

Accuracy, precision, usability, and cost of free chlorine residual testing methods

Anna Murray and Daniele Lantagne

ABSTRACT

Chlorine is the most widely used disinfectant worldwide, partially because residual protection is maintained after treatment. This residual is measured using colorimetric test kits varying in accuracy, precision, training required, and cost. Seven commercially available colorimeters, color wheel and test tube comparator kits, pool test kits, and test strips were evaluated for use in low-resource settings by: (1) measuring in quintuplicate 11 samples from 0.0–4.0 mg/L free chlorine residual in laboratory and natural light settings to determine accuracy and precision; (2) conducting volunteer testing where participants used and evaluated each test kit; and (3) comparing costs. Laboratory accuracy ranged from 5.1–40.5% measurement error, with colorimeters the most accurate and test strip methods the least. Variation between laboratory and natural light readings occurred with one test strip method. Volunteer participants found test strip methods easiest and color wheel methods most difficult, and were most confident in the colorimeter and least confident in test strip methods. Costs range from 3.50–444 USD for 100 tests. Application of a decision matrix found colorimeters and test tube comparator kits were most appropriate for use in low-resource settings; it is recommended users apply the decision matrix themselves, as the appropriate kit might vary by context.

Key words | chlorination, developing countries, drinking water, field testing methods, free chlorine residual, household water treatment

Anna Murray (corresponding author)
Daniele Lantagne
Department of Civil and Environmental
Engineering,
Tufts University,
200 College Avenue,
Medford,
MA 02155,
USA
E-mail: anna.murray@tufts.edu

INTRODUCTION

Chlorine is the most common drinking water disinfectant worldwide, and has been used in municipal water treatment in the United States and Europe since the early 20th century. Chlorination of drinking water is considered one of the advances that virtually eradicated epidemic diarrhea in the United States and Europe (Cutler & Miller 2005). The advantages of chlorine disinfection are that it is inexpensive; simple to use; effective at inactivating most disease-causing pathogens in water; and residual chlorine is maintained in the water, which protects against recontamination in water distribution, transport, and storage (WHO 2011). The drawbacks of chlorine disinfection include user taste acceptability, chlorine's ineffectiveness against the protozoa *Cryptosporidium*, and the formation of potentially cancer-causing disinfection byproducts such as trihalomethanes

and haloacetic acids. However, 'the risks to health from these byproducts are extremely small in comparison with the risks associated with inadequate disinfection, and it is important that disinfection efficacy not be compromised in attempting to control such byproducts' (WHO 2011). Although alternative disinfectants exist (ozone, ultraviolet light, and halogens like iodine and bromine), 98% of water treatment facilities in the United States that disinfect water do so with chlorine-based disinfectants (Black & Veatch Corporation 2010).

When chlorine is added to water for disinfection, it reacts irreversibly with various constituents that exert chlorine demand; including natural organic matter, ammonia, nitrogen, hydrogen sulfide, and metals such as iron and manganese. The chlorine that has reacted becomes

doi: 10.2166/wh.2014.195

unavailable for disinfection. What remains after chlorine demand is met is known as total chlorine residual (TCR), and consists of: (1) combined chlorine, which is chlorine combined with ammonia to form chloramines (monochloramine: NH_2Cl , dichloramine: NHCl_2 , and trichloramine: NCl_3); and (2) free chlorine residual (FCR), consisting primarily of hypochlorous acid and hypochlorite (OCl^-). Free chlorine is a more effective disinfectant than is combined chlorine; combined chlorine concentration must be increased 25-fold, or the contact time 100-fold, to achieve the same inactivation of various microorganisms as that of free chlorine (APHA/AWWA/WEF 2005; Black & Veatch Corporation 2010).

The efficiency of chlorine disinfection depends on several factors, including the type and concentration of organisms being inactivated, chlorine dosage and contact time, water temperature and pH, and the presence of interfering substances in the water that exert a chlorine demand (Black & Veatch Corporation 2010). Initial chlorine dosages for water treatment are calibrated to meet chlorine demand and maintain a FCR sufficient for adequate disinfection during water distribution, transport, and storage. Thus, FCR presence in drinking water indicates two things: (1) the water was treated with a chlorine dose sufficient to inactivate most viruses and bacteria that cause diarrheal disease; and (2) the system is operating effectively and the water is safe against recontamination (CDC 2008). The World Health Organization (WHO) recommends a minimum FCR concentration in drinking water of 0.2 mg/L and a maximum of 5.0 mg/L. The maximum recommendation is a conservative health-based maximum guideline, as 'no specific adverse treatment-related effects have been observed in humans and experimental animals exposed to chlorine in drinking-water' (WHO 2011). WHO also notes that a taste barrier exists for FCR well below 5.0 mg/L, and sometimes as low as 0.3 mg/L. The United States Environmental Protection Agency (USEPA) maximum contaminant level (MCL) for FCR is 4.0 mg/L (USEPA 2006).

In addition to continuous dosing in piped water supplies in areas with infrastructure systems, chlorine is also used to disinfect pipes and installations after construction, repair, or cleaning, and is used to directly disinfect stored household drinking water in low-resource settings, such as developing

countries and in emergencies where there is little reliable infrastructure (WHO n.d.). For household water treatment programs using chlorine in these environments, the Centers for Disease Control and Prevention (CDC) recommends a maximum FCR of 2.0 mg/L 1 h after chlorine addition (to not exceed taste acceptability concerns) and a minimum of 0.2 mg/L 24 h after chlorine addition to ensure protection against recontamination during transport and storage (CDC 2008). One limitation with these recommendations is how to accurately test FCR in these low-resource environments.

There are several methods available, both titrimetric and colorimetric, for testing FCR and TCR in water. These include iodometric titration, amperometric titration, *N,N*-diethyl-*p*-phenylenediamine (DPD) ferrous titrimetric and DPD colorimetric methods, the syringaldazine (FACTS) method (APHA/AWWA/WEF 2005), and orthotolidine (OTO) method. Colorimetric methods, such as DPD, OTO, and FACTS, are operationally simplest and best-suited to field applications. The OTO colorimetric method for measuring TCR was developed in 1913 and is still used today, although orthotolidine was listed as potentially carcinogenic in 1966 and the method was excluded from 'Standard Methods' beginning in 1980. Introduced in 1957, the DPD method can measure both FCR and TCR and is popular for field testing worldwide (Black & Veatch Corporation 2010). These colorimetric methods utilize a color change to determine residual concentration against a set of visual comparator standards or by use of a colorimeter. OTO uses a color change in the yellow spectrum, while DPD uses a red spectrum.

A variety of commercially available test kits – ranging in accuracy, precision, training necessary to use, and cost – is commonly used to test the presence of TCR and FCR in mg/L. These include colorimeters, color wheel comparator kits, test tube comparator kits, pool test kits, and test strips (Figure 1, Table 1). Colorimeters are hand-held electronic meters that analyze a sample's color intensity by measuring light absorbance at a particular wavelength. Test tube kits, color wheel kits, and pool test kits each require the user to fill a vial with sample, add a reagent, and visually compare the sample color to a standard chart. To use test strips, users submerge a test strip in water sample and visually compare the resultant test strip color to a standard color chart.

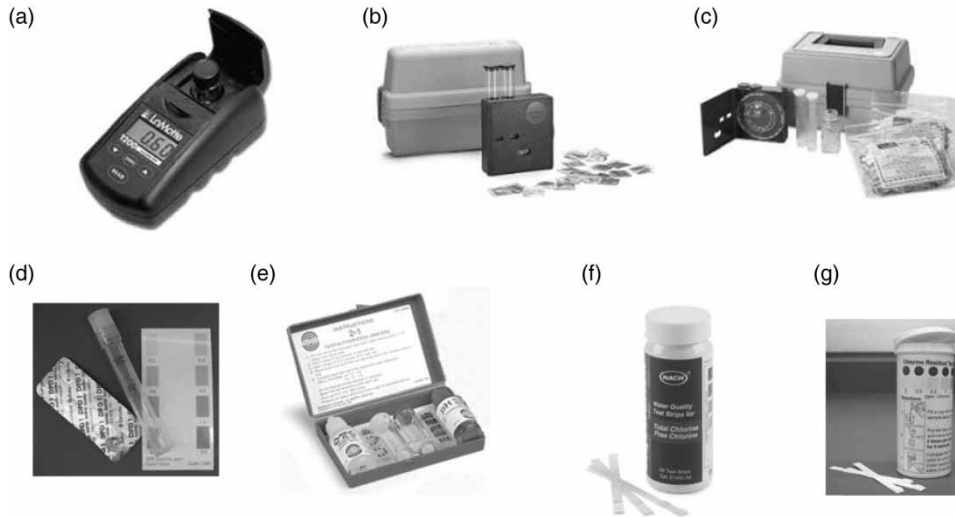


Figure 1 | Seven commercially-available FCR test kits evaluated. (a) LaMotte Colorimeter, model 1200 (www.lamotte.com); (b) Hach Color Wheel Test Kit, model CN-66 (www.hach.com); (c) Hach Color Wheel Test Kit, model CN-70 (www.hach.com); (d) LaMotte Test Tube Kit; (e) Pentair Rainbow Pool Chlorine Test kit (www.pentairpool.com/pdfs/rainbowmaintB.pdf); (f) Hach AquaChek Free and Total Chlorine Test Strips (www.hach.com); (g) Precision Laboratories Very Low Level Chlorine Test Strips.

Field experience in developing countries indicates that users often make errors performing FCR tests or reading results. Some common errors include (Tom Armitage, personal communication, February 26, 2013): (1) the test tube or vial and cap are not rinsed before or after testing, allowing contamination between samples; (2) users do not fully dissolve the reagent before reading the result; (3) test samples are left in tubes/vials, which stain them and render subsequent results inaccurate; (4) the test tube/vial is stored without fully drying and is stained by water droplets; (5) color wheels or comparator charts are left in the sun, thus fading the calibrated standard; (6) incorrect reagent types or quantities are substituted for the correct reagent; (7) users fill the tube/vial to the improper level, which uses an incorrect dilution factor for the test kit reagent; (8) the colorimeter sample vial becomes scratched and yields inaccurate readings; (9) condensation forms on the colorimeter vial with cold water samples, which may affect the colorimeter reading; and (10) users forget to calibrate the colorimeter, which may result in inaccurate readings.

Several studies and reports have provided valuable information comparing chlorine residual test methods (Lishka & McFarren 1971; Wilde 1991; Derrigan *et al.* 1993), and further reviews of such studies have reported on the relative advantages and disadvantages of test methods

(Gordon *et al.* 1987; Harp 2002). Most of these studies, however, focus solely on method accuracy, precision, and measurement interference, primarily on test methods appropriate for laboratory settings. While some evaluations discuss operator training required (Gordon *et al.* 1987; Harp 2002), test method usability has not been quantified, and costs were not discussed. Lastly, the most recent of these studies was conducted over 10 years ago, and as such more recent test methods (such as test strips) were not included in the analysis.

In this research, we evaluated seven FCR and TCR field test methods applicable for use in low-resource settings. In addition to considering accuracy and precision at a variety of sodium hypochlorite (NaOCl) concentrations and light settings, this study includes data on usability and cost.

METHODS

Testing location

The full testing regime was conducted in the Civil and Environmental Engineering laboratory, classroom, and exterior locations on the Tufts University campus in Medford, Massachusetts, USA.

Table 1 | Comparison of seven FCR and TCR test kits used in study

	Measurement range (mg/L)	Measurement increment (mg/L)	Test method	Measure of FCR/TCR	Equipment cost, initial	Incremental test cost	Total cost, 100 tests	Total cost, 1,000 tests
LaMotte Colorimeter, model 1200	0.00–4.00	0.01	DPD instrument grade tablets	FCR/TCR	\$435.00	\$0.09	\$444.40	\$529.00
Hach Color Wheel Test Kit, model CN-66	0.0–3.4	0.2	DPD powder	FCR/TCR	\$55.00	\$0.16	\$71.10	\$216.00
Hach Color Wheel Test Kit, model CN-70	Low range: 0.00–0.68 Mid range: 0.0–3.4	Low range: 0.04 Mid range: 0.2	DPD powder	FCR/TCR	\$72.00	\$0.16	\$88.10	\$253.00
LaMotte Test Tube Kit	0–3.0	0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0	DPD rapid, visual grade tablets	FCR	\$5.85	\$0.09	\$15.25	\$99.85
Pentair Rainbow Pool Chlorine Test Kit	0–3.0	0, 0.3, 1.0, 1.5, 3.0	OTO liquid	TCR	\$14.00	\$0.05	\$19.00	\$64.00
Hach AquaChek Free and Total Chlorine Test Strips	0–10.0	0, 0.5, 1.0, 2.0, 4.0, 10.0	Test strips	FCR/TCR	–	\$0.51	\$30.58	\$305.80
Precision Laboratories Very Low Level Chlorine Test Strips	0–5	0, 0.3, 0.5, 1, 2, 5	Test strips	FCR	–	\$0.04	\$3.50	\$35.00

Test kit selection

Seven commonly used, commercially available chlorine test kits in the United States were selected for comparison in this study; representing colorimeters, color wheel comparator kits, test tube comparator kits, pool test kits, and test strips manufactured by various companies (Figure 1 and Table 1).

Test solution preparation

Eleven sodium hypochlorite (NaOCl) solutions of varying concentration were prepared in the Environmental Sustainability Laboratory at Tufts University. The concentration of NaOCl in Clorox® bleach was verified by Hach 8209 iodometric titration method (APHA/AWWA/WEF 2005). Bleach was added to deionized, chlorine demand-free water in plastic containers to create solutions at the following concentrations (subsequently referred to as Doses 1–11): 0.0, 0.1, 0.2, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 mg/L. Solutions were prepared immediately before testing and discarded at the completion of testing.

Laboratory testing

Each of the seven test kits was used according to the manufacturer's instructions to measure the FCR and/or TCR (depending on test) in each test solution in quintuplicate, and an arithmetic mean was calculated for each set of readings. For test kits that measure both FCR and TCR, only FCR results were analyzed, as sample water was free of chlorine demand.

The Hach color wheel model CN-70 kit was used at both the mid- and low ranges to measure chlorine concentrations in Doses 1–4 (0.0–0.5 mg/L). The remaining doses (1.0–4.0 mg/L) were tested using the mid-range procedure only. The LaMotte colorimeter was calibrated each test day using non-expired calibration solutions at 0, 0.2, 1.0, and 2.5 mg/L. When the instrument read 'Er2', indicating a chlorine concentration above the meter's range, the sample was diluted with deionized water and retested. For tests requiring visual color matching, FCR or TCR values were recorded at the closest matched color shade on the comparator standard; Hach Aquachek test strips and

Hach color wheel kits were read at intermediate increments where applicable, as per manufacturer instructions.

Data were entered into Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA) and analyzed in Excel and R statistical package, version 3.0.1 (R Foundation for Statistical Computing). Percent measurement error was calculated for mean readings at each dose, and a composite mean percent measurement error was calculated for each test kit across all doses. Error was defined as low (<10% measurement error), medium (10–25% error), or high (>25% error). In addition, the percentages of false negative and false positive readings were calculated, considering whether samples and their readings were within the CDC-recommended FCR range of 0.2–2.0 mg/L. False negative readings were defined as a measured value outside the recommended FCR range, where the actual sample FCR was within the range. False positive readings were those within the recommended FCR range, where the water sample FCR was outside the range. Values of 0.15–0.19 mg/L were rounded to 0.2 mg/L. Standard error was calculated across replicate readings at each dose for all test kits. The set of mean FCR measurements across all doses for each test kit was compared to that of a reference method (the LaMotte colorimeter) using a nonparametric test for equality of medians in a paired sample, the Wilcoxon signed rank test, to determine if the results differed at a 0.05 significance level.

Lighting conditions

The full laboratory procedure described above was repeated outside on a sunny day, taking care to store samples out of direct sunlight. Measurement errors from readings in sunlight were compared to those of the laboratory results, considering a measurement error differential threshold of 10%. A Wilcoxon signed rank test was used to compare the mean laboratory results of each test kit to mean outdoor results across all doses at a 0.05 significance level.

Reagent testing

Three test kits that use DPD tablet or powder reagents – Hach color wheel model CN-66, LaMotte test tube kit, and LaMotte colorimeter – were used in the laboratory

with a variety of reagent combinations to test the FCR of three sample doses prepared as described above (0.2, 0.5, and 2.0 mg/L). Nineteen test kit/reagent combinations were evaluated, accounting for potential user errors such as: substituting a different manufacturer's reagent, interchanging instrument-grade and rapid-grade DPD tablets, using the LaMotte DPD #3 (total chlorine) tablet without the first use of DPD #1 (free chlorine) tablet, and correcting for a reagent's volume dilution factor.

Measurements were performed in quintuplicate and an arithmetic mean and measurement error were calculated. Measurement errors were compared to the laboratory testing errors, and test method/reagent combinations were categorized as: (1) effective (measurement error equal to, or lower than, laboratory error); (2) somewhat effective (measurement error higher than laboratory error, but within 25% error); or (3) ineffective (measurement error greater than 25%).

Volunteer testing

Three water samples at 0.2, 0.5, and 2.0 mg/L FCR were prepared using the procedures described above. Eight volunteer participants followed written instructions and used each kit to measure the FCR in all three water samples, with the exception of the Hach color wheel model CN-70 kit in the low range, which was used to test the 0.2 and 0.5 mg/L concentrations only. Each water sample was tested in duplicate with each test kit, for a total of 46 measurements for each volunteer participant. Sample concentrations were unknown to the participants, who were observed and photographed performing the test procedures. Following the water testing, participants completed a questionnaire including: (1) prior laboratory and water testing experience; (2) Likert-scale questions on relative difficulty of test procedures and their confidence in the results; (3) open-ended comments on difficulty of test kit procedures; (4) indication of the easiest kit, which results they were most confident in, and why; and (5) which test kit they would recommend for a variety of contexts and why. Standard error, average measurement error, and false negative readings were calculated for the volunteer test results, as described above. Free and informed consent of the participants was obtained and the study protocol was found to be exempt by the Social, Behavioral, and

Educational Research Institutional Review Board at Tufts University (Protocol #1210007, October 12, 2013).

Cost

Costs for test equipment were calculated by adding fixed equipment and reagent costs from the manufacturers' websites in September 2012. The costs do not include shipping or handling.

Decision matrix

A template decision matrix was developed to determine which FCR test method was most appropriate for testing chlorine-treated drinking water in low-resource settings. Each test method was rated 0, 1, or 2 (where a lower number is more favorable) in each of five categories representing accuracy, usability, and cost. Precision was not included, as precision results were similar for all tests. Accuracy was ranked by the mean composite measurement error across all eleven doses tested, considering laboratory and volunteer test results. Method difficulty and confidence in results were ranked according to the mean reported values from volunteer testing questionnaires on those themes. Cost was ranked on the total equipment and reagent cost for performing 1,000 tests. Values were summed, and the tests were listed according to the total score, where lower score is more favorable.

RESULTS

Laboratory testing

Laboratory test results for FCR and/or TCR are presented in [Figure 2](#) and [Table 2](#). The LaMotte colorimeter was the most accurate in the laboratory across all doses (5.1% average composite measurement error), followed by the Hach color wheel models CN-70 and CN-66, and the LaMotte test tube kit (13.0–14.8% measurement error). The least accurate methods were the Hach AquaChek test strips, Pentair pool kit, and Precision Laboratories Very Low Level test strips (30.6–40.5% measurement error). In looking at frequencies of low, medium, and high measurement error

occurring at each FCR dose ([Table 2](#)), the LaMotte colorimeter and LaMotte test tube kit had the most doses with low measurement error (nine and seven of 11 doses, respectively). The Hach color wheel kits had the highest frequency of readings with medium measurement error (10 of 11 doses for model CN-66 and six of 11 doses for model 70). The Pentair pool kit and Hach AquaChek strips had high measurement error at three FCR doses each, and the Precision Laboratories Very Low Level test strips had high measurement error at seven of 11 doses. Both color wheel methods gave readings consistently above the sample FCR, and the Precision Laboratories test strips gave readings primarily below the sample FCR ([Figure 2](#)).

Considering only the FCR/TCR doses for which a value exists on the comparator standard for those tests that rely on visual color matching, the measurement error decreased substantially for the LaMotte test tube kit (from 14.8 to 1% average measurement error) and Pentair pool test kit (from 34.0 to 0% average measurement error).

The most precise method was the Pentair pool kit (standard error of 0.0 across all doses), and the least precise was the Precision Laboratories test strip method, particularly at high FCR ranges (standard error of 0.73 and 0.60 at Doses 10 and 11, respectively). The remaining test methods had standard errors under 0.20 for all readings ([Figure 2](#)).

The two test strip methods had the highest levels of false positive and false negative readings; with 32% false negative and 13% false positive for the Hach strips, and 20% false negative and 43% false positive for the Precision Laboratories strips. All other test kits exhibited false negative and positive readings at or below 20% ([Table 2](#)).

None of the results from other test kits differed significantly from those of the reference test (LaMotte colorimeter). The tests kits with the highest probability of producing different results (i.e. the least accurate tests) were the Precision Laboratories test strips ($p = 0.054$) and Hach test strips ($p = 0.08$), followed by Hach color wheel CN-66 and CN-70 ($p = 0.18$), Pentair pool kit ($p = 0.52$), and LaMotte test tube kit ($p = 0.64$).

Lighting conditions

The Hach AquaChek test strips were the only visual test that showed greater than 10% average difference between the

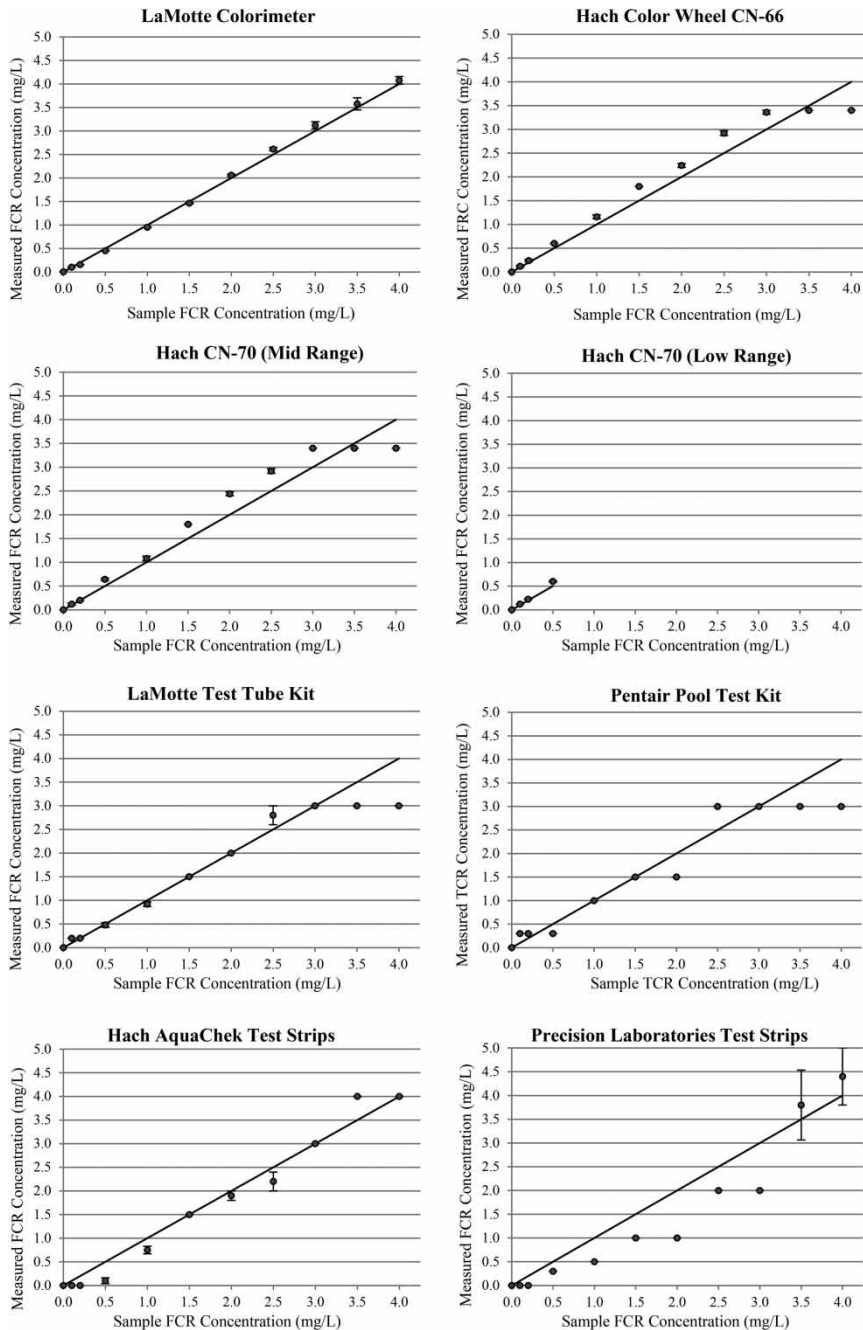


Figure 2 | Average laboratory FCR measurement results with standard error bars for each test kit and chlorine dose. Results in FCR, except pool test kit, which measures total chlorine residual (TCR). Lines represent 'ideal' readings.

values determined under fluorescent light and under direct sunlight, considering all tested doses. FCR readings differed by at least 10% between sunlight and laboratory lighting for four sample doses: 0.5 mg/L (100% measurement error), 1.0 mg/L (53% error), 2.5 mg/L (36% error), 3.5 mg/L

(10% error). Additionally, TCR measurement error was greater than 10% for four doses. None of the test kits, however, showed statistically significantly different FCR measurements between outdoor and laboratory results when considering all doses ($p > 0.05$).

Table 2 | Measurement accuracy of laboratory and volunteer FCR test measurement results

	FCR concentration (mg/L)										Mean error (%)	False negative (%)	False positive (%)	
	0.0	0.1	0.2	0.5	1.0	1.5	2.0	2.5	3.0	3.5				4.0
a. Laboratory testing accuracy														
LaMotte Colorimeter											5.1	20	0	
Hach Color Wheel (CN-66)											14.1	20	3	
Hach Color Wheel (CN-70) Low range											13.0	0	0	
Hach Color Wheel (CN-70) Mid range											13.3	20	3	
LaMotte Test Tube Kit											14.8	0	20	
Pentair Pool Test Kit											34.0	0	17	
Hach AquaChek Test Strips											30.6	32	13	
Precision Labs Very Low Level Test Strips											40.5	20	43	
b. Volunteer Testing Accuracy														
LaMotte Colorimeter	-	-			-	-		-	-	-	-	4.5	17	-
Hach Color Wheel (CN-66)	-	-			-	-		-	-	-	-	21.5	13	-
Hach Color Wheel (CN-70) Low range	-	-			-	-	-	-	-	-	-	14.9	0	-
Hach Color Wheel (CN-70) Mid range	-	-			-	-		-	-	-	-	28.8	17	-
LaMotte Test Tube Kit	-	-			-	-	-	-	-	-	-	6.6	0	-
Pentair Pool Test Kit	-	-			-	-	-	-	-	-	-	34.2	2	-
Hach AquaChek Test Strips	-	-			-	-		-	-	-	-	54.0	44	-
Precision Labs Very Low Level Test Strips	-	-			-	-		-	-	-	-	28.8	21	-

Low Measurement Error: <10%
 Medium Measurement Error: 10–25%
 High Measurement Error: >25%
 Not Tested

Volunteer testing

Eight Tufts University undergraduate and graduate students participated in volunteer testing. All participants (8/8) reported some general laboratory experience, and 88% (7/8) reported prior experience testing for water quality parameters. Of those who had water testing experience, 57% (4/7) self-reported a beginner level, and 43% (3/7) self-reported an intermediate level of experience.

Volunteer testing measurement results are presented in Table 2. The most accurate test kits across three sample doses were the LaMotte colorimeter and LaMotte test tube kit (4.5–6.6% measurement error), followed by the Hach

color wheel kits and Precision Laboratories test strips (14.9–28.8% error). The Pentair pool test kit and Hach AquaChek test strips were least accurate (34.2–54.0% error). The Hach AquaChek test strips had 44% false negative readings, while the remaining kits were at or below 21% (Table 2). The most precise test method was the colorimeter (0.00–0.02 standard error). The least precise were the test strip methods, particularly at the highest FCR concentration, with standard errors of 0.26 and 0.12 for the Precision Laboratories and Hach AquaChek test strips, respectively. Accuracy in the volunteer testing was lower than that of the laboratory testing for the Hach color wheel kits and Hach AquaChek test strips, but higher for

the Precision Laboratories test strips and LaMotte test tube kit. Precision remained the same or decreased in volunteer testing as compared to that of the laboratory testing for all methods but the colorimeter.

Participants found the Precision Laboratories very low level test strips easiest to use, and the Hach color wheel model CN-70 the most difficult (Table 3). Participants judged the easiest test based on it being simple to perform (5/8), quick to complete (3/8), and easy to match the indicator colors (3/8). All participants (100%) wrote that both test strip methods were 'easy', 'simple', or 'straightforward'. Similar comments were written by 75% of volunteers about the LaMotte test tube kit, 63% about the Pentair pool kit, 50% for the Hach color wheel CN-66, 38% for the LaMotte colorimeter, and 25% for the Hach color wheel CN-70 procedure. Eighty-eight percent of respondents (7/8) described the Hach color wheel CN-70 as having 'difficult directions', 'many steps', or being 'difficult to set up', and 75% (6/8) of volunteers thought the LaMotte colorimeter 'took too long', was 'hard or confusing at first', or had 'a lot of instructions and a confusing calibration procedure'.

Participants also commented on the difficulty of matching the sample color to the color comparator standard. Half of respondents (4/8) had difficulty matching color shades for each of the following test methods: Hach AquaChek test strips, LaMotte test tube kit, Pentair pool test kit, and the

Hach color wheel kits, while only 25% (2/8) of respondents had difficulty matching color with the Precision Laboratories very low level test strips.

All participants were most confident in the colorimeter results (Table 3) due to: measurement precision/numerical feedback (5/8), calibrating to a standard (4/8), and trust in technology over human error in judging color (4/8). While the responses were divided regarding test kit preference for developing countries (Table 3), 88% (7/8) people chose a particular test because of its simplicity or ease of use. Thirty-eight percent (3/8) chose a particular test because it was quick to complete, and 38% (3/8) chose because it is precise or seemed to 'give good data'.

Users had more confidence in the tests that they perceived to be more difficult, when the colorimeter results are removed ($R^2 = 0.90$). The average reported confidence level was weakly correlated with average measurement error ($R^2 = 0.44$), where tests with higher confidence have lower error. Average reported difficulty was not associated with average measurement error of the volunteer testing results ($R^2 = 0.14$).

Reagent testing

Of the 19 combinations tested, six were effective, six were somewhat effective, and seven of the combinations were

Table 3 | Summary of volunteer testing questionnaire responses to several questions about test kit usability ($n = 8$)

	Difficulty of test procedure^a Ave (Min, Max, SD)	Confidence in test results^b Ave (Min, Max, SD)	Easiest^{c,d} (Frequency)	Most confident^e (Frequency)	Choice for use in US laboratory^f (Frequency)	Choice for use in developing countries^g (Frequency)
LaMotte Colorimeter	2.6 (1, 5, 1.4)	5.0 (5, 5, 0.0)	1	8	8	1
Hach Color Wheel (CN-66)	3.3 (1, 5, 1.2)	3.1 (2, 4, 0.8)	1	0	0	1
Hach Color Wheel (CN-70)	4.3 (3, 5, 0.9)	3.4 (1, 4, 1.1)	0	0	0	1
LaMotte Test Tube Kit	2.4 (1, 3, 0.7)	2.9 (1, 4, 0.8)	1	0	0	1
Pentair Pool Test Kit	2.4 (1, 3, 0.7)	3.0 (2, 4, 0.8)	0	0	0	0
Hach AquaChek Test Strips	1.5 (1, 2, 0.5)	2.3 (1, 3, 0.9)	1	0	0	1
Precision Laboratories Very Low Level Test Strips	1.1 (1, 2, 0.4)	2.3 (1, 4, 1.0)	5	0	0	3

^aIndicate the difficulty of each test procedure on a scale of 1–5, where 1 is the simplest and 5 is the most difficult.

^bIndicate your confidence level in each test's results on a scale of 1–5, where 1 is least confident and 5 is the most confident.

^cWhich test did you find to be the easiest to perform today?

^dOne respondent chose two responses (Hach test strips and Precision Laboratories test strips).

^eWhich of the test results are you most confident in?

^fIf you were to receive chlorine test results from a laboratory in the United States, which test would you be most confident receiving results from?

^gIf you were to train local people in a developing country to take chlorine measurements and report them to you, which test would you be most confident receiving results from?

ineffective. Substituting a different manufacturer's reagent was effective with the Hach color wheel and the LaMotte colorimeter; however, the LaMotte test tube kit was only somewhat effective with the Hach DPD reagent. Interchanging instrument-grade and rapid-grade tablets was effective for the test tube kit, but the rapid-grade tablets were ineffective with the colorimeter. Using the LaMotte DPD #3 tablet without the first use of DPD #1 tablet was ineffective in all cases. Tests which corrected for the reagent's intended volume were less effective than those which did not alter the test kit volume. Data are available from the authors upon request.

Cost

The total cost for performing 100 and 1,000 tests with each test kit varied from 3.50–444.40 USD for 100 tests and 35–529 USD for 1,000 tests, with the colorimeter being the most expensive and the Precision Laboratories test strips the least expensive (Table 1). Note that at just over 2,000 samples, the Hach AquaChek test strips surpass the colorimeter as the most expensive method.

Decision matrix

A decision matrix ranking each test based on these results is displayed in Table 4. The LaMotte colorimeter and test tube kit scored most favorably, and the Hach AquaChek test strips scored least favorably.

DISCUSSION

Commercially available FCR test kits vary in terms of measurement accuracy, precision, usability, and cost. Accuracy in the laboratory varied between 5 and 40%. The colorimeter was most accurate in both laboratory and volunteer testing and the two test strip methods were least accurate, considering measurement error, statistical analysis, and false positive/false negative readings in both laboratory and volunteer testing. Accuracies of laboratory results were largely comparable to those of volunteer testing; however, some differences are due to variability in operator experience, training, and visual judgment. No test method results were statistically significantly different in different lighting conditions, although the largest average measurement error between results occurred with the Hach AquaChek test strips. Precision did not vary widely between test kits, although the Precision Laboratories test strips were least precise, and the Pentair pool kit was the most precise. Usability was evaluated in terms of difficulty and confidence in results from a group of relatively inexperienced volunteer users. Participants found test strip methods easiest to perform, but were least confident in those results. They found color wheel methods most difficult, and were most confident in the colorimeter results. Overall costs for FCR test kits depend on the number of tests performed, as some test kits have high fixed equipment costs (colorimeter), and some only have a cost per test (test strip methods). For a lifetime of 1,000 tests, costs vary between 35 USD for the Precision Laboratories test strips and 529 USD for the LaMotte colorimeter. In substitute test reagents testing, using a different

Table 4 | Summary of recommended FCR test methods; considering accuracy, usability, and cost. Each category is rated as 0, 1, or 2, where low numbers are more favorable, and test methods are listed in preferential order

	Accuracy, laboratory	Accuracy, volunteer testing	Difficulty, volunteer testing	Confidence, volunteer testing	Cost	Total
LaMotte Colorimeter	0	0	1	0	2	3
LaMotte Test Tube Kit	1	0	1	1	0	3
Precision Laboratories Test Strips	2	1	0	2	0	5
Pentair Pool Test Kit	2	2	1	1	0	6
Hach Color Wheel Test Kit (CN-66)	1	1	2	1	1	6
Hach Color Wheel Test Kit (CN-70)	1	1	2	1	1	6
Hach AquaChek Test Strips	2	2	0	2	1	7

manufacturer's FCR reagent can yield results within the error of measuring with the intended reagent, although it does not in the majority of cases. While not recommended, a substitution could be made when options are limited and concurrent and retrospective tests are completed to show consistent results.

While test kit measurements did not significantly differ from the reference method (0.05 level), the nonparametric test significance levels confirm and correspond with the accuracy evaluated with mean measurement error. Using a significance level of 0.10, both test strip methods, which also exhibited the highest mean error and false positive and negative rates, differed significantly from the reference method.

When considering all the results presented here, a ranking system (Table 4) indicates that a test tube comparator kit or colorimeter is most appropriate for measuring FCR concentrations in drinking water in low-resource settings. While neither method was considered easiest in volunteer testing, both tests ranked well on measurement accuracy and user confidence. The primary difference between these two test methods is equipment cost.

Each of the five categories is weighted equally in this decision matrix; however, category weights, or additional categories, may be applied to reflect particular program priorities or constraints. Researchers and practitioners should understand differing FCR test kit options in order to evaluate trade-offs and make an informed decision about the preferred method for their circumstances. Some factors that may affect this decision-making process are: (1) testing application, including the range of expected FCR readings; (2) the intended use of collected data; (3) the accuracy and precision required; (4) who will be performing the tests and how they will be trained; (5) how many readings will be made; (6) the available budget; and (7) project location in terms of equipment portability and availability of replacement parts or reagents. For example, the decision-maker should evaluate the measurement precision needed in light of equipment cost and the final use of the data. The colorimeter provides a digital reading to the hundredths place, but at a high cost, and the user may only be interested to know if water has an FCR within a wide acceptable range. Additionally, different methods have strengths in different measurement ranges. Both test strip methods were less accurate at low FCR concentrations and had a high rate of false negative readings on the low end of the acceptable FCR

range (Table 2). In the laboratory testing, color wheel methods were also more likely to overestimate FCR, while the Precision Laboratories test strips were more likely to underestimate (Figure 2).

All but one of the evaluated kits relies on the user to judge color intensity of the water sample. With these kits, the researcher found that the colors had a different character under fluorescent lighting versus sunlight, and they were easier to match under sunlight conditions. Despite this, the color intensity was mostly judged equally in both settings. When volunteer test users were asked to comment on the procedures, several expressed difficulty comparing colored water samples to the standards, regardless of the test procedure's simplicity. The volunteers had high confidence in colorimeter results because 'the machine is designed to test differences in color intensity, so it can do it much better than I can'. This suggests that training on visual colorimetric FCR test kits use may benefit from 'eye calibration', where users practice measuring samples with a known FCR. However, it is known that individuals view color shades differently from one another, and individuals may change their color perceptions from one time to another (Culpepper 1970). This is a point of variation for both the results of this study and for general use of this type of test kit.

The limitations of this research include the following: (1) there is potential bias in laboratory measurements due to reliance on subjective visual color matching, and FCR doses being unblinded to the researcher; (2) volunteer testing results should be cautiously interpreted, as all participants were well-educated university students who all reported previous laboratory experience; and (3) as these tests were performed with chlorine-demand-free water, we cannot comment on the efficacy of using total chlorine reagents, and testing was not performed with real waters that may have other interferences affecting FCR measurements. While these limitations are real, the data and recommendations are still valid: several FCR test kits were evaluated over a wide range of chlorine concentrations, accuracy was evaluated by several different measures, and volunteer test participants provided valuable data on measurement accuracy and quantitative information on ease-of-use for users new to FCR testing.

There are, however, opportunities for future research to refine these data and recommendations. Potential future research includes: (1) more rigorous laboratory testing with

real-world waters, considering both FCR and TCR measurements; (2) additional volunteer testing with individuals of various educational backgrounds; (3) evaluating the effect of more extensive user training on the accuracy of readings among various users; and (4) consideration of environmental and/or human toxicological effects of different test reagents.

CONCLUSIONS

Chlorine is used worldwide to directly disinfect stored household drinking water in low-resource settings, and testing for FCR presence is important to ensure chlorine levels fall within recommended guidelines. Users are faced with a variety of commercially available FCR test kits with varying accuracy, precision, usability, and costs. Based on a ranking system developed to rate seven commonly used tests, a test tube comparator kit or colorimeter were shown to be most appropriate for measuring FCR concentrations in drinking water in low-resource settings. This decision matrix considers accuracy by experienced and inexperienced users, ease-of-use, confidence in the results, and cost. Decision-makers may consider additional criteria or weighting schemes to choose an appropriate FCR test to match their individual priorities.

REFERENCES

- APHA/AWWA/WEF 2005 *Standard Methods for the Examination of Water and Wastewater*. 21st edn, American Public Health Association/American Water Works Association/Water Environmental Federation, Washington, DC, USA.
- Black & Veatch Corporation 2010 *White's Handbook of Chlorination and Alternative Disinfectants*. 5th edn, John Wiley & Sons, Inc., Hoboken, NJ, USA.
- CDC 2008 *Chlorine Residual Testing Fact Sheet*. Centers for Disease Control and Prevention, Atlanta, GA, USA. www.cdc.gov/safewater/publications_pages/chlorineresidual.pdf (accessed 21 September 2012).
- Culpepper, W. D. 1970 *A comparative study of shade matching procedures*. *J. Prosthet. Dentist.* **24** (2), 166–173.
- Cutler, D. & Miller, G. 2005 *The role of public health improvements in health advances: the twentieth-century United States*. *Demography* **42** (1), 1–22.
- Derrigan, J., Lin, L. & Jensen, J. 1993 *Comparison of free and total chlorine measurement methods in municipal wastewaters*. *Water Environ. Res.* **65** (3), 205–212.
- Gordon, G., Cooper, W. J., Rice, G. & Pacey, G. E. 1987 *Disinfectant Residual Measurement Methods*. AWWA Research Foundation, Denver, CO.
- Harp, D. L. 2002 *Current Technology of Chlorine Analysis for Water and Wastewater. Technical Information Series – Booklet No. 17*. Hach Company, Loveland, CO. www.hach.com/cms-portals/hach_com/cms/documents/pdf/LIT/L7019-ChlorineAnalysis.pdf.
- Lishka, E. F. 1971 *Water Chlorine (Residual) No. 2. Report No. 40*. United States Environmental Protection Agency Office of Water Programs, Cincinnati, Ohio.
- USEPA 2006 *List of Drinking Water Contaminants and MCLs: National Primary Drinking Water Regulations*. United States Environmental Protection Agency, Washington, DC, USA. <http://www.epa.gov/safewater/mcl.html> (accessed 13 June 2013).
- WHO 2011 *Guidelines for Drinking-Water Quality*. 4th edn, WHO chronicle. World Health Organization, Geneva, Switzerland.
- WHO n.d. *Environmental Sanitation Fact Sheet 2.17: Chlorination Concepts*. World Health Organization, Geneva, Switzerland. www.who.int/water_sanitation_health/hygiene/emergencies/fs2_17.pdf (accessed 13 June 2013).
- Wilde, E. W. 1991 *Comparison of three methods for measuring residual chlorine*. *Water Res.* **25** (10), 1303–1305.

First received 7 October 2013; accepted in revised form 12 June 2014. Available online 1 July 2014