020, 1.025, and 1.030) represent commonly used cutoff values in the literature^{5,21} for hydration status (euhydration, dehydration, and severe dehydration, respectively). At a higher cutoff value (ie, 1.025 or 1.030), a substantial number of measurements were classified below the cutoff for all tools. As indicated by color coding in Table 2, the misclassification of samples increased from 0 to 2 at 1.020 to 1 to 3 samples classified lower than the cutoff at 1.025 and 1.030 USG. All other results are presented in Table 3. The measured median (interquartile range) USG value for the reference $^{\circ}\text{Bx}$ solution was 1.025 (1.020–1.030). No median differences were seen between the $^{\circ}\text{Bx}$ solution and the DIG and MAN at any temperature or for the HYD 5°C ($P > .05$), reflecting similar reporting for averages. Only HYD 20°C (1.019 [1.022–1.030], $P = .040$, and HYD 37°C (1.017 [1.013–1.022], $P = .005$) demonstrated clearly lower values than those of the reference.

Ranking for all measurements of the DIG, MAN, and HYD with the $^{\circ}\text{Bx}$ solution can be considered strong for all temperatures, with $r = 0.99$ to 1.00, except for the HYD

Table 3. Standard ²BRIX-Based Urine Specific Gravity Values Compared with Outcomes for Different Tools and Temperatures

Measure	Urine Specific Gravity											
	Digital Refractometer				Manual Optical Refractometer				Hydrometer			
	5°C	20°C	37°C	37°C	5°C	20°C	37°C	37°C	5°C	20°C	37°C	37°C
Outcome												
Median	1.024	1.023	1.022	1.024	1.024	1.024	1.024	1.024	1.026	1.019	1.017	1.017
Interquartile range	1.018–1.028	1.018–1.028	1.017–1.027	1.019–1.029	1.019–1.029	1.019–1.029	1.019–1.029	1.019–1.029	1.022–1.030	1.016–1.025	1.013–1.022	1.013–1.022
Difference	-0.0016	-0.0016	-0.0027	-0.0010	-0.0010	-0.0010	-0.0020	-0.0020	-0.0005	-0.0055 ^a	-0.0075 ^a	-0.0075 ^a
Correlation												
<i>r</i>	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.89	0.99	0.99	0.99
95% CI	0.99–1.00	0.99–1.00	0.99–1.00	0.99–1.00	0.99–1.00	0.99–1.00	0.99–1.00	0.99–1.00	0.76–0.96	0.96–0.99	0.98–1.00	0.98–1.00
Bland-Altman analysis												
Bias	-0.002	-0.001	-0.003	-0.001	-0.001	-0.002	-0.002	-0.002	0.001	-0.004	-0.007	-0.007
95% LoA	-0.003; 0.000	-0.003; 0.000	-0.004; -0.001	-0.002; 0.001	-0.002; 0.001	-0.004; 0.001	-0.004; 0.000	-0.004; 0.000	-0.006; 0.008	-0.010; 0.001	-0.011; -0.003	-0.011; -0.003
N < or > 95% LoA	0	1 < LoA	0	1 < LoA	1 < LoA	1 < LoA	1 < LoA	1 < LoA	0	2 > LoA	0	0
R; <i>P</i> values	-0.34, .127	-0.97, <.001	-0.92, <.001	-0.37, <.001	-0.27, <.01	-0.27, <.01	-0.52, <.001	-0.52, <.001	-0.32, .004	-0.21, .012	0.45, <.001	0.45, <.001
Receiver operating curve												
Area under the curve	1.000	1.000	0.996	0.996	1.000	1.000	1.000	1.000	0.867	0.977	1.000	1.000
Sensitivity	1.000	1.000	0.938	0.938	1.000	1.000	1.000	1.000	0.750	0.875	1.000	1.000
Specificity	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Abbreviation: LoA, limits of agreement.

^a Indicates a difference from standard BRIX for the hydrometer at 20°C ($P < .05$) and 37°C ($P < .005$). $P < .001$ for all receiver operating characteristic curve analyses.

5°C, for which we calculated a slightly lower correlation of $r = 0.89$.

Bland-Altman analysis revealed that the HYD data consistently underreported hydration status at all temperatures for the DIG and MAN, ranging from -0.001 to -0.003 . The DIG 5°C underreported slightly without reporting bias for low and high measurements ($r = -0.344$, $P = .13$); all other plots for DIG, MAN, and HYD exhibited considerably larger underreporting biases at higher concentrations ($P \leq .012$) at all temperatures except for the HYD 5°C, which displayed overreporting bias at lower values.

The overall diagnostic ability of all tools for measuring high USG values was excellent, with an AUC from 0.977 to 1.00, except for HYD 5°C, which demonstrated a good AUC of 0.867 for identifying a USG threshold of 1.020. Sensitivity was high in all cases, with slightly lower values for the HYD 5°C (0.750) and HYD 20°C (0.875); specificity based on a cutoff value of 1.020 was optimal at 1.000 in all cases.

Experiment 2: Comparison of Refractometers

We compared the MAN and HYD outcomes with the DIG measurement at 20°C that was selected as the standard (Table 4). The measured median (interquartile range) USG value for the standard DIG at 20 °C was 1.021 (1.010–1.024). No differences were seen in comparison with the DIG 5°C (1.021 [1.009–1.024], $P = .93$), DIG 37°C (1.020 [1.009–1.024], $P = .81$), MAN 5°C (1.020 [1.011–1.025], $P = .71$), MAN 20°C (1.021 [1.010–1.025], $P = .81$), or MAN 37°C (1.021 [1.010–1.025], $P = .71$). Although not different from the selected standard DIG 20°C, the median reported values for HYD 5°C (1.018 [1.008–1.022], $P = .20$) and HYD 20°C (1.017 [1.008–1.021], $P = .20$) were lower. The only significant difference was between the standard and HYD 37°C (-0.0049 , $P = .008$). All correlations with the DIG 20°C can be considered strong at all temperatures, as they ranged from 0.97 to 1.00. This result indicated that DIG measurements at 5°C and 37°C and MAN and HYD measurements at all temperatures can be ranked as consistently similar to the DIG 20°C standard.

Bland-Altman plots showed that the HYD data consistently underreported hydration status at all temperatures. At 37°C, the measurements of all tools were as biased as the MAN 5°C, which was indicated by the moderate correlations between the different methods and the means of both methods ranging from $r = 0.37$ to 0.52, with $P \leq .034$. No difference was observed for DIG 5°C or MAN 20°C; the absolute differences between the DIG 20°C standard and DIG 37°C (-0.001), MAN 5°C (0.001), and MAN 37°C (0.001) measurements were small, but the HYD outcomes were considerably larger: -0.002 for the HYD 5°C and HYD 20°C and -0.006 for the HYD 37°C.

The overall diagnostic ability of all tools to classify low and high urine concentrations, based on receiver operating characteristics, was excellent, and all were able to identify a USG threshold of 1.020. The AUC scores for the DIG and MAN were ≥ 0.98 and the HYD scores ranged from 0.92 to 0.96. Sensitivity and specificity for the DIG and MAN at all temperatures were >0.85 . For the HYD, sensitivity ranged from 0.75 to 0.80 and specificity from 0.85 to 0.90.

Table 4. Standard DIG 20°C Urine Specific Gravity Values Compared With Outcomes for Different Tools and Temperatures

Measure	Urine Specific Gravity											
	DIG				Manual Optical Refractometer				Hydrometer			
	5°C	37°C	5°C	37°C	5°C	20°C	37°C	5°C	20°C	37°C		
Outcome												
Median	1.021	1.020	1.020	1.021	1.020	1.021	1.021	1.018	1.017	1.014		
Interquartile range	1.009–1.024	1.009–1.024	1.011–1.025	1.010–1.025	1.010–1.025	1.010–1.025	1.010–1.025	1.008–1.022	1.008–1.021	1.004–1.018		
Difference	0	-0.0005	-0.0006	-0.0005	-0.0006	-0.0005	-0.0009	-0.0003	-0.0012 ^a	-0.0049 ^a		
Correlation												
r	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.97	0.99		
95% CI	0.99–1.00	0.99–1.00	0.98–1.00	0.99–1.00	0.98–1.00	0.99–1.00	0.99–1.00	0.95–0.99	0.94–0.99	0.98–1.00		
Bland-Altman analysis												
Bias	0.000	-0.001	0.001	0.000	0.001	0.000	0.001	-0.002	-0.002	-0.006		
95% LoA	-0.001:0.001	-0.002:0.001	-0.002:0.003	-0.001:0.002	-0.002:0.003	-0.001:0.002	-0.001:0.002	-0.006:0.002	-0.006:0.002	-0.008:0.003		
n < or > 95% LoA	1 > LoA	1 > LoA	1 < and 1 > LoA	1 > LoA	1 > LoA	1 > LoA	1 > LoA	1 > LoA	1 > LoA	0		
R, P values	-0.01, .955	0.41, .017	0.37, .034	0.27, .133	0.37, .034	0.27, .133	0.52, .002	-0.03, .861	0.21, .251	0.45, .009		
Receiver operating curve												
Area under the curve	0.975	0.981	0.987	0.985	0.987	0.985	0.987	0.921	0.954	0.964		
Sensitivity	0.846	0.900	0.873	0.850	0.873	0.850	0.873	0.746	0.773	0.796		
Specificity	1.000	0.900	0.950	0.850	0.950	0.850	0.950	0.900	0.850	0.950		

Abbreviations: DIG, digital refractometer; LoA, limits of agreement for the Bland-Altman analysis.

^a Indicates a difference from standard BRX for the hydrometer at 37°C ($P < .05$).

DISCUSSION

When we compared the °Bx solution with the USG measurements, reporting bias and a misclassification of samples at higher cutoff points (1.025 and 1.030) were evident, especially for the samples measured at 37°C. Overall, the HYD displayed the highest degree of misreporting at all temperatures. Experiment 2, in which we measured urine samples using the 20°C DIG as the standard, showed that the DIG and MAN displayed no mean differences when ranked against the outcomes of both tools and the receiver operating characteristic analysis, whereas the HYD reflected clear differences from the selected standard, as present in the greater bias.

We suggest that the DIG and MAN performed similarly, which contradicts the earlier findings of Minton et al,¹⁶ who, using Bland-Altman analyses, suggested an underestimation bias for 2 types of DIGs compared with a MAN as USG values increased. In the current study, the HYD underreported in both experiments versus the selected standards. Rudinsky et al¹⁵ compared 4 refractometers and noted measurement differences of ± 0.002 . We observed that the urine measurement differences were smaller (± 0.001). Athletic trainers, dietitians, and physicians expect consistent USG results, regardless of the refractometer used; yet it is important to understand that refractometers may report differently.¹⁵ To our knowledge, only 1 other group of researchers investigated the effect of temperature on mean outcomes for USG. Jin et al⁹ compared MAN and HYD measurements at 4°C, 20°C, and 37°C. They obtained similar results, with no effect of temperature on mean outcomes of the refractometer but described underreporting of the HYD at 37°C. We added to this previous report by validating the refractometers and HYD in different ways (median difference, ranking of outcomes, magnitude of underestimation, and accuracy of reporting) using both urine samples and an independent °Bx reference solution. The inclusion of an independent marker (°Bx) was critical, as the Bland-Altman analysis revealed that all tools tended to underreport when USG increased.

Despite the suggestions of some manufacturers, we did not correct for temperature differences with any of the tools. The HYDs are known to be temperature sensitive,²⁴ and for each ± 3 °C of standard temperature (20°C), 0.001 should be added or subtracted from the original reading. Considering that the HYD measurements showed large internal variations (Table 2), we recommend that HYDs should not be used to detect severe dehydration, as the tool underreported even with the temperature correction.

As illustrated by the results in Table 2 and shown statistically in the Bland-Altman analysis in Table 3, all tools except the HYD 5°C misreported at higher USG concentrations. Although the absolute difference in misclassification increased only by 1 to 2 cases per set of measurements, the relevance of misreporting at high USG values must be considered, as this is the point at which athletes are classified into groups with a low or high urine concentration. Some sports, such as boxing, wrestling, and other weight-class-bound sports, use specific cutoffs as part of the weigh-in process.^{4,19,25} “Cutting weight” may influence the fluid balance, resulting in many athletes displaying USG values around 1.030.^{4,19,25} In other sports, such as cycling, lacrosse, soccer, and triathlon, many athletes present high USG values.^{5,26,27} For example,

Armstrong et al⁵ determined that 25% of all collected morning urine samples exceeded 1.029. Montazer et al²⁸ found that 38% of construction workers had USG values larger than 1.025. As high USG values around frequently used cutoff values for dehydration (1.026) and severe dehydration (1.030)¹⁰ are described in a wide range of high-risk individuals, the misreporting of refractometers is likely to affect the proper classification of hydration status among these populations.

We believe that one of the strengths of this study was our design, which was a combined approach with measurements of reference °Bx solutions as an independent marker and urine samples covering the same range of measurement values. This enabled us to investigate the accuracy of USG measurement over the spectrum from 1.000 to 1.040 that is normally used for measuring human urine. Another strength is that we used only fresh urine samples measured within 5 days of collection during experiment 2. The USG in fresh urine remains stable for at least 7 days when stored at 5°C, and freezing the samples negatively alters USG measurements.²⁷

A limitation is that we did not correct for large molecules in the urine, such as high levels of protein and glucose, which may influence USG readings.²⁴ However, our use of an independent marker (°Bx) was a strength; if the urine samples had been dried to determine the precise weight of the total solids,^{14,29} large molecules would have affected the calculated density of the urine sample based on the dry matter obtained.

This study is clinically relevant, as the assessment of chronic hydration status via USG is often used in athlete education.¹ A correct diagnosis requires accurate measurements. As we have shown here, values near clinically relevant dehydration (≥ 1.030) tended to indicate lower values, which were amplified by a 37°C sample temperature. At lower cutoff values, such as 1.020 and 1.026, which are often used in the literature,⁵ there is also a chance of underreporting. The underreporting is likely to be 0.001 to 0.003, depending on the tool, temperature, and concentration. Although these small differences may not be clinically relevant, they may have a large effect when professionals use the measurement to classify athletes as being euhydrated versus underhydrated. These differences would also affect the hydration advice given, with more individuals being classified on the lower side of the cutoff (toward being hydrated), whereas a substantial number should be classified as underhydrated. Therefore, despite the manufacturers' instructions about the temperature-correction abilities of their devices, we suggest that measurement temperatures should be standardized at 5°C or 20°C and that measurements at approximately 37°C should be discouraged. Finally, the DIG was the most time-efficient device studied, as it could be dipped directly into a sample to obtain a measurement, whereas the MAN required an additional action of pipetting a small drop of urine on a transparent glass plate or lens that then needed to be closed. The HYD was the most time-consuming and least accurate measurement. The specific HYD we used had only 1 scale printed on the side and often kept spinning in the glass jar after it was dropped in the sample, making it more difficult to obtain a good reading.

In conclusion, the HYD produced unreliable results and should not be used. The DIG and MAN provided

comparable results at multiple levels against an independent standard, and no clear differences were detected for medians, correlations, or receiver operating characteristics for the USG cutoff value of 1.020, regardless of sample temperature. Furthermore, USG reporting on individual urine samples may be slightly biased when urine concentration increases toward or above the threshold value of 1.025, particularly at higher sample temperatures. Thus, USG measurements should be standardized for temperature and preferably measured consistently at 5°C or room temperature (20°C) using a DIG or MAN.

ACKNOWLEDGMENTS

Stavros Kavouras, PhD, has served as a scientific consultant for Quest Diagnostics, Standard Process, and Danone Research and has active grants with Danone Research and Standard Process. We thank research nurse Ginger Hook, RN, CDE, in the College of Health Solutions Nutrition Laboratories for her support during this study.

REFERENCES

1. McDermott BP, Anderson SA, Armstrong LE, et al. National Athletic Trainers' Association position statement: fluid replacement for the physically active. *J Athl Train.* 2017;52(9):877–895. doi:10.4085/1062-6050-52.9.02
2. Armstrong LE. Assessing hydration status: the elusive gold standard. *J Am Coll Nutr.* 2007;26(5 suppl):575S–584S. doi:10.1080/07315724.2007.10719661
3. American College of Sports Medicine, Sawka MN, Burke LM, et al. American College of Sports Medicine position stand. Exercise and fluid replacement. *Med Sci Sports Exerc.* 2007;39(2):377–390. doi:10.1249/mss.0b013e31802ca597
4. Fernández-Eliás VE, Martínez-Abellán A, López-Gullón JM, et al. Validity of hydration non-invasive indices during the weightcutting and official weigh-in for olympic combat sports. *PLoS One.* 2014;9(4):e95336. doi:10.1371/journal.pone.0095336
5. Armstrong LE, Pumerantz AC, Fiala KA, et al. Human hydration indices: acute and longitudinal reference values. *Int J Sport Nutr Exerc Metab.* 2010;20(2):145–153. doi:10.1123/ijsnem.20.2.145
6. Armstrong LE. Hydration assessment techniques. *Nutr Rev.* 2005;63(6 pt 2):S40–S54. doi:10.1111/j.1753-4887.2005.tb00153.x
7. Schmidt M, Brinkerhoff R, Stocking C. The effect of education on hydration status of female collegiate tennis players at Southern Utah University. *J Acad Nutr Diet.* 2018;118(9 suppl):A50.
8. Wyness SP, Hunsaker JHH, Snow TM, Genzen JR. Evaluation and analytical validation of a handheld digital refractometer for urine specific gravity measurement. *Pract Lab Med.* 2016;5:65–74. doi:10.1016/j.plabm.2016.06.001
9. Jin BK, Choi YA, Yoon YA, Shin SH, Kim ES, Lee JH. The effects on urine specific gravity according to temperatures and concentrations of glucose, protein, sodium chloride and urea. *Korean J Med Technol.* 1993;25:160–167.
10. Adams JD, Capitan-Jiménez C, Huggins RA, Casa DJ, Mauro-moustakos A, Kavouras SA. Urine reagent strips are inaccurate for assessing hypohydration: a brief report. *Clin J Sport Med.* 2019;29(6):506–508. doi:10.1097/ISM.0000000000000555
11. Bennett AD, McKnight GE, Dodkin SJ, Simpson KE, Schwartz AM, Gunn-Moore DA. Comparison of digital and optical hand-held refractometers for the measurement of feline urine specific gravity. *J Feline Med Surg.* 2011;13(2):152–154. doi:10.1016/j.jfms.2010.09.012
12. Paris JK, Bennett AD, Dodkin SJ, Gunn-Moore DA. Comparison of a digital and an optical analogue hand-held refractometer for the

- measurement of canine urine specific gravity. *Vet Rec.* 2012;170(18):463. doi:10.1136/vr.100348
13. Tvedten HW, Norén A. Comparison of a Schmidt and Haensch refractometer and an Atago PAL-USG Cat refractometer for determination of urine specific gravity in dogs and cats. *Vet Clin Pathol.* 2013;43(1):63–66. doi:10.1111/vcp.12110
 14. Tvedten HW, Ouchterlony H, Lilliehöök IE. Comparison of specific gravity analysis of feline and canine urine, using five refractometers, to pycnometric analysis and total solids by drying. *N Z Vet J.* 2015;63(5):254–259. doi:10.1080/00480169.2014.1002553
 15. Rudinsky AJ, Wellman M, Tracy G, Stoltenberg L, DiBartola SP, Chew DJ. Variability among four refractometers for the measurement of urine specific gravity and comparison with urine osmolality in dogs. *Vet Clin Pathol.* 2019;48(4):702–709. doi:10.1080/00480169.2014.1002553
 16. Minton DM, O’Neal EK, Torres-McGehee TM. Agreement of urine specific gravity measurements between manual and digital refractometers. *J Athl Train.* 2015;50(1):59–64. doi:10.4085/1062-6050-49.3.47
 17. Ercan M, Akbulut ED, Abuşoğlu S, et al. Stability of urine specimens stored with and without preservatives at room temperature and on ice prior to urinalysis. *Clin Biochem.* 2015;48(13–14): 919–922. doi:10.1016/j.clinbiochem.2015.05.016
 18. Pethick WA, Murray HJ, McFadyen P, Brodie R, Gaul CA, Stellingwerff T. Effects of hydration status during heat acclimation on plasma volume and performance. *Scand J Med Sci Sports.* 2019;29(2):189–199. doi:10.1111/sms.13319
 19. Pettersson S, Berg CM. Hydration status in elite wrestlers, judokas, boxers, and taekwondo athletes on competition day. *Int J Sport Nutr Exerc Metab.* 2014;24(3):267–275. doi:10.1123/ijsnem.2013-0100
 20. Muñoz CX, Johnson EC, Demartini JK, et al. Assessment of hydration biomarkers including salivary osmolality during passive and active dehydration. *Eur J Clin Nutr.* 2013;67(12):1257–1263. doi:10.1038/ejcn.2013.195
 21. Hamouti N, Del Coso J, Ávila A, Mora-Rodriguez R. Effects of athletes’ muscle mass on urinary markers of hydration status. *Eur J Appl Physiol.* 2010;109(2):213–219. doi:10.1007/s00421-009-1333-x
 22. Introduction to ROC curves. University of Nebraska Medical Center. <http://gim.unmc.edu/dxtests/ROC1.htm>. Accessed November 11, 2019.
 23. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39(2):175–191. doi:10.3758/bf03193146
 24. Oppliger RA, Bartok C. Hydration testing of athletes. *Sports Med.* 2002;32(15):959–971. doi:10.2165/00007256-200232150-00001
 25. Jetton AM, Lawrence MM, Meucci M, et al. Dehydration and acute weight gain in mixed martial arts fighters before competition. *J Strength Cond Res.* 2013;27(5):1322–1326. doi:10.1519/JSC.0b013e31828a1e91
 26. Olzinski S, Beaumont J, Toledo M, Yudell A, Johnston CS, Wardenaar FC. Hydration status and fluid needs of Division I female collegiate athletes exercising indoors and outdoors. *Sports.* 2019;7(7):155. doi:10.3390/sports7070155
 27. Jacobs BM, Patience JF, Dozier WA III, Stalder KJ, Kerr BJ. Effects of drying methods on nitrogen and energy concentrations in pig feces and urine, and poultry excreta. *J Anim Sci.* 2011;89(8):2624–2630. doi:10.2527/jas.2010-3768
 28. Montazer S, Farshad AA, Monazzam MR, Eyvazlou M, Yaraghi AA, Mirkazemi R. Assessment of construction workers’ hydration status using urine specific gravity. *Int J Occup Med Environ Health.* 2013;26(5):762–769. doi:10.2478/s13382-013-0143-x
 29. Adams JD, Kavouras SA, Johnson EC, et al. The effect of storing temperature and duration on urinary hydration markers. *Int J Sport Nutr Exerc Metab.* 2017;27(1):18–24. doi:10.1123/ijsnem.2016-0098

Address correspondence to Floris Wardenaar, PhD, College of Health Solutions, Arizona State University, 425 North 5th Street, Phoenix, AZ 85004. Address email to Floris.wardenaar@asu.edu.