Standardizing Direct Observation for Assessing Compliance to a Daily Chlorhexidine Bathing Protocol Among Hospitalized Patients

An efficacious intervention for preventing health care-associated infections is daily bathing with chlorhexidine gluconate (CHG). Consequently, many hospitals in the United States have implemented CHG bathing in their intensive care units (ICUs) and non-ICU units. With the increasing implementation of CHG bathing in healthcare facilities, it is important to monitor compliance to ensure that CHG baths are appropriately conducted and to identify potential opportunities for improving the process.

Most studies on compliance with CHG bathing procedures have used bathing product purchasing data or inventory assessments as proxy measures of compliance. Direct observations of CHG bathing may be a preferred method for assessing compliance and for understanding the overall process. Direct observation is an effective method for collecting real-time, naturalistic behavioral information about a specific process, and this method is commonplace in infection prevention.

Herein, we describe our experience training observers to conduct CHG bathing observations, and we present findings from pilot observations.

METHODS

Training of New Observers

An experienced observer trained 2 new observers using a CHG bathing training manual (with components for both ICU and non-ICU CHG bathing) created by a multidisciplinary team. The training manual is available on our website (http://cqpi.wisc.edu/1758.htm). New observers (trainees) were given 3 days to read the training manual; then they met with the experienced observer, who reviewed each item on the data collection tool, a checklist similar to one used in our previous study. Pilot observations were conducted after this training session. The University of Wisconsin Minimal Risk Institutional Review Board exempted this project as a quality improvement project.

Pilot Observations

Each observer completed 5 observations. The first observation was educational in nature; the experienced observer introduced trainees to the observation methodology. Trainees watched the experienced observer collect data, paying attention to how the data collection instrument was completed. The second observation was similar, but the trainees were more independent and completed the data collection instrument side-by-side with the experienced observer. The subsequent 3 observations were standard inter-rater reliability (IRR) testing observations (no communication between observers) used to assess the extent of agreement between observers.

Assessment of Agreement Between Observers

We assessed the extent of agreement (or IRR) between observers using Cohen’s κ, a measure of agreement between 2 raters that accounts for agreement due to chance alone. Kappa interpretations were defined as follows: values ≤ 0 indicated no agreement; 0.01–0.20 indicated no agreement to slight agreement; 0.21–0.40 indicated fair agreement; 0.41–0.60 indicated moderate agreement; 0.61–0.80 indicated substantial agreement; and 0.81–1.00 indicated almost perfect agreement.

Calculation of IRR

Our data collection instrument (Online Supplementary Material) was a checklist. It included a section where observers could record notes. Except for the observer notes, IRR was calculated based on all data collection items because this is where variation was expected to occur. Responses to items were arbitrarily assigned numbers, and the data for both experienced and trainee observers were entered in an Excel spreadsheet (Microsoft, Redmond, WA). For example, a response of “No” was assigned 0, “Yes” was assigned 1, “Not Applicable” was assigned 2, and so on. Items that involved recording time (eg, timing of the bath) were coded as follows: A score of 1 was assigned if the time recorded was the same or differed by ±1 minute. For example, if observer A recorded “Total time CHG is left on chest before rinsing” of 2:35, to assign a score of 1, observer B should have assigned a time including and between 2:35 and 3:35; otherwise, a score of 0 was assigned. These data were exported to SPSS (IBM SPSS Statistics, version 22.0, Armonk, NY) where the IRR was calculated. Acceptable IRR in this project was κ ≥ 0.8 on 3 consecutive pilot observations.

RESULTS

Each observer conducted a total of 5 pilot observations, 2 of these observations were purely training observations, while 3 were IRR observations were used to assess the extent of agreement between observers. Comparing trainee with experienced observer, all observations for both trainees had a κ ≥ 0.8, and all were statistically different from 0. Results are summarized in Table 1.

DISCUSSION

In this paper we present our experience with training for CHG bathing observations and provide findings from our pilot observations. Following training, trainee observers were able to
achieve substantial to almost perfect agreement with an experienced observer. We believe that our project is the first to report measurement of IRR for CHG bathing using direct observations.

High levels of IRR in this project may be attributed to the use of a standard, detailed training manual, which all trainees had to review and understand, as well as in-person training on use of the instrument.

Adherence to and standardization of a CHG bathing protocol is important to ensure consistent application and optimal skin coverage and to achieve sufficient concentrations of CHG on the patient’s skin. These measures benefit clinical outcomes while avoiding adverse effects of contact with mucous membranes.

We recommend that institutions planning to conduct direct observations of CHG bathing perform training and evaluation of observer competency prior to collecting data. Periodic IRR should also be conducted to ensure that observers are continually within acceptable agreement.

In conclusion, with standardized and systematic training, high levels of IRR can be achieved for CHG bathing observations.

**Acknowledgments**

Financial support: This project was supported by the Research and Development Service Quality Enhancement Research Initiative project no. PEC 15-248, a Patient Safety Center of Inquiry United States (US) Department of Veterans Affairs and grant number R18HS024039 from the Agency for Healthcare Research and Quality.

Potential conflicts of interest: All authors report no conflicts of interest relevant to this article.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality and the United States (US) Department of Veterans Affairs.

**TABLE 1. Cohen’s κ Comparing Experienced and Trainee Observers**

<table>
<thead>
<tr>
<th>Observation</th>
<th>κ</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced observer vs trainee 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First IRR observation</td>
<td>0.91</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Second IRR observation</td>
<td>0.81</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Third IRR observation</td>
<td>0.88</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Experienced observer vs trainee 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First IRR observation</td>
<td>0.80</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Second IRR observation</td>
<td>0.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Third IRR observation</td>
<td>0.92</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**NOTE.** IRR, inter-rater reliability.

**References**