

## Comparison of Methods for Assessing Quality of Care for Community Case Management of Sick Children: An Application with Community Health Workers in Malawi

Cristina V. Cardemil,\* Kate E. Gilroy, Jennifer A. Callaghan-Koru, Humphreys Nsona, and Jennifer Bryce

Department of Pediatrics, Children's National Medical Center, Washington, District of Columbia; Institute for International Programs, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; Integrated Management of Childhood Illnesses Unit, Ministry of Health, Lilongwe, Malawi

**Abstract.** Direct observation (DO) with re-examination (RE) by a skilled clinician is a rigorous method for assessing health worker performance, but is not always feasible. We assessed the performance of 131 community health workers in Malawi in community case management of sick children with cough and fast breathing, fever, and diarrhea. We compared estimates of correct treatment measured through DO with RE ( $n = 382$  cases) to DO only ( $n = 382$  cases), register review ( $n = 1,219$  cases), and case scenarios ( $n = 917$  cases). Estimates of correct treatment of uncomplicated fever and diarrhea measured through DO only, register review, and case scenarios were within 9 percentage points of DO with RE estimates, while estimates for uncomplicated cough and fast breathing, and severe illness were substantially higher than DO with RE (12–51 percentage points above the estimate). Those planning for community health worker assessments in community case management can use these results to make an informed choice of methods on the basis of their objectives and the local context.

### INTRODUCTION

Programs to train community health workers (CHWs) in resource-poor settings to deliver life-saving interventions to sick children have grown in number and scale, increasing the importance of assessing the quality of the care they deliver.<sup>1,2</sup> The assessment methods used to measure quality of care should be valid, feasible, and an accurate reflection of performance of CHWs.<sup>3</sup> However, few studies have examined the methods used to assess quality of health worker performance, and most previous studies focused on facility-level care.<sup>3–6</sup> Assessments of CHW performance pose unique challenges because of their remote locations, diverse responsibilities, and variable caseloads,<sup>7</sup> and it remains unclear which method provides the optimal balance between data validity, reliability and feasibility of implementation at the community level.

Methods that have been used to assess health workers' performance in community case management (CCM) of sick children delivered by CHWs are summarized in Table 1. Direct observation (DO) with re-examination (RE) by a higher-level trained clinician is a common and well-regarded method to assess clinical performance of health workers at first-level health facilities.<sup>3,5,6</sup> Direct observation enables assessment of clinical cases by the CHW in the presence of a silent observer, and DO with RE involves independent verification of the case by a skilled clinician. One disadvantage of DO is the influence of an observer, known as the Hawthorne effect,<sup>8,9</sup> which can introduce a positive bias in quality of care assessments.<sup>7,10,11</sup>

Although DO with or without RE can be feasible at facilities, assessment of a CHW stationed hours or days in travel time from a referral center can be time-consuming, expensive, and logistically challenging, especially because in many settings CHWs have responsibilities that require leaving their post.<sup>12</sup> Caseloads seen by each CHW can also vary widely<sup>7</sup> depending on disease burden, catchment area size, geographic barriers, and other factors; in low caseload settings, DO can be resource intensive. As a result, few studies to date have

attempted DO or RE in the community.<sup>13–18</sup> Instead, many choose to bring CHWs to a central location, such as a district hospital, for DO or role-play assessments.<sup>7,19–24</sup> However, removing CHWs from their home environment changes the physical space in which they work (e.g., lighting and facility structure) and the resources they have at hand (e.g., timers, drug supply), removes the interaction with the caretaker as a respected member of the community, and may bias the case mix towards more severely ill children.

Many researchers have avoided these practical problems by using other assessment methods, such as reviewing clinical records or registers (register review [RR]) or assessing CHW knowledge using case scenarios (CS) that present a vignette, followed by a series of questions.<sup>7,13,14,19,21,25</sup> Register review is less time-consuming and resource-intensive than DO<sup>5</sup> and could be an ideal proxy for clinical actions, but records of CHWs are often incomplete, subject to reporting bias (e.g., CHWs may only record uncomplicated illnesses they can manage and not record others), and may only be possible in settings with high CHW literacy. Case scenarios can measure CHW knowledge and cover selected illnesses representative of those in the community, including rare cases that are difficult to observe directly, but do not document performance in real clinical cases.

As part of an assessment of quality of CCM services in Malawi,<sup>26</sup> we measured CHW performance in classification and treatment of cough and fast breathing (suspected pneumonia), fever (suspected malaria), and diarrhea cases using various measurement methods. This report examines the bias associated with measuring CHW performance in CCM by using RR, CS, and DO only methods compared with direct observation with re-examination (DO with RE) by a higher-level clinician, and discusses the relative strengths and weaknesses of the four assessment methods in the Malawi CCM context.

### MATERIALS AND METHODS

**Assessment tool development.** We adapted detailed checklists for DO and RE from the World Health Organization (WHO) Health Facility Survey<sup>27</sup> to include required CCM

\*Address correspondence to Cristina V. Cardemil, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333. E-mail: ccardemil@cdc.gov

TABLE 1

Methods used to assess community health worker technical performance in case management of sick children in the community, Malawi\*

Method	Description	Advantages	Disadvantages
Observation of clinical encounter with re-examination	In person, silent observation of real case by surveyor with independent re-examination by higher level clinician	Complete picture of provider's actions in real setting, including steps in case management process where errors occur; collection of real-time data	Intrusive; observer influence (Hawthorne effect); time consuming, complex, and generally more expensive because of need for careful training and clinical staff; usually cross-sectional, trend analysis less feasible
Observation of clinical encounter with actor simulation	In person, observation of simulated case played by an actor unbeknownst to CHW	Complete picture of provider's actions in real setting, including steps in case management process where errors occur; reduction of Hawthorne effect compared with known observation	Usually cross-sectional, trend analysis less feasible; time consuming, complex, and generally more expensive because of need for careful training of actors/observers; may be difficult in CCM programs where CHWs know their patients from the community
Observation of clinical encounter	In person, silent observation of real case by surveyor	Complete picture of provider's actions in real setting, can be ideal for actions required regardless of classification, such as counseling	Careful training needed for observers; no independent re-examination makes it difficult to determine ideal, gold-standard management actions
Register review	Review of register or other records of individual cases and summaries	Review large number of cases relatively quickly; able to perform in any setting; inexpensive, can be performed by non-clinical staff at different points in time for trend analysis	Quality and quantity of documentation varies; data subject to health worker accuracy in reporting
Provider knowledge testing	Test of knowledge using 1) oral or written examination; 2) case scenarios/vignettes; or 3) scenario using video/audio recording of case	Assessment of knowledge relatively quickly; transferable between settings (can adjust questions/cases to reflect local case mix); focus on severe disease or broaden scope to include more common cases; can take place in any setting (community vs. facility based)	Questions and cases must accurately reflect child illnesses; measure of health worker knowledge, which may be different than practice
Exit interview with child's caretaker	Oral or written questionnaire of health worker performance	Determination of patient satisfaction; allows collection of additional data (such as costs incurred with visit or understanding of counseling messages)	Subject to recall bias, especially depending on timing of interview after consultation; tendency to over-report tasks
Provider self report	Questions/checklist of items performed on a routine basis	Ability to conduct in any setting; inexpensive and simple to administer	Biased assessment of one's own performance; potential for large bias means it may only be able to identify large performance gaps
Re-examination of child (without observation)	Gold-standard clinician re-examines child after CHW	Allows for accurate determination of child's clinical status for signs and symptoms that have not changed since the CHW consultation	Examination can change dramatically from initial presentation to CHW, especially if treatment is administered in the interim

\*CHW = community health worker; CCM = community case management.

tasks according to the WHO manual *Caring for the Sick Child in the Community*.<sup>28</sup> A standard form for RR abstracted the child's age, sex, classification, treatment and referral decision from CCM patient registers. At the time of the assessment, Malawi's CHWs (known as Health Surveillance Assistants [HSAs]) were using ad hoc registers; often these were old diarrhea management registers that included the name, age and sex of child, classification and management decisions. Eight case scenarios were designed based on the WHO Health Facility Survey, reviewed by independent CCM experts, and adapted to reflect common and rare clinical illness expected in children in the community in Malawi (five scenarios required

referral, three could be managed by the HSA in the community, three included diarrhea, three included cough and fast breathing, and three included fever). Copies of all assessment forms are available at [www.jhsph.edu/dept/ih/IIP/projects/catalyticinitiative.html](http://www.jhsph.edu/dept/ih/IIP/projects/catalyticinitiative.html).<sup>29</sup> All data collection tools were piloted over a one-week period and revised accordingly.

**Study setting.** In 2008, Malawi's Ministry of Health started to train and equip a cadre of community-based health workers (HSAs), to provide CCM for sick children 2 months up to 5 years of age by using the WHO strategy in which CHWs assess symptoms, briefly examine the patient, classify and treat disease, and refer complex cases to a health facility.<sup>28</sup> The HSAs

are a paid cadre of CHWs who are required to have a 10th grade education and work in communities in water and sanitation and other programs such as child vaccination, tuberculosis treatment, family planning, or assisting at a health facility.<sup>30</sup> The CCM six-day training course, comprising three days of clinical practice, trained HSAs to treat cough and fast breathing (suspected pneumonia) with cotrimoxazole, fever (suspected malaria) with artemether/lumefantrine and paracetamol, and diarrhea with oral rehydration salts. For more complicated cases, HSAs were taught to recognize signs and symptoms and refer to the nearest health facility.<sup>30</sup> The Ministry of Health CCM strategy first targeted the training and deployment of HSAs to serve hard-to-reach communities (> 8 km from a health facility). The HSAs treat sick children in their communities at a village health clinic site designated by the community, often a central structure or outdoor place in the village, or sometimes located at the home of an HSA. The caseload of HSAs is relatively high; an average of > 40 sick children are seen per month.<sup>26,30</sup>

The HSAs were randomly sampled from a listing of CCM-trained HSAs with available drug supply and serving in hard-to-reach areas in 6 of the 28 districts in Malawi with early CCM implementation, resulting in a total sample of 132 HSAs. A detailed description of the sampling methods has been reported.<sup>26</sup>

**Surveyor training and data collection.** We trained surveyors, nurses, or clinical officers who were CCM master trainers in the use of the tools and methods for one week. Inter-observer reliability measurements were made for the direct observation tool during the initial training period and halfway through the assessment, continuing until 90% agreement was reached for all surveyors. Teams of three surveyors visited HSAs at their village health clinics in their communities and collected data for five weeks during October–November 2009.

The teams spent one day for each HSA visit, and carried out the following steps. First, the survey team obtained written consent from the HSA for all data collection methods. Second, as caretakers brought sick children to the HSA, typically early morning, surveyor 1 identified eligible children (sick children 2–59 months of age presenting for initial consultation) and obtained oral consent from their caretakers for DO and RE. Direct observation and RE were conducted for ≤ 5 sick children per HSA. Third, surveyor 2 observed sick child consultations and recorded whether specific steps in the CCM algorithm were carried out for each child. The classification of the HSA was determined by what the HSA wrote in the register; in instances where this was unclear, surveyors were instructed to ask the HSA for the classification. Treatments were determined by observing those given to the sick child. Fourth, without witnessing the initial encounter between the HSA and the sick child, surveyor 3 (also the team leader) re-examined the sick child privately and recorded his or her gold-standard assessment. Fifth, after all sick children had been seen by the HSA, surveyor 3, with help from the other surveyors, abstracted required data from the clinical registers for the last 10 cases seen by the HSA before the assessment. Surveyor 3 also read eight case scenarios to each HSA, who held a written copy. After each scenario, the HSA was asked open ended questions about what he or she would do for the child. The scenarios were administered without prompting and answers were recorded on a checklist of possible responses including HSA treatment and referral

decisions. The HSAs were allowed to use any job aids available at their site to formulate their responses.

**Definitions.** We focused our analyses on three classifications (cough with fast breathing [suspected pneumonia and referred to as fast breathing throughout this report], fever [suspected malaria], and diarrhea) because collectively fast breathing, fever and diarrhea formed most clinical diagnoses seen by HSAs<sup>26</sup> and together are estimated to account for 45% of all deaths in children 1–59 months of age in Malawi.<sup>31</sup> We divided these cases into uncomplicated illness and severe illness, the latter defined as any illness with danger signs (Figure 1).

**Measures and statistical analyses.** Measures used to define correct classification and treatment of each method are shown in Table 2. These same methods were used to calculate misuse of antibiotics and antimalarial drugs; incorrect performance was defined as giving an antimicrobial drug in cases in which treatment was not indicated. In this analysis, we excluded cases for DO with RE and DO only in which the HSA was experiencing a stock-out of the drug needed to treat a classification correctly. Children who were not able to tolerate oral medications (i.e., vomiting everything or not able to drink or eat anything) were also excluded from analysis because the CCM guidelines recommend against treatment with drugs in these cases. We calculated correct treatment across the methods for all HSAs and cases and for a restricted sample of HSAs that had managed fast breathing, fever, or diarrhea during the DO. We present the former here because we found no appreciable differences in the results produced.

We considered the DO with RE as our gold-standard method. We first examined the classification and treatment decisions by the HSAs collected by DO compared with the independent re-examiner for uncomplicated illness and severe illness and report sensitivity and specificity estimates. Agreement between DO and RE was assessed by percent agreement and kappa statistics, with values < 0, 0.00–0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.00 considered as poor, slight, fair, moderate, substantial, and almost perfect agreement, respectively.<sup>31</sup> We calculated point estimates of the proportion of cases correctly treated as measured by DO only, RR and CS for uncomplicated and severe illness using the measurement definitions above. We then compared estimates of correct treatment measured by DO only, RR (historical cases) and CS (fictional cases) to cases with the same classifications as measured by DO with RE, calculating the arithmetic difference in point estimates and testing for significant differences by using chi-square or Fisher's exact, as appropriate.

Proportions, 95% confidence intervals, and *P* values were calculated by using Stata version 11 (StataCorp LP, College Station, TX) with adjustments for HSA clustering.<sup>32</sup> Differences were considered statistically significant if *P* < 0.05. The Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health and the Malawi National Health Sciences Review Committee provided ethical review and approval for the study.

## RESULTS

A total sample of 131 HSAs was visited; 1 HSA could not be found by the survey team and was dropped from the sample. One HSA did not consent to participate in the study; this HSA and seven others who did not have initial drugs stocks (3) or were absent (4) were replaced during the assessment by

### Definitions

**Uncomplicated illness** includes cough and fast breathing (referred to throughout as fast breathing), diarrhea or fever without any danger signs.

**Severe illness** includes fast breathing, diarrhea or fever with one or more danger signs requiring referral according to community case management (CCM) guidelines. Danger signs included chest indrawing, cough for  $\geq 21$  days, fever for  $\geq 7$  days, diarrhea for  $\geq 14$  days, blood in stool, inability to drink or feed, vomits everything, convulsions, palmar pallor, red on mid-upper arm circumference (MUAC) tape, or bipedal edema.

**Classification** is the presumptive illness assigned to the child after examination. Fast breathing, diarrhea and fever are included in this analysis.

**Correct classification** is identification or recording by the Health Surveillance Assistant (HSA) of the same classification as recorded by the re-examiner, and is based on the application of the Malawi adaptation of the World Health Organization CCM guidelines.

**Correct treatment<sup>a</sup>** for uncomplicated illness is prescription of the appropriate medicine by the HSA according to CCM guidelines for the classification. Children with cough and fast breathing are treated with cotrimoxazole, children with fever are treated with antimalarials, and children with diarrhea are treated with oral rehydration salts.<sup>b</sup> Correct treatment for severe illness is defined as correct treatment by the HSA for the classification, as well as appropriate referral to a health facility, based on the CCM guidelines. Referral is required for any danger signs reported by the caretaker or identified by the HSA.

**Misuse of antibiotics or antimalarials** refers to a medication given to a child by the HSA that is not indicated in CCM guidelines for that classification.

#### Sensitivity and specificity

##### Classification of Illness

**Sensitivity:** proportion of children who truly had a classification (e.g., fever) as defined by the gold-standard re-examiner that were classified as such by the HSA.

**Specificity:** proportion of children who did not have the classification (e.g., did not have fever) who were correctly identified as not having that classification by the HSA.

##### Treatment of Uncomplicated Illness

**Sensitivity:** proportion of children who truly had a classification as defined by the gold-standard re-examiner who were treated correctly by the HSA.

**Specificity:** proportion of children who did not have the classification as defined by the gold standard re-examiner who were not treated by the HSA as if they did have that classification (e.g., child does not have fever, HSA did not give Antimalarial Combination Therapies (ACTs)).

##### Treatment of Severe Illness

**Sensitivity:** proportion of children who truly had an illness (fast breathing, fever, or diarrhea) with danger sign as defined by the gold-standard re-examiner who were treated correctly and referred for that illness by the HSA.

**Specificity:** proportion of children who did not have the illness (fast breathing, fever, or diarrhea), but did have a danger sign, as defined by the gold standard re-examiner, who were not treated by the HSA with a drug as if they had that illness (fast breathing, fever, or diarrhea), but were referred.

<sup>a</sup> Dose, frequency, and duration were not included in this analysis because this information was not available from register reviews or case scenarios.

<sup>b</sup> Zinc is also included as treatment for diarrhea in CCM guidelines, but was not available for any HSAs at the time of this assessment.

FIGURE 1. Definitions used in comparison of methods for assessing quality of care for community case management of sick children, Malawi.

TABLE 2  
Measures used to define correct classification and treatment of each method, Malawi\*

Measure	HSA performance measure	Standard for correct performance
Direct observation with re-examination (DO with RE)	Classification given by HSA as directly observed; treatment given by HSA as directly observed	RE classification; treatment recommended in CCM guidelines, according to RE classification
Direct observation (DO) only	Treatment given by HSA as directly observed	Treatment recommended in CCM guidelines, given HSA's classification observed
Register review (RR)	Treatment given by HSA as recorded in register	Treatment recommended in CCM guidelines, given HSA's classification recorded in register
Case scenarios (CS)	Treatment recommended by HSA in response to CS	Treatment recommended in CCM guidelines, given signs and symptoms of child included in CS

\*HSA = health surveillance assistant; DO = direct observation; RE = re-examination; CCM = community case management; CS = case scenarios.

using random selection. Data from 382 sick child consultations with DO and RE were available for analysis after exclusion of six cases that did not meet eligibility criteria or had incomplete documentation. All caretakers approached to participate in the study consented. Surveyors directly observed and re-examined a median of 3 (range = 1–5) sick child consultations for each of 131 HSAs. Of 1,310 RR sick child cases recorded, 91 (6.9%) were excluded because they lacked information on the classification assigned by the HSA. Two of the eight scenarios administered to all HSAs were excluded from this analysis because they lacked analogous cases in RE, resulting in a total of 262 suspected pneumonia, 262 fever, and 393 diarrhea cases.

The percentage of cases by type of classified illness included in our sample for each method is shown in Table 3. Fever was the most common (59–75%) uncomplicated illness in DO with RE, DO only, and RR. A higher percentage of uncomplicated fast breathing cases were recorded in RR (32%) than observed in DO only (21%) or DO with RE (15%). Many fewer severe illnesses (with danger signs) were recorded in RR than observed in DO only or DO with RE. The most common danger signs classified by RE were fever for  $\geq 7$  days, palmar pallor, and blood in the stool. Most cases of uncomplicated fast breathing, fever, or diarrhea seen were

classified with one (74–92%) or two of these illnesses (28–34%). A smaller proportion of cases were recorded as referred in RR (1–3%) than observed in DO only (19–22%) or DO with RE (19–21%).

**Sensitivity, specificity, percent agreement, and kappa.** The sensitivity, specificity, percent agreement, and kappa for HSA classification and treatment through DO relative to RE are shown in Table 4. For uncomplicated fever and diarrhea cases, sensitivity and specificity of HSA's classification and treatment (DO) compared with RE were high (87–99%), and kappa estimates indicated substantial or almost perfect agreement (0.70–0.92). Sensitivity was lower for classification and treatment of uncomplicated fast breathing (59% and 63%), as was specificity (82% and 75%). In further exploration, a contributing factor to the low sensitivity of classification and treatment of fast breathing cases was co-morbid fever and/or diarrhea. The HSAs more frequently missed the fast-breathing classification (18 of 24 cases incorrectly classified) or giving cotrimoxazole (19 of 21 cases with incorrect treatment) in the presence of other illnesses. For treatment of cases that did not have fast breathing, most errors leading to low specificity (46 of 57) were caused by incorrect classification with fast breathing.

For severe illness, sensitivity and specificity of HSA classification of fever with danger signs was highest (86% and 94%, respectively) (Table 4); sensitivity of fast breathing with danger signs (68%) and diarrhea with danger signs (57%) were lower. For treatment of severe illnesses, sensitivity and specificity were low for all three illnesses, ranging from 32% sensitivity for fast breathing to 65% for specificity for fever. Percent agreement and kappa estimates correlated with sensitivity and specificity estimates. Errors for fast breathing cases with danger signs were often caused by lack of referral (9 of 13) and or lack of treatment with cotrimoxazole (6 of 13). For fever cases with danger signs, 23 of 30 were not referred and 11 of 30 were not treated with artemether/lumefantrine.

**Correct treatment of illness: comparisons of methods.** Correct treatment of cases as measured by DO only, RR, and CS compared with DO with RE is shown in Table 5. For uncomplicated fast breathing, the proportion of cases with correct treatment as measured by DO only (100%), RR (97%), and CS (85%) were all more than 20 percentage points significantly higher than that recorded by DO with RE (63%). Measured levels of correct treatment by HSAs for uncomplicated fever and diarrhea cases in DO, RR, and CS were all within 9 percentage points of the DO with RE estimate. The aggregate indicator, correct treatment of fast breathing, fever and/or diarrhea, was different for DO only and RR compared with the gold-standard DO with RE (79%).

TABLE 3

Percentage cases by type of illness and referral, by measurement method, Malawi\*

Illness or referral	Direct observation with re-examination (n = 382)		Direct observation only (n = 382)		Register review (n = 1,219)	
	%	95% CI	%	95% CI	%	95% CI
<b>Uncomplicated illness</b>						
Fast breathing	15	11–19	21	16–27	32	27–37
Fever	63	58–69	59	53–65	75	72–79
Diarrhea	25	21–29	23	19–27	19	16–22
<b>Severe illness</b>						
Fast breathing	5	3–7	6	3–8	0.2	0–0.4
Fever	13	10–17	12	8–16	0.5	0–1
Diarrhea	4	2–6	4	2–6	0.1	0–0.2
<b>No. uncomplicated illnesses†</b>						
1	74	69–78	75	70–80	92	89–99
2	29	24–33	28	23–33	34	29–38
3	3	1–5	7	0.3–2.8	1	0.5–2
<b>Referrals‡</b>						
Fast breathing	20	71–90	19	10–27	1	0–3
Fever	21	15–26	19	14–25	2	0.3–4
Diarrhea	19	12–27	22	14–31	3	0.6–6

\*CI = confidence interval. Case scenarios were designed by investigators as described in the text and included cases of fast breathing, fever, and diarrhea.

†One classification includes either fast breathing, fever, or diarrhea. Two is any combination of two of these illnesses per child; three is all three illnesses. Danger signs excluded.

‡Children referred for any reason.

TABLE 4

Sensitivity specificity, percent agreement, and kappa for direct observation of health surveillance assistant classification and treatment of illness compared with re-examination, Malawi\*

Illness		Sensitivity		Specificity		% Agreement†	Kappa statistic (95% CI)
		No.	% (95% CI)	No.	% (95% CI)		
<b>Uncomplicated illness</b>							
Fast breathing	Classification	34/58	59 (46–72)	209/256	82 (75–88)	77	0.35 (0.23–0.47)
	Treatment	36/57	63 (50–77)	173/230	75 (68–83)	73	0.31 (0.19–0.43)
Fever	Classification	226/242	93 (90–97)	71/72	99 (96–100)	95	0.86 (0.79–0.92)
	Treatment	198/229	87 (81–92)	52/53	98 (94–100)	89	0.70 (0.60–0.79)
Diarrhea	Classification	86/94	92 (86–97)	218/220	99 (98–100)	97	0.92 (0.88–0.97)
	Treatment	63/70	90 (82–98)	149/150	99 (98–100)	96	0.91 (0.86–0.97)
<b>Severe illness</b>							
Fast breathing	Classification	13/19	68 (41–95)	42/49	86 (75–96)	81	0.53 (0.31–0.76)
	Treatment	6/19	32 (3–60)	23/45	51 (35–67)	45	–0.15 (–0.38 to 0.07)
Fever	Classification	44/51	86 (76–96)	16/17	94 (82–100)	88	0.72 (0.54–0.90)
	Treatment	17/47	36 (22–51)	11/17	65 (38–91)	44	0.01 (–0.18 to 0.19)
Diarrhea	Classification	8/14	57 (26–89)	46/54	85 (74–96)	79	0.40 (0.14–0.66)
	Treatment	5/10	50 (12–88)	18/43	42 (23–61)	43	–0.05 (–0.24 to 0.15)

\*CI = confidence interval. See Figure 1 for definitions of sensitivity and specificity for classification and treatment.

†Percent agreement is a weighted average of sensitivity and specificity. For example, 77% agreement for uncomplicated fast breathing (first row) was calculated as  $(34 + 209)/(58 + 256)$ .

For fast breathing, fever, or diarrhea with danger signs, DO only (56–83%) and CS (37–70%) showed higher levels of correct treatment than DO with RE (24–40%) (Table 5). As the number of uncomplicated illness cases treated by HSAs increased from one to two to three per child, correct treatment by HSAs significantly decreased as measured by DO with RE, RR, and CS (one illness = 79–96%, two illnesses = 37–80%, and three illnesses = 11–60%). The proportion of cases managed correctly as measured by DO only did not decrease in the same manner and always remained  $\geq 89\%$ .

**Misuse of antibiotics and antimalarial drugs.** The proportion of cases that should not have received antibiotics or antimalarial drugs is shown in Table 6. The DO only, RR, and CS methods all showed lower proportions of cases without fast breathing that received cotrimoxazole compared with DO with RE; i.e., misuse of this antibiotic was recorded as lower (2–18% for DO only, RR, and CS than for DO with RE (24%). A high percentage of cases with cough but not fast breathing received cotrimoxazole, as measured by the DO with RE (48%). RR measured an even higher level of misuse of antibiotics in the cases with cough (73% and +25 points from DO with RE), and DO only measured a lower proportion (16%); both were significantly different from the DO with RE.

## DISCUSSION

We found that directly observed HSA classification and treatment of uncomplicated fever and diarrhea had high validity and reliability compared with RE, as indicated by high sensitivity, specificity, and kappa estimates. The DO only, RR, and CS methods provided estimates of the proportion of children correctly treated for uncomplicated fever and diarrhea that were similar to those from DO with RE (within nine absolute percentage points). In contrast, directly observed HSA classification and treatment of uncomplicated fast breathing showed lower sensitivity, specificity, and fair kappa estimates compared with RE. The DO only, RR, and CS methods also estimated significantly higher correct treatment of uncomplicated fast breathing compared with DO with RE. The same trend was true for HSA performance in severe illness

cases for fast breathing, fever, and diarrhea, in which HSAs directly observed treatment had particularly low sensitivity, specificity, and kappa as compared with RE, and the proportion of correctly treated cases in DO only and CS were higher compared with DO with RE. Collectively, these results suggest that HSA performance for uncomplicated fever and diarrhea cases, but not fast breathing or severe illness, can be accurately reported through assessments using DO only, RR, or CS in the Malawi CCM context.

Using DO with RE enabled exploration of errors made by HSAs through an in-person, direct assessment of HSA performance in real-world clinical cases. We identified that most classification and treatment errors for uncomplicated fast breathing in DO were associated with a co-classification of fever, and most errors in severe illness cases were caused by a lack of appropriate referral. In addition, DO with RE and RR identified significant misuse of antibiotics. These errors would have been missed if only using the HSA classifications seen in DO, without the independent RE classifications.

Despite these advantages, DO with RE has been used in few published CHW assessments,<sup>15,22–24</sup> perhaps because it is resource intensive.<sup>5,7</sup> In our assessment, it required removal of clinicians from regular duties for a six-week period. In settings with chronic health worker shortages, this limitation poses a temporary drain on resources. The DO method is also subject to the Hawthorne effect.<sup>6,8–11</sup> Despite the limitations of DO with or without RE, ideally the long-term benefits from valid and reliable information gained from DO with RE on program improvements will outweigh the upfront costs.

Register review enabled us to retrospectively analyze data from a larger number of cases than was possible through DO. Register review identified a high proportion of HSAs giving cotrimoxazole inappropriately for children with cough but no fast breathing (73%). However, DO with RE also indicated a lower, but still high proportion (48%) of HSAs making this error, and DO only did not (16%). In addition, substantially more fast breathing cases were recorded and abstracted in RR than were directly observed. Before introduction of the WHO CCM algorithm, guidelines in Malawi indicated that giving antibiotics for cough was appropriate. Although CCM guidelines no longer endorse this policy,

TABLE 5  
Correct treatment of illness, by method, Malawi\*

Feature	Uncomplicated illness, correct treatment				Severe illness, correct treatment				No. uncomplicated illness classifications†		
	Fast breathing	Fever	Diarrhea	Fast breathing, fever, and/or diarrhea	Fast breathing	Fever	Diarrhea	Fast breathing, fever, and/or diarrhea	One	Two	Three
Direct observation with re-examination	No. 57 % correct 63 (49–76) (95% CI)	229 87 (80–91)	70 90 (79–96)	265 79 (73–84)	17 24 (0.06–58)	44 32 (20–46)	10 40 (13–76)	50 26 (15–40)	167 89 (80–94)	71 65 (53–75)	9 11 (0.01–63)
Direct observation only	No. 88 % correct 100 (95% CI) Difference in % points from DO with RE	229 92 (87–95) +5	66 97 (88–99) +7	287 93 (88–96) +14	8 75 (19–98) +51	27 56 (35–74) +24	12 83 (44–97) +43	35 63 (43–79) +37	164 94 (87–97) +5	79 89 (76–95) –24	4 100 +89
Register review	No. 384 % correct 97 (94–98) (95% CI) Difference in % points from DO with RE	898 93 (89–95) +6	225 82 (74–88) –8	1098 89 (86–92) +10	– – –	– – –	– – –	– – –	688 96 (93–97) +7	387 80 (73–85) +15	15 60 (33–82) +49
Case scenarios	No. 131 % correct 85 (79–91) (95% CI) Difference in % points from DO with RE	131 95 (90–98) +8	131 99 (94–99.6) +9	131 79§ (72–96) 0	131 37 (4–29) +13	131 44 (35–52) +12	131 70 (61–78) +30	131 25§ (18–33) –1	131 79§ (72–96) –10	131 37 (29–46) –28	–
	$P$ ‡	0.003	0.008	0.973	0.323	0.132	0.057	0.912	< 0.001	< 0.001	< 0.001

\* CI = confidence interval; DO = direct observation; RE = re-examination.

† One classification is either fast breathing, fever, or diarrhea. Two is any combination of two of these illnesses per child; three is all three. Danger signs excluded.

‡  $P$  value represents chi-square test, or Fisher's exact test as appropriate, for difference in proportions of the method to direct observation with re-examination adjusting for health surveillance assistant clustering.

§ Aggregate indicator (fast breathing, fever, and/or diarrhea) for case scenarios calculated at the health surveillance assistant level, i.e., health surveillance assistant correctly treated all three case scenarios for uncomplicated illness.

TABLE 6  
Misuse of antibiotics and antimalarial drugs, by method, Malawi\*

Feature	Misuse of antibiotics and antimalarial drugs			
	No fast breathing, received cotrimoxazole	Cough (no fast breathing), received cotrimoxazole	No fever, received AL	
Direct observation with re-examination	No.	272	113	64
	% Detected (95% CI)	24 (17–31)	48 (36–60)	3 (0–8)
Direct observation only	No.	247	84	80
	% Detected (95% CI)	6 (2–9)	16 (6–25)	2 (0–6)
	Difference in % points from DO with RE	–18	–32	–1
	<i>P</i> †	< 0.001	< 0.001	0.869
Register review	No.	791	125	273
	% Detected (95% CI)	18 (13–23)	73 (59–87)	5 (3–8)
	Difference in % points from DO with RE	–6	+25	+2
	<i>P</i> †	0.192	< 0.001	0.746
Case scenarios	No.	131	NA	131
	% Detected (95% CI)	2 (0–5)	NA	9 (4–14)
	Difference in % points from DO with RE	–22	NA	+6
	<i>P</i> †	< 0.001	NA	0.15

\* AL = artemether/lumefantrine; CI = confidence interval; DO = direct observation; RE = re-examination; NA = no case scenarios tested this classification or combination of classifications.  
† *P* value represents chi-square test, or Fisher's exact test as appropriate, for difference in proportions of the method to direct observation with re-examination adjusting for health surveillance assistant clustering.

HSAs may feel pressured to include a diagnosis of cough with fast breathing and dispense antibiotics either because of previous habits or because the caregiver expects it, as reported in studies of patient influence on prescribing practices for acute respiratory infections.<sup>33,34</sup> Therefore, the HSAs might justify prescribing antibiotics when not indicated and record these cases, but change their behavior in the presence of an observer.

Unfortunately, registers do not exist in all settings where CHWs work, and the quality and quantity of documentation varies in those that are available.<sup>5,35</sup> Although most HSAs did have registers, minimal information was recorded and did not enable a complete assessment of decisions made. For example, we found that fewer severe cases but more uncomplicated fast-breathing cases were recorded in RR compared with DO, and cannot be sure if this is associated with poor documentation or is a true difference. We directly observed substantial errors in recognizing and referring severe illness and HSAs may not have recorded severe cases in registers because of lack of recognition and/or discomfort managing more severe cases. Our assessment took place early in the CCM implementation in Malawi and extracted data from ad hoc registers created by each HSA for his or her own records. Since the time of our assessment, new, standardized registers have been introduced that include more detailed information on classifications, counted breaths per minute, danger signs, and treatments given. These registers could increase the validity of information gleaned from the registers, and serve as a job aid that could improve HSA performance.

Case scenarios assess clinical knowledge of CHWs, rather than actual practice, which is their primary limitation. Conversely, one advantage in our assessment was that our CS included more severe cases requiring complex management; DO of presenting cases could not capture substantial numbers of rare, severe illnesses, and RR also had small sample sizes and potential documentation bias for severe cases. Also, CS were not resource intensive to administer and can elucidate gaps in CHW knowledge in management of severe illnesses. Although training is still required for surveyors, scenarios can be applied in any setting (community or facility) and in CHW programs with low patient volume. A

study directly comparing knowledge of CHWs in the management of severe cases to their actual practice in clinical settings would provide useful further information on the value of this method.

Our comparison of correct treatment across all methods indicated that increasing complexity of illness led to lower HSA performance. First, HSAs performed better for uncomplicated illnesses than severe illnesses in DO with RE, DO only, and CS methods. Second, HSAs performed better in managing cases with a single classification than cases with two or three classifications across DO with RE, RR, and CS methods. Compared with DO with RE, these results suggest that RR and CS, and in some cases DO only, may be able to detect a large decrease in HSA performance caused by the presence of danger signs or increasing number of classifications. A difference in CHW performance by severity of disease has been reported.<sup>20</sup> However, the correlation between increasing number of classifications and decreasing CHW performance should be confirmed in future studies.

This analysis has limitations. The study was not specifically designed to compare these methods, especially for severe illnesses, and small sample sizes limit the conclusions that can be drawn in some cases. For RR, we did not abstract the registers for the directly observed cases, which would have enabled more robust estimates of the reliability and validity of RR compared with DO with RE. This additional process should be considered in future studies of this kind. Case scenarios, by necessity, must be designed ahead of time, and therefore the comparison to actual cases seen is inflexible once the assessment begins. Finally, we were not able to assess the relative costs of these methods because all were implemented simultaneously as a part of a larger research study.

The results reported are based on one application in Malawi, and may differ in other settings. First, HSAs saw a median of three sick children per day in their communities during our observations, which is a relatively high caseload not always seen in other CHW programs. Second, HSAs were observed in the same settings where they practice, and they were assessed managing sick children from their



communities who were likely to reflect a case mix similar to the routine practice of the HSA. Although our assessment did not directly compare HSA performance in the community versus their performance if brought to a health facility for the assessment, we believe that conducting this quality of care assessment in the community likely produced results with higher validity. In other settings with lower care-seeking to CHWs, DO at the community level may not be feasible or affordable, given that surveyors will need to spend long periods in the field. Third, HSAs in Malawi have more education than CHWs providing CCM in other countries in Africa, and receive a government salary, and may therefore have conceptual skills to handle CS and be better motivated to perform their jobs well. Fourth, HSAs maintain relatively detailed clinical registers, but such documentation is not standard practice in all CHW settings. Fifth, CCM in Malawi focuses on fast breathing, fever, and diarrhea cases, but other CHWs may be trained in one disease only or have other responsibilities that require adaptation of the assessment method. Future assessments should investigate these methods, and any alternatives, in different contexts.

Calls for greater accountability in maternal, newborn, and child health programs<sup>36</sup> present new challenges to governments and their development partners about how to conduct regular assessments of the quality of health services delivered at community level. In this study, we aimed to assess the quality of care received by sick children and HSA performance at the community level, and the added validity and reliability of information gained by using DO with RE outweighed the intensity of resources required to carry out this method. However, if the objective is to determine effectiveness of a training program in developing knowledge, CS may be sufficient, or for skills building or reinforcement, DO with RE in a centralized setting may be sufficient. In all settings, if written documentation of cases is available, RR can provide additional information on historical cases, and CS can supplement assessment of routine clinical cases with more severe, rare cases with danger signs. In our assessment, a more comprehensive, valid and reliable picture of HSA performance was created by using all methods together in the community, but use of all methods together may not always be feasible for program managers or researchers. Our analysis of method alternatives to assess the performance of CHWs in delivering community case management of childhood illnesses indicates that the assessment objectives and feasibility of each method should be considered carefully in the context of the CHW program being assessed before making a selection.

Received June 20, 2012. Accepted for publication August 19, 2012.

**Acknowledgments:** We thank the health surveillance assistants and surveyors for participating in this assessment, the CCM-Malawi Quality of Care Working Group for their contributions, and Agbessi Amouzou for assistance during the analysis and write-up. The CCM-Malawi Quality of Care Working group includes Samira Aboubaker, Agbessi Amouzou, Enoch Bonongwe, Jennifer Bryce, Jennifer Callaghan, Cristina Cardemil, Bernadette Daelmans, Kate Gilroy, Asha George, Timothy Kachule, Susan Kambale, Regis Katsande, Leslie Mgalula, Angella Mtimuni, Edwin Nkhono, Humphreys Nsona, Thoko, Sambakunsi, Newton Temani, and Texas Zamasiya.

**Financial support:** This study was supported by the Bill and Melinda Gates Foundation through a contract to the World Health Organization to support the independent evaluation of the Maternal New-

born Child Health Rapid Scale-Up in Malawi. Cristina V. Cardemil received additional support from the Child Health Center Board and the Robert H. Parrott REACH Program.

**Authors' addresses:** Cristina V. Cardemil, Centers for Disease Control and Prevention, Atlanta, GA, E-mail: ccardemil@cdc.gov. Kate Gilroy, Jennifer Callaghan-Koru, and Jennifer Bryce, Institute for International Programs, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, E-mails: kgilroy@jhsph.edu, jcallagh@jhsph.edu, and jbryce@jhsph.edu. Humphreys Nsona, Integrated Management of Childhood Illnesses Unit, Ministry of Health, Lilongwe, Malawi, E-mail: hnsona@gmail.com.

## REFERENCES

- Haines A, Sanders D, Lehmann U, Rowe AK, Lawn JE, Jan S, Walker DG, Bhutta Z, 2007. Achieving child survival goals: potential contribution of community health workers. *Lancet* 369: 2121–2131.
- Rowe AK, Savigny D, Lanata CF, Victoria CG, 2005. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet* 366: 1026–1035.
- Hermida J, Nicholas DD, Blumenfeld SN, 1999. Comparative validity of three methods for assessment of the quality of primary health care. *Int J Qual Health Care* 11: 429–433.
- Bryce J, Toole MJ, Waldman RJ, Voigt A, 1992. Assessing the quality of facility-based child survival services. *Health Policy Plan* 7: 155–163.
- Franco LM, Franco C, Kumwenda N, Nkhoma W, 2002. Methods for assessing quality of provider performance in developing countries. *Int J Qual Health Care* 14 (Supp 1): 17–24.
- Hrisos S, Eccles MP, Francis JJ, Dickinson HO, Kaner EF, Beyer F, Johnston M, 2009. Are there valid proxy measures of clinical behaviour? A systematic review. *Implement Sci* 4: 37.
- Rowe SY, Olewe MA, Kleinbaum DG, McGowan JE, McFarland DA, Rochat R, Deming MS, 2006. The influence of observation and setting on community health workers' practices. *Int J Qual Health Care* 18: 299–305.
- Campbell JP, Maxey VA, Watson WA, 1995. Hawthorne effect: implications for prehospital research. *Ann Emerg Med* 26: 590–594.
- Leonard KL, Masatu MC, 2005. The use of direct clinician observation and vignettes for health services quality in developing countries. *Soc Sci Med* 61: 1944–1951.
- Rowe AK, Lama M, Onikpo F, Deming MS, 2002. Health worker perceptions of how being observed influences their practices during consultations with ill children. *Trop Doct* 22: 166–167.
- Rowe AK, Onikpo F, Lama M, Deming MS, 2012. Evaluating health worker performance in Benin using the simulated client method with real children. *Implement Sci* 7:95 (Published online 8 Oct 2012). doi: 10.1186/1748-5908-7-95
- World Health Organization, 2007. *Community Health Workers: What Do We Know about Them? The State of the Evidence on Programmes, Activities, Costs and Impact on Health Outcomes of Using Community Health Workers*. Available at: [http://www.who.int/healthsystems/round9\\_7.pdf](http://www.who.int/healthsystems/round9_7.pdf). Accessed June 10, 2012.
- Ashwell HES, Freeman P, 1995. The clinical competency of community health workers in the Eastern Highlands Province of Papua New Guinea. *P N G Med J* 38: 198–207.
- Mehnaz A, Billoo AG, Yasmeen T, Nankani K, 1997. Detection and management of pneumonia by community health workers: a community intervention study in Rehri Village, Pakistan. *J Pak Med Assoc* 47: 42–45.
- Hadi A, 2003. Management of acute respiratory infections by community health volunteers: experience of Bangladesh Rural Advancement Committee (BRAC). *Bull World Health Organ* 81: 183–189.
- Baqui AH, Arifeen SE, Rosen HE, Mannan I, Rahman SM, Al-Mahmud AB, Hossain D, Das MK, Begum N, Ahmed S, Santosham M, Black RE, Darmstadt GL, Projahnmo Study Group, 2009. Community-based validation of assessment

- of newborn illnesses by trained community health workers in Sylhet district of Bangladesh. *Trop Med Int Health* 14: 1448–1456.
17. Darmstadt GL, Baqui AH, Choi Y, Bari S, Rahman SM, Mannan I, Ahmed ASM, Sahar SK, Rahman R, Chang S, Winch PJ, Black RE, Santosham M, El Arifeen S, for the Bangladesh Projahnmo-2 (Mirzapur) Study Group, 2009. Validation of community health workers' assessment of neonatal illness in rural Bangladesh. *Bull World Health Organ* 87: 12–19.
  18. Puett C, Coates J, Alderman H, Sadler K, 2012. Quality of care for severe acute malnutrition delivered by community health workers in southern Bangladesh. *Matern Child Nutr* (Published online 19 Apr 2012). doi: 10.1111/j.1740-8709.2012.00409.x
  19. Zeitz PS, Harrison LH, Lopez M, Cornale G, 1993. Community health worker competency in managing acute respiratory infections of childhood in Bolivia. *Bull Pan Am Health Organ* 27: 109–119.
  20. Kelly JM, Osamba B, Garg RM, Hamel MJ, Lewis JJ, Rowe SY, Rowe AK, Deming MS, 2001. Community health worker performance in the management of multiple childhood illnesses: Siaya District, Kenya, 1997–2001. *Am J Public Health* 91: 1617–1624.
  21. Rowe SY, Olewe MA, Kleinbaum DG, McGowan JE, McFarland DA, Rochat R, Deming MS, 2007. Longitudinal analysis of community health workers' adherence to treatment guidelines, Siaya, Kenya, 1997–2002. *Trop Med Int Health* 12: 651–663.
  22. Rowe SY, Kelly JM, Olewe MA, Kleinbaum DG, McGowan JE Jr, McFarland DA, Rochat R, Deming MS, 2007. Effect of multiple interventions on community health workers' adherence to clinical guidelines in Siaya, Kenya. *Trans R Soc Trop Med Hyg* 101: 188–202.
  23. Kallander K, Tomson G, Nsabagasani X, Sabiiti JN, Pariyo G, Peterson S, 2006. Can community health workers and caretakers recognise pneumonia in children? Experiences from western Uganda. *Trans R Soc Trop Med Hyg* 100: 956–963.
  24. Mukanga D, Babirye R, Peterson S, Pariyo GW, Ojiambo G, Tibenderana JK, Nsubuga P, Kallander K, 2011. Can lay community health workers be trained to use diagnostics to distinguish and treat malaria and pneumonia in children? Lessons from rural Uganda. *Trop Med Int Health* 16: 1234–1242.
  25. Yasuoka J, Poudel KC, Poudel-Tandukar K, Nguon C, Ly P, Socheat D, Jimba M, 2010. Assessing the quality of service of village malaria workers to strengthen community-based malaria control in Cambodia. *Malar J* 9: 109.
  26. Gilroy KE, Callaghan-Koru JA, Cardemil CV, Nsona H, Amouzou A, Mtimuni A, Daelmans B, Mgalula L, and Bryce J on behalf of the CCM-Malawi Quality of Care Working Group, 2012. Quality of sick child care delivered by Health Surveillance Assistants in Malawi. *Health Policy Plan*: first published online October 13, 2012. doi:10.1093/heapol/czs095
  27. World Health Organization, 2003. *Health Facility Survey: Tool to Evaluate the Quality of Care Delivered to Sick Children Attending Outpatient Facilities*. Available at: [http://www.who.int/maternal\\_child\\_adolescent/documents/9241545860/en/index.html](http://www.who.int/maternal_child_adolescent/documents/9241545860/en/index.html). Accessed June 10, 2012.
  28. WHO/UNICEF, 2011. *Caring for the Sick Child in the Community*. Geneva: World Health Organization.
  29. Institute for International Programs, Johns Hopkins Bloomberg School of Public Health. *Quality of Care Tools (Malawi)*. Available at: [www.jhsph.edu/dept/ih/IIP/projects/catalyticinitiative.html](http://www.jhsph.edu/dept/ih/IIP/projects/catalyticinitiative.html).
  30. Nsona H, Mtimuni A, Daelmans B, Callaghan-Koru JA, Gilroy K, Mgalula L, Kachule T, Zamasiya T, 2012. Scaling up integrated community case management of childhood illness: update from Malawi. *Am J Trop Med Hyg* 87 (Suppl 5): 54–60.
  31. Landis JR, Koch GG, 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
  32. Stata Corp LP, 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.
  33. Macfarlane J, Holmes W, Macfarlane R, Britten N, 1997. Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: questionnaire study. *BMJ* 315: 1211–1214.
  34. Scott JG, Cohen D, DiCicco-Bloom B, Orzano AJ, Jaen CR, Crabtree BF, 2001. Antibiotic use in acute respiratory infections and the ways patients pressure physicians for a prescription. *J Fam Pract* 50: 853–858.
  35. Gilroy K, Winch P, 2006. *Management of Sick Children by Community Health Workers: Intervention Models and Program Examples*. UNICEF and WHO. Available at: [www.coregroup.org/storage/Program\\_Learning/Community\\_Health\\_Workers/Management\\_of\\_Sick\\_Children\\_by\\_Community\\_Health\\_Workers.pdf](http://www.coregroup.org/storage/Program_Learning/Community_Health_Workers/Management_of_Sick_Children_by_Community_Health_Workers.pdf). Accessed April 25, 2012.
  36. Commission on Information and Accountability for Women's and Children's Health, 2011. *Keeping Promises, Measuring Results*. Geneva: World Health Organization.