Reliability of three urinalysis methods used in the assessment of hydration

Aric J. Warren1, Matthew S. O’Brien1, Doug B. Smith2
1 Associate Professor, Department of Athletic Training, Oklahoma State University Center for Health Sciences, 1111 W. 17th Street, Tulsa, OK 74107, USA
2 Professor, Oklahoma State University, School of Kinesiology, Applied Health and Recreation, Stillwater, OK 74078, USA

Abstract

Objective: The purpose of this study was to determine the intratester and intertester reliability of three common methods of assessing hydration; urine color chart (UCC), dipstick reagent strip (DRS) and refractometer (REF).

Methods: Twenty-three male collegiate wrestlers (n = 23, age = 20.09 ± 1.35, weight = 78.73 ± 11.25 kg, height = 174.49 ± 7.23 cm) provided urine samples on three separate occasions totaling 69 samples (n = 69). Samples were analyzed with repeated measures by three testers, three trials, and three separate methods of assessment. Results: All methods had very high intertester reliability as demonstrated by Cronbach’s Alpha coefficients; DRS (r = .985), UCC (r = .973) and REF (r = .968). Intraclass correlation coefficient ranges were also very high; DRS .983 - .994, UCC 964 - 983, and REF .829 - .996. Conclusions: The three modes of urine hydration assessment are highly reliable, and are a practical, noninvasive method to evaluate hydration status in the field.

Keywords: Hydration; Reliability; Urinalysis.

INTRODUCTION

Athletes that compete in body weight restrictive sports such as wrestling, judo, and rowing sometimes go to extreme measures to qualify for a specific weight class or event. Wrestlers lose weight frequently, using rapid weight reduction techniques [1]. It is a common practice in the sport of wrestling to control weight through fasting, dietary restriction, excessive exercise, and dehydration methods including fluid restriction, the use of saunas, laxatives, diuretics, vomiting, and vapor-impermeable suits [1-6]. One of the dangers involved with rapid weight loss practices, and exercising in hot, humid environments is the risk of dehydration. Consequences of dehydration include elevated heart rate [6], stroke volume reduction [6-8], compromised thermoregulation [7], resulting in increased resting body temperature and elevated core temperatures throughout mild and extreme exercise [9]. Increased sweating during exercise can cause core body temperatures to climb between 0.15 – 0.20° C for every 1% of body weight lost [10].

Adequate hydration has an important function in the performance of athletes. Loss of 1-2% of body weight due to dehydration compromises physiological function and can influence performance negatively [11]. Dehydration with body weight loss of greater than 3% creates an increased for exertional heat illness [11]. An accepted assessment technique of monitoring hydration status is by evaluating the specific gravity of urine. Specific gravity of urine indicates the relative density of the urine sample (mass per volume) in contrast to pure water. Properties of urine reflect body water status, with the urine concentrated in a more dehydrated state and dilute in a normal hydrated state [8].

Urine specimens of normal, healthy adults usually have a specific gravity within the range of 1.013 to 1.029 [8]. Euhydration, or adequate hydration has been identified as having a urine specific gravity value of ≤ 1.020 [11]. In a dehydrated condition, the specific gravity of urine exceeds 1.030 [8]. Urine specific gravity is commonly assessed with a refractometer and dipstick reagent strip. Refractometry involves reading the specific gravity of the urine specimen on a scale that ranges from 1.000 to 1.040. Specific gravity measured by dipstick reagent strips involves identifying the value of specific gravity by comparing the change in color that takes place on the reagent strip to a color-coded specific gravity scale. Another common method of measuring hydration status is to compare the color of one’s urine to a color chart...
scale that can indicate the level of hydration [8]. In all of these commonly used techniques of hydration assessment, the specific gravity or level of hydration is dependent on a subjective interpretation by the tester, either by reading the level of specific gravity on a scale, or to match colors on a chart. While refractometry has been identified as the standard measure for urine specific gravity [12] and has been deemed a valid indicator of measuring hydration status in a non-laboratory or field setting [11, 13-15], multiple or repeated tests are often performed to measure temporal variations in the hydration level of a patient. In many cases, multiple tests are performed on a patient during acute management, or multiple clinicians assess the same patient. With a multitude of testers and the subjective interpretation of the visual scales on the equipment, these commonly used techniques of hydration assessment need to be examined for reliability between testers and amongst a single tester to ensure quality patient management in the field.

While some of these methods have been determined a reliable measure of hydration status by some [10], others have determined them to remain subjective [17]. Unless a digital refractometer is used, these techniques rely on subjective readings of color, color change, and scale measurement readings by the examiner. The subjectivity of examiner interpretation is concerning, as are the factors that may affect the color of urine. Armstrong et al [8] noted that day-to-day reliability of urine color measurements is enhanced if meals, fluid consumption during exercise, training, and the time of urine collection are consistent. While researchers indicate that urine color validity reflects body hydration [9,18], it is the subjective nature of this assessment that questions the reliability of this technique. Since dehydration occurs before the need to drink is perceived [19], and is likely to impair athletic performance and is linked to a myriad of health issues, determining reliable methods of identifying dehydration in the field is essential.

Therefore, the aim of the current study was to examine the intertester and intratester reliability of 3 common methods of assessing hydration; urine color chart (UCC), evaluation of specific gravity of urine using a dipstick reagent strip (DRS) and a visual analog refractometer (REF) between 3 testers and 3 separate trials.

MATERIALS and METHODS

Participants

Twenty-three male collegiate-level wrestlers (n = 23, age = 20.09 ± 1.35, height = 174.49 ± 7.23, weight = 78.73 ± 11.25) participated in the study. Each subject provided a urine specimen at three different time points for a total of 69 urine samples (n = 69). Written informed consent was approved by the institution’s Review Board and was provided by each subject prior to participation. All subjects who provided urine specimens were of good health and without a history of renal disorders, diabetes, or heat illness.

Investigators

One certified athletic trainer, one exercise physiologist, and one exercise physiology graduate student who were experienced in urine analysis via color chart comparison, dipstick reagent strip, and refractometer measurements served as testers.

Instruments

Urine color (UCC) was assessed using The Urine Color Chart [8] (Human Kinetics, Champaign, IL). The Urine Color Chart consists of a Likert scale of eight colors described as varying from very pale yellow, to brownish green. The lightest of the urine colors on the chart equates to number 1 on the Likert scale while the darkest of urine colors is equal to number 8 on the scale. Urine specific gravity was measured using two methods; urine dipstick reagent strips (DRS), (Bayer® Multistix 10SG Reagent Strips for Urinalysis, Bayer Corporation, Elkhart, IN), and a clinical hand-held visual analog refractometer (REF) (Model 300CL, Atago Inc., Japan).

Procedures

Our research team collected data over the course of one competitive wrestling season. Urine specimens were obtained before the beginning of the competitive season, during mid-season, and upon completion of the regular season. Instructions not to modify dietary and fluid intake, and to eat and drink ad libitum were provided to participants. Research assistants collected fresh, midstream urine specimens in the morning in a clean, dry, sterile container. We collected urine specimens that were the first void of the day for consistency and accuracy. Fresh urine samples were examined three individual times by three separate researchers using the urine color chart (UCC), a urinalysis dipstick reagent strip (DRS), and a hand-held refractometer (REF). We performed all analyses on the same day that the urine was collected, followed manufacturer and suggested guidelines for each analysis method, and performed all urinalyses in the same well-lit room with fluorescent bulbs. We compared each specimen to a urine color chart held in front of a white background [10,14]. For consistency sake, prior to analysis we agreed to choose the higher of two numbers if UCC reading fell between two colors on the chart. Likewise, when performing the specific gravity testing if a reading fell between two values on the refractometer or color chart on the dipstick reagent strips, the higher of the two numbers would be recorded. Per study protocol, any oddly discolored urine (reddish in color) was to be discarded. A recorder transcribed all measurements on a data-recording sheet to prevent bias comparison of previous measurement values. In addition, each urine specimen was blinded by name to the examiners and randomized during analysis.

Data analysis

We calculated intertester and intratester reliability coefficients for UCC, DRS, and REF to identify reliability of the methods between testers and within testers. Cronbach’s alpha coefficients were calculated for the three methods of hydration assessment, with three testers and three trials each. Intertester and intratester relationships among the three trials for each method of hydration assessment were calculated using Pearson correlation coefficients. Intertester and intratester correlation coefficient strength was established according to Vincent [20]. All statistics were calculated using SPSS 14.0 (SPSS Inc., Chicago, IL), with an alpha level of p ≤ .05 established a priori for all tests.

RESULTS

The results of our reliability analysis showed that the three measures of hydration assessment (UCC, DRS, and REF) are highly reliable. Cronbach’s Alpha coefficients show very good reliability for the three methods of hydration evaluation with DRS having the most consistent measures (r = 0.985, p = .000), followed by UCC (r = 0.973, p = .000) and REF (r = 0.968, p = .000) (Table 1). Mean values and standard deviations for the three methods of hydration assessment, three testers and three trials are provided (Table 2). We found that intertester reliability was good for REF (r = 0.870 – 0.954, p = .000), moderate for DRS (r = 0.823 – 0.900, p = .000), and moderate for UCC (r = 0.775 – 0.855, p = .000) (Table 3). Intraclass correlation coefficients for the three testers were also found to be good for DRS (r = 0.983 – 0.994, p = .000) and UCC (r = 0.964 – 0.983, p = .000), and moderate to good for REF (r = 0.829 – 0.996, p = .000) (Table 4).
Table 1: Reliability coefficients for each method using three testers and three trials

<table>
<thead>
<tr>
<th>Method</th>
<th>Cronbach’s Alpha coefficient</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRS</td>
<td>.985</td>
<td>p = .000</td>
</tr>
<tr>
<td>UCC</td>
<td>.973</td>
<td>p = .000</td>
</tr>
<tr>
<td>REF</td>
<td>.968</td>
<td>p = .000</td>
</tr>
</tbody>
</table>

Table 2: Hydration indices for Method x Trial x Tester (Mean ± SD)

<table>
<thead>
<tr>
<th>Method</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tester 1</td>
<td>5.348 ± 1.589</td>
<td>5.406 ± 1.556</td>
<td>5.319 ± 1.613</td>
</tr>
<tr>
<td>Tester 2</td>
<td>5.348 ± 1.474</td>
<td>5.377 ± 1.446</td>
<td>5.348 ± 1.423</td>
</tr>
<tr>
<td>Tester 3</td>
<td>5.723 ± 1.403</td>
<td>5.667 ± 1.502</td>
<td>5.725 ± 1.371</td>
</tr>
<tr>
<td>DRS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tester 1</td>
<td>1.024 ± 0.006</td>
<td>1.024 ± 0.006</td>
<td>1.024 ± 0.007</td>
</tr>
<tr>
<td>Tester 2</td>
<td>1.023 ± 0.006</td>
<td>1.023 ± 0.006</td>
<td>1.023 ± 0.006</td>
</tr>
<tr>
<td>Tester 3</td>
<td>1.024 ± 0.006</td>
<td>1.024 ± 0.006</td>
<td>1.024 ± 0.006</td>
</tr>
<tr>
<td>REF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tester 1</td>
<td>1.028 ± 0.009</td>
<td>1.027 ± 0.005</td>
<td>1.027 ± 0.005</td>
</tr>
<tr>
<td>Tester 2</td>
<td>1.027 ± 0.007</td>
<td>1.027 ± 0.006</td>
<td>1.027 ± 0.005</td>
</tr>
<tr>
<td>Tester 3</td>
<td>1.027 ± 0.005</td>
<td>1.027 ± 0.005</td>
<td>1.027 ± 0.005</td>
</tr>
</tbody>
</table>

N = 69

Table 3: Intertester reliability correlation coefficients (Pearson r correlations)

<table>
<thead>
<tr>
<th>Tester interaction</th>
<th>Reliability coefficient *</th>
</tr>
</thead>
<tbody>
<tr>
<td>REF</td>
<td></td>
</tr>
<tr>
<td>Tester 1 – Tester 2</td>
<td>.870</td>
</tr>
<tr>
<td>Tester 2 – Tester 3</td>
<td>.954</td>
</tr>
<tr>
<td>Tester 3 – Tester 1</td>
<td>.902</td>
</tr>
<tr>
<td>DRS</td>
<td></td>
</tr>
<tr>
<td>Tester 1 – Tester 2</td>
<td>.889</td>
</tr>
<tr>
<td>Tester 2 – Tester 3</td>
<td>.823</td>
</tr>
<tr>
<td>Tester 3 – Tester 1</td>
<td>.900</td>
</tr>
<tr>
<td>UCC</td>
<td></td>
</tr>
<tr>
<td>Tester 1 – Tester 2</td>
<td>.781</td>
</tr>
<tr>
<td>Tester 2 – Tester 3</td>
<td>.775</td>
</tr>
<tr>
<td>Tester 3 – Tester 1</td>
<td>.855</td>
</tr>
</tbody>
</table>

* Pearson r correlation coefficients all significant at the p ≤ .01 level.

Table 4: Intratester reliability coefficients (Pearson r correlations)

<table>
<thead>
<tr>
<th>Method</th>
<th>Tester 1 *</th>
<th>Tester 2 *</th>
<th>Tester 3 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRS</td>
<td>.986</td>
<td>.994</td>
<td>.983</td>
</tr>
<tr>
<td>UCC</td>
<td>.971</td>
<td>.983</td>
<td>.964</td>
</tr>
<tr>
<td>REF</td>
<td>.829</td>
<td>.936</td>
<td>.996</td>
</tr>
</tbody>
</table>

* Pearson r correlation coefficients all significant at the p = .000 level.
DISCUSSION

Urinary indices are a common measure to assess hydration levels in various patient populations, as well as collegiate and high school athletes. Urinary indices of specific gravity and color are more receptive to small variations in hydration status and are often the earliest signs of imminent dehydration [21]. Heat-related deaths that have occurred in sports such as football and wrestling have demonstrated that the risks associated with dehydration are high. Due to the multidisciplinary nature in caring for individuals with acute dehydration, and the potential severity of physiological ramifications that dehydration imposes, it is crucial that healthcare clinicians employ the most reliable techniques to recognize and treat dehydration properly.

The determination of hydration level is commonly assessed through examination of urine and blood.

Tests of urine include measures of osmolality, specific gravity, protein and potassium. Blood serum counts for sodium, potassium, chloride, and serum osmolality, as well as various ratios of urine-to-plasma osmolality, extracellular water-to-intracellular water ratios, and creatinine ratios can be performed. Other hematological tests include hematocrit and hemoglobin concentration. The hematological indices, while valid indicators of hydration measurement, require venipuncture and are better suited as a late indicator of dehydration [17]. Acute exercise, particularly endurance training, has been associated with decreases in plasma volume [22]. Serum measurements of sodium concentration, plasma osmolality, and hematocrit are not as sensitive as urinary evaluation of color, osmolality, or specific gravity [14]. In addition, many of these indices of hydration require laboratory testing, and are not practical for use in the field or on-site at athletic events.

The most commonly used techniques for urinary indices of hydration are urine osmolality, urine specific gravity, and urine color [23]. Urine color, specific gravity of urine, and urine osmolality are more sensitive indices of determining moderate dehydration than measurements of blood [23]. An osmometer quantifies the amount of osmoles of solute per kilogram of solution [24]. Particles like urea, proteins, and glucose are not detected by the osmometer, but particles that dissociate in the solution such as sodium chloride are detected. While osmolality is an accurate measure of hydration, its use is not practical in the field and is more suited for laboratory testing. Two common methods of measuring specific gravity of urine are through refractometry and dipstick reagent strips. Specific gravity of urine is defined as the density comparison between a urine sample and pure water [24]. The specific gravity of a urine specimen can be dependent on the concentration of glucose, urea, and protein particles within the solution [24]. Refractometry involves measuring the refraction of a beam of light on a scale as it passes through a urine sample [25].

Athletes have been advised to evaluate their own levels of hydration by observing the color of urine after evacuation, and monitoring sweat loss, fluid intake and body weight [11]. Some authorities have recommended urinalysis as a screening tool because it involves a relatively easy, inexpensive, and noninvasive evaluation of a body fluid [23,24]. The National Collegiate Athletic Association (NCAA) and National Federation of State High School Associations (NFHS) in the United States have implemented hydration testing as an integral part of minimum weight assessment of competitive wrestlers [26,27]. With respect to high school wrestling, guidelines from NFHS have mandated that every state athletic association develop and employ a weight-management program that includes examination of urine specific gravity not to exceed 1.025 [27]. While standards exist for hydration testing in high school wrestling weight management programs, to date the NFHS has only ruled that urine specific gravity be assessed during the weight certification process with no specifications as to which method is to be used [27]. Other athletic associations have determined that urine specific gravity must be measured, and that refractometry be used as the criterion measure when evaluating wrestlers’ body composition to determine minimum wrestling weight [26,28].

Refractometry has been suggested as the proper technique of hydration status assessment [11-16] and thus has been adopted as the method of choice to evaluate specific gravity of urine in collegiate wrestling programs [26] and some high school wrestling associations [28]. Monitoring measurements of urine specific gravity by refractometer or reagent strip is appealing to many because of its cost, ease of use, and convenience [26,29]. Our findings indicate that specific gravity of urine, when measured by refractometer, has moderate to good reliability. Inter-tester reliability ranged from \( r = .870 \) to .954 between testers with moderate to good intratest reliability ranging from \( r = .829 \) to .996. Others have also found the refractometer to have good reliability between and among testers and trials. Steumpfel et al [16] found intratest reliability to be very high using a refractometer to assess specific gravity (\( r = .998 \)). While the reliability coefficients in our study are quite high among some of the testers for REF, when compared to DR5 and UCC, refractometer ranked third overall among the three methods tested. This could be explained by a lack of clarity in the viewer on the refractometer. At times, the scale and reading of the specific gravity became blurred if a large volume of urine was placed on the test slide, or if it the test slide was not completely dry after cleaning between uses. Even with the high degree of reliability that our results have shown for the refractometer, the use of a digital refractometer can eliminate the subjective nature involved with the hand held refractometer.

Some athletic agencies recommend the refractometer as the only method to assess urine specific gravity for the collegiate wrestling weight certification process [16,26,28]. Our results indicate that the refractometer has good intertester reliability and moderate to good intratest reliability, but dip stick reagent strips have a slightly higher correlation coefficient (\( r = .985 \) versus \( r = .968 \)), and have better intratest reliability than UCC or REF. Our data show that dipstick reagent strip measurements of urine specific gravity are as reliable as urinary indices using the refractometer. According to the NCAA in 1998 [30] both the refractometer and dip stick reagent strips are acceptable methods of measuring specific gravity in urine. However, according to the NCAA rules modification in 1999 [31], only the refractometer was the acceptable method of obtaining urine specific gravity. Our findings support the NCAA recommendation from 1998 and suggest that agencies such as the NCAA and high school athletic associations could rely on urinary specific gravity assessments using both the refractometer and dipstick reagent strips. Other researchers agree with our findings and have suggested that reagent strips be an acceptable substitute to refractometry [32,34], however some have determined that dipstick reagent strips are not reliable between examinations for individual testers and among groups of testers [16].

It has been reported that a comparison of urine color with a urine color chart [8] is the simplest method to use to determine dehydration status [13], is strongly correlated to urinary plasma indices [18] and with urine specific gravity [14]. While the NCAA or NFHS do not consider hydration status measured by urine color to be an acceptable method of hydration assessment for their weight certification programs, urine color may be used as a meaningful index of hydration when other methods (urine osmolality and urine specific gravity) are not practical or readily available [14]. Despite not being utilized by many athletic authorities as a means of hydration assessment, the National Athletic Trainers’ Association recommends that a color of urine less than or equal to “4” on the color chart be the acceptable upper range at weigh-ins for events that use weight classes [14]. In athletic practice, the use of urine color is often used to determine the need for fluid replacement [35] and to offer advice to athletes as a means of self hydration assessment. It has also been reported that urine color is a poor index of determining hydration status and the need for fluid
intake during the first six hours post exercise [35]. While many would consider urine color as a subjective assessment of hydration status, researchers have found the intertester reliability of urine color assessment to range from average to good [17]. Our results demonstrate that observing urine color as a means of hydration assessment is moderately reliable with interclass correlation coefficients ranging between r = 0.775 – 0.855 for the 3 testers. Intraclass correlation coefficients for urine color were also found to be good (r = 0.964 – 0.983). Compared to the two specific gravity measures, urine color had the lowest intertester reliability of the three methods tested, but fared much better than the refractometer in reliability among testers.

This study was not without limitations. One weakness pertains to the absence of a recent history of the subjects' physical activity, dietary intake, and fluid consumption. Since all subjects that were sampled participated on the same competitive wrestling team, it was assumed that physical activity levels remained relatively constant between subjects. Subjects were free to eat and drink ad libitum throughout the study and were not instructed to adhere to a specific caloric or fluid intake protocol. Many of the subjects used in the study were attempting to maintain a specific body weight required for their respective sport. Since this study focused on the reliability of conventional methods used to monitor hydration levels in athletes, fluid consumption of subjects was not relevant to determine the reliability of the instruments in question.

Because urine color can be altered by dietary intake [36], medications [37,38], illness [39], vitamin supplementation, and exercise [12,14,40] as well as poor renal function [17], the color of urine in the samples may have not accurately reflected true specific gravity of the urine. Fresh fruits such as some berries, or vegetables (rhubarb, carrots, beets, and fava beans) can affect the color of urine turning it a pink or orange tint, but have been found to create a minimal effect comparing the color to the urine color chart [17]. Some medications, including over-the-counter pain medications such as ibuprofen and aspirin can also alter urine color into a readily identifiable reddish color [17]. Although medication and vitamin intake was not assessed in the subjects, any reddish colored urine was discarded from analysis in the study. Multivitamin usage also has the potential to discolor urine making it more yellow in color. However, Mentes et al [17] found no significant differences in urine color in participants who were taking multivitamins and those who were not. Future research in hydration monitoring should include a history of recent illness, dietary intake, and vitamin supplementation to help screen data for foods and vitamins that can potentially alter urine color so those respective samples are not included in the analysis. An additional limitation of the study is that we did not include an assessment of protein and glucose levels in the urine samples. Some researchers have proposed that protein [23] and glucose [34] may have an impact on urinary specific gravity measurements using reagent strips while others found glucose and protein levels do not affect specific gravity [34].

**CLINICAL IMPLICATIONS AND CONCLUSIONS**

Several methods examining blood and urine are used in the determination of hydration status in physically active people. Our findings suggest that three common methods of body fluid assessment to evaluate hydration are highly reliable and a noninvasive method of hydration status. Our results indicate that refractometry or dipstick reagent strips to assess specific gravity of urine, and hydration assessed by the use of a urine color chart are reliable techniques both between testing sessions and among multiple testers.

This information can be useful to athletic trainers, athletic therapists, and other health care professionals when monitoring the hydration status of patients in a field or clinical setting.

**Conflicts of Interest**

The authors declare no conflicts of interest.

**REFERENCES**