Validity and Reliability of an Alcohol Withdrawal Clinical Assessment Scale for Use with Acutely Ill Patients: An Abbreviated Version of the CIWA-Ar

Perry M. Gee, Herschel Knapp, Kevin P. Miller, Michael J. Sieczka, Jill Welton, Tricia Cavallero, Regina Truong, Cathy Chiu, Odette Horsthius, Michelle Hallisy and Frances L. Patmon

Nursing Research and Analytics, Dignity Health, San Francisco, CA
College of Nursing, University of Utah, Salt Lake City, UT, USA
Betty Irene Moore School of Nursing, University of California, Davis, Sacramento, CA, USA
Department of Social Work, University of Southern California, Los Angeles, CA
Clinical Informatics, St Francis Memorial Hospital, San Francisco, CA
Department of Nursing, St Francis Memorial Hospital, San Francisco, CA
Department of Pharmacy, St Francis Memorial Hospital, San Francisco, CA

Abstract: The number of people in the United States admitted to hospitals with alcohol use disorder is increasing. Screening for alcohol dependence is a critical component to prevent alcohol withdrawal syndrome which places both the patient and staff at risk for injury. Management of symptoms includes early identification of at-risk individuals and continuous assessment of symptoms and treatment plans. In light of the growing number of patients in need of care for alcohol disorders, our goal was to assess the reliability and validity of a more concise version of the Clinical Institute Withdrawal Assessment of Alcohol Scale, Revised (CIWA-Ar) to enhance clinical efficiency without compromising the integrity of this diagnostic instrument. Members of an interdisciplinary team researched and revised the full 10-item CIWA-Ar to six items. Clinicians were trained on how to conduct and document their diagnostic findings using this abridged 6-item instrument. The research team then examined the psychometric properties of the new brief instrument to determine reliability and validity. The research team confirmed content validity. Based on data drawn from multiple sites, we compared the 10-item version of this instrument to the 6-item version; this produced a strong positive correlation, thereby satisfying concurrent criterion validity. Correlational analysis of 71 cases confirmed robust interrater reliability. This pilot study suggests that the new 6-item scale is a valid and reliable instrument, essentially performing equivalently to the 10-item version. Our research team therefore suggests further multi-site studies to confirm our findings of this more concise diagnostic instrument.

Keywords: Alcohol Withdrawal Scale, Alcohol Withdrawal Syndrome, CIWA-Ar, Symptoms, Assessment, Alcoholism

Introduction

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (APA, 2013) has combined alcohol abuse and alcohol dependence into its own disorder called the Alcohol Use Disorder (AUD). In the United States over 16 million adults have AUD with the incidence in males (10.8 million) nearly twice that of females (5.8 million) (NIAAA, 2015). The number of patients admitted to the hospital in the United States with AUD is increasing. In 2010 over 960,000 adults were admitted to hospitals with AUD (NIAAA, 2013). Screening for alcohol dependence is a critical component to prevent Alcohol Withdrawal Syndrome...
(AWS) which places both the patient and staff at risk for injury. As many as 30% of patients with AWS are admitted to Intensive Care Units (ICU) and while in the ICU those affected tend to have more clinical complications (Adams and Ferguson, 2017). Management of symptoms include early identification of at-risk individuals and continual assessment of symptoms and treatment plans (Melson et al., 2014; Perry, 2014; Salottolo et al., 2017). Benzodiazepines are shown to be an effective therapy for AWS and are a main component of treatment for patients experiencing AWS symptoms particularly because of their efficacy in reduction or prevention of seizures (Amato et al., 2010; Perry, 2014; Adams and Ferguson, 2017). Benzodiazepine therapy was traditionally dosed in fixed, scheduled regimens, but more recent recommendations utilize a variable dosing schedule that is driven by a structured protocol dependent on assessment scale scoring in an effort to utilize the minimum dose necessary, but complexity of these procedures leads to increased training and allocation of nursing resources (Amato et al., 2010; Perry, 2014). Another scale used with the critically ill patient, the Richmond Agitation-Sedation Scale (RASS) score is commonly administered to facilitate Benzodiazepine therapy and care coordination (Ely et al., 2003; Sessler et al., 2002).

An urban hospital located in a large Northern California city was challenged with a high number of patients admitted with acute alcohol withdrawal or Delirium Tremens (DTs). Complicated AWS in the acute and critical care environment is associated with significant clinical sequela and even mortality (Maldonado, 2017). Based on ongoing clinical staff concerns, focus groups were initiated by the facility clinical leaders who found that the patient care teams experience was that executing the full 10 element scale was too time consuming and interfered with their ability to provide the interventions necessary in this patient population/treatment scenario. An interdisciplinary team at the facility determined the usual scale for AWS, the Clinical Institute Withdrawal Assessment of Alcohol Scale, Revised (CIWA-Ar), was not adequate for their acute patient population. The team determined that the face validity, the scale’s perceived ability to measure the phenomenon at hand, did not adequately capture AWS for their population of acute medical-surgical and critically ill patients. Some researchers have expressed concerns that the “standard” CIWA-Ar scale has not been sufficiently studied in the emergency department and critical care environments (Bostwick and Lapid, 2004; Maldonado et al., 2015; Sarff and Gold, 2010). Specifically, the hospital-based team was concerned that some of the more subjective findings in the scale (headache, nausea and tactile stimulation) were difficult to assess with patients who are agitated, critically ill, or sedated. In summary, the team was in search of a scale that could be administered in less time and not as dependent on the elements that they were not always able to elicit.

An interdisciplinary group of clinical experts at the facility familiar with AWS and the patient population suggested a modified version of the 10-item CIWA-Ar scale. The new Alcohol Withdrawal Clinical Assessment (AWCA) scale had been developed earlier with a consensus of the expert clinical team at a different hospital in Oakland, CA. This team successfully used the AWCA for the management and treatment of acutely ill patients with AWS (Rosenson et al., 2012). The AWCA is a shortened version of the valid and reliable CIWA-Ar instrument (Stuppaec et al., 1994; Sullivan et al., 1989) where the factors headache, nausea, vomiting and tactile disturbances were removed. Additionally auditory and visual disturbances findings were combined into a single factor. These changes formed a new 6-item scale that included: Tremor, paroxysmal sweats, agitation, auditory or visual disturbances, anxiety and orientation/clouding of senses that the team concluded they could more objectively assess and measure.

Our literature review revealed that there is a rational history of utilizing an abbreviated version of the AWCA instrument (Foy et al., 2006); specifically the standard CIWA-Ar itself was derived from the original 15-item CIWA-A scale (Sullivan et al., 1989). A comparison study found that a shorter 8-item CIWA-AD worked as well as the standard CIWA-Ar and was more acceptable to the clinicians (Reoux and Oreskovich, 2006). In a double blind, randomized, placebo controlled study of the use of single-dose phenobarbital for the treatment of AWS the research team successfully used the shorter 6-item AWCA (Rosenson et al., 2012).

Because the six-question version of the AWCA is new and has only been trialed at two facilities, the research questions addressed in this study focus on the reliability and validity in quantifying the alcohol withdrawal syndrome for patients with acute illness. Sound psychometrics is essential to assure new or revised instruments will generate accurate research findings (DeVon et al., 2007). Therefore the objectives addressed in this study are to analyze the face validity and content validity of the AWCA scale and to evaluate inter-rater reliability.

Methods

Setting

After receiving human subjects’ protection (IRB) authorization, the study was performed in a large, urban, hospital in a downtown area of a Northern California city. The research was conducted with clinical staff in three domains within this facility: Medical-surgical, telemetry and critical care units.

Sample

The research team first conducted and confirmed content validity tests using an interdisciplinary group of
nine expert clinical staff at the facility who are familiar with the assessment, care and management of patients who have, or who are at risk for AWS. These clinicians, who were hospital-based registered nurses, pharmacists and physicians, attended a training session detailing the correct use and administration of the AWCA.

Over the next five months, the team administered the AWCA to acutely ill patients who were at risk for or were experiencing AWS following the SFMH Alcohol Withdrawal Protocol. Considering the newness of this implementation at this site, clinicians followed procedures for assessing interrater reliability: Instead of solo assessments, the trained clinicians independently completed the examination simultaneously or within five minutes of each other using the revised 6-metric scale and then scored the AWCA blinded to each other’s results. In addition to the AWCA scores, the researchers recorded patient age, gender, ethnicity, primary diagnosis, medical unit and the RASS following the confidentiality protocol as specified in the IRB application.

Instruments

An index of Content Validity (CVI) was evaluated using a 4-point ordinal scale and each of the six factors in the AWCA will be rated on a scale of 1 (not relevant) through 4 (extremely relevant) (Lynn, 1986; Polit and Beck, 2006). Using this tool, we evaluated both individual item validity and scale validity (Polit and Beck, 2006). The 6-item AWCA scale was then used by expert clinicians to assess AWS in a selected group of subjects. Because the RASS scale is routinely administered at the same time as the patients are assessed for AWS, the score from this tool was also collected as part of the process. To enhance the precision of the data collection process, a clinician and research team member who is also familiar with the care of this patient population supervised the data gathering process and carefully collected and recorded the findings using the specified confidentiality protocols.

Analysis

Members of the research team collected the data and carefully entered it into a database that was stored on a secure password-protected network server. The data was analyzed by qualified research staff using IBM SPSS® v23 statistical analysis software to compute descriptive and inferential statistics including interrater reliability, concurrent criterion validity comparing the data gathered from the 10-metric version of the ACWA to the 6-metric version and correlational analysis comparing the RASS to the AWCA scores.

Results

Instrument Reduction: Concurrent Criterion Validity Evaluation

In light of prior efforts to create a more parsimonious version of the ACWA, initially, we needed to determine how the shorter (6-metric) version of the AWCA would perform, compared to the original longer (10-metric) version. Specifically, we considered the viability of reducing the 10-metrics on the AWCA form, which measures (1) nausea/vomiting, (2) tremors, (3) anxiety, (4) agitation, (5) paroxysmal sweats, (6) orientation/clouding of sensorium, (7) headache/fullness in head, (8) tactile disturbances, (9) visual disturbances, (10) auditory disturbances, to a 6-metric instrument by removing the following three metrics: (1) nausea/vomiting, (7) headache/fullness in head, (8) tactile disturbances and combining items (9) visual disturbances and (10) auditory disturbances into a single (audiovisual) item; considering that these signs and symptoms are traditionally be gathered and coded elsewhere in the Electronic Health Record (EHR).

To determine the concurrent criterion validity of this abbreviated version of the CIWA-Ar, we extracted only the data points used in the AWCA data from our EHR from 10 similar hospitals and care units associated with a large Western United States health system of the 219 records gathered, nine were eliminated from our analysis due to missing data, leaving a viable sample (n) of 210 spanning eight facilities. Based on this archival data, we computed correlational analysis, comparing the full (10-metric) version to the abbreviated (6-metric) version, with the following three items removed: (1) nausea/vomiting, (7) headache/fullness in head, (8) tactile disturbances. We computed correlational analyses for each site, as well as an aggregated total of all sites (Table 1).

<table>
<thead>
<tr>
<th>Hospital</th>
<th>n</th>
<th>M (ACWA&lt;sub&gt;10&lt;/sub&gt;)</th>
<th>M (ACWA&lt;sub&gt;6&lt;/sub&gt;)</th>
<th>r (ACWA&lt;sub&gt;10&lt;/sub&gt;: ACWA&lt;sub&gt;6&lt;/sub&gt;)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>3.00 (2.83)</td>
<td>3.00 (2.83)</td>
<td>1.000</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>8.63 (7.29)</td>
<td>6.25 (5.57)</td>
<td>0.948</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>11.17 (9.91)</td>
<td>9.17 (9.02)</td>
<td>0.942</td>
<td>0.005*</td>
</tr>
<tr>
<td>D</td>
<td>154</td>
<td>2.99 (3.01)</td>
<td>2.60 (2.72)</td>
<td>0.932</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>2.25 (1.89)</td>
<td>2.25 (1.89)</td>
<td>1.000</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>F</td>
<td>21</td>
<td>6.48 (8.64)</td>
<td>5.62 (6.67)</td>
<td>0.969</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>G</td>
<td>8</td>
<td>2.88 (3.09)</td>
<td>2.88 (3.09)</td>
<td>1.000</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>H</td>
<td>6</td>
<td>11.83 (7.06)</td>
<td>9.67 (6.12)</td>
<td>0.888</td>
<td>0.018*</td>
</tr>
<tr>
<td>Total</td>
<td>210</td>
<td>4.08 (5.06)</td>
<td>3.50 (4.23)</td>
<td>0.954</td>
<td>&lt; 0.001*</td>
</tr>
</tbody>
</table>

*p<0.05
Each correlation (r) comparing the original 10-metric version of this instrument to the abbreviated 6-metric version was greater than 8 with a statistically significant p value (α = 0.05), indicating strong concurrent criteria validity between the full version and the abbreviated version of this instrument. These findings suggest that the 6-metric version of this instrument performed equivalently to the (full) 10-question version. As mentioned earlier, data regarding the three omitted metrics (nausea/vomiting, headache/fullness in head and tactile disturbances) are recorded elsewhere in the electronic medical records for such cases, hence, this reduced version of the instrument serves to enhance efficiency by reducing unnecessary assessment redundancies.

**Interrater Reliability**

Considering that this instrument was new to this facility, appropriate training was given to the providers who would be using it. The clinicians (raters) administered the examinations in the telemetry and intensive care units. We calculated and confirmed that interrater reliability was robust for both the ACWA (r = 0.900, p < 0.001) and the RASS (r = 0.904, p < 0.001); the AWCA is the primary focus of this paper.

**Patient Characteristics**

Our sample, consisting of 72 adult patients, was gathered over five months (10/26/15 through 3/24/16), during the day shift, between 11:00 and 14:50 (M = 12:57, SD = 1:08).

The majority (73.6%) of the patients in this study were males. Ages ranged from 31 to 74 (M = 57.94, SD = 8.88). On average, women were 2.36 years older than the men, however this difference was found to be statistically insignificant (p = 0.323, α = 0.05) (Table 2). The majority of the patients were Caucasian (69.4%), followed by Hispanic (13.9%), African American (11.1%) and Other (5.6%).

A variety of admitting diagnoses were encountered; as expected, the majority of the diagnoses (54.2%) were conditions typically associated with long-term alcohol consumption (e.g., esophageal/GI bleed/stricture, pancreatitis, status post-fall, TIA/CVA, atrial fibrillation, acute decompressed diastolic and systolic heart failure), 22.2% were acute alcohol withdrawal and 23.6% involved diagnoses not necessarily associated with alcohol (ab)use (e.g., cellulitis, Dilantin toxicity, pneumonia, sepsis).

**RASS and ACWA**

RASS scores ranged from -2 to +2 (M = -0.1366, SD = 0.705). No statistically significant differences in the RASS scores were detected among the genders (p = 0.360, α = 0.05) or races in our sample (p = 0.268, α = 0.05) (Table 3).

**Discussion**

This instrument reliability and validity data gathered in this study confirmed the historical progression of reducing the items the AWCA over the past three decades while maintaining sound psychometric properties (Foy et al., 2006; Reoux and Oreskovich, 2006; Rosenson et al., 2012; Sullivan et al., 1989). The clinical facility in this study has continued to use the

### Table 2: Patient ages

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>M Age (SD)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>53</td>
<td>57.32 (9.55)</td>
<td>0.323*</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
<td>59.68 (6.59)</td>
<td></td>
</tr>
</tbody>
</table>

*No significant difference in age stratified by gender (α = 0.05)

### Table 3: RASS scores

<table>
<thead>
<tr>
<th>Gender</th>
<th>M RASS (SD)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-0.182 (0.635)</td>
<td>0.360*</td>
</tr>
<tr>
<td>Female</td>
<td>-0.009 (0.877)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>-0.170 (0.707)</td>
<td>0.268*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.400 (0.699)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.333 (0.735)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0 (0.000)</td>
<td></td>
</tr>
</tbody>
</table>

*No significant differences in Gender or Race (α = 0.05)

### Table 4: ACWA scores

<table>
<thead>
<tr>
<th>Gender</th>
<th>M ACWA (SD)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2.503 (1.716)</td>
<td>0.235*</td>
</tr>
<tr>
<td>Female</td>
<td>3.105 (2.291)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>2.840 (2.002)</td>
<td>0.279*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2.883 (2.883)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.979 (2.157)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.250 (0.500)</td>
<td></td>
</tr>
</tbody>
</table>

*No significant differences in Gender or Race (α = 0.05)

ACWA scores ranged from 0 to 9 (M = 2.662, SD = 1.887). No statistically significant differences in the ACWA scores were detected among the genders (p = 0.235, α = 0.05) or races (p = 0.279, α = 0.05) (Table 4).

**Limitations**

This study has several limitations: This pilot study was implemented at a single site, which, expectedly, acquired a relatively small sample size (n = 72). Based on these findings, we are considering conducting a multisite implementation of this 6-metric instrument using the same protocol. Our team computed interrater reliability scores on an ongoing basis, starting at the onset of the study in order to monitor the quality of the data collection process and proper usage of the instrument. We are considering continuing this practice in our next study; potential inconsistencies in the interrater reliability (e.g., r<0.8) could signal the research team to provide supplemental training(s) to the clinicians.
AWCA and incorporated the new tool into their current assessment protocol. The new, more concise instrument and increased awareness have led to several notable improvements of AWS at the facility involved in this study. Overall the care of the AWS patient at the hospital in this study has seen (1) a reduction of intensive care length of stay by 1.5 days; (2) a decrease of overall hospital length of stay by 1.7 days; and (3) a decrease of Benzodiazepine dosing by 26%. Additionally the clinical staff is reporting an increased confidence in using the AWCA and is assured that they are truly measuring AWS. The improvements in AWS care cannot be directly attributed to the implementation of the AWCA; however the new instrument is plausibly considered an integral component in a successful program.

Conclusion

A reliable and valid alcohol withdrawal assessment tool is necessary for the care and safe management of acutely ill patients who are at risk for, or are experiencing acute alcohol withdrawal syndrome. One of the first logical steps in the revision of a research instrument is determining if the tool has appropriate translation validity (face and content) and has robust interrater reliability. This paper has effectively answered the recommendation of Rosenson et al. (2012) to determine the validity of the instrument (2012). Our findings revealed that the shortened version of this diagnostic instrument performed equivalently to the longer version, which is consistent with the findings of other research on this topic (Foy et al., 2006; Sullivan et al., 1989; Reoux and Oreskovich, 2006; Rosenson et al., 2012). As such, this briefer version has the potential to accurately characterize patients with alcohol disorder, expedite the diagnostic process without compromising the precision of the assessment and help to streamline clinical processes. Considering that these pilot findings are based on an implementation at one intercity hospital, our next step will be to expand to a larger, multi-site study to confirm the validity and reliability of the AWCA as a viable more compact essential diagnostic instrument.

Author Contributions

Perry Gee: Developed research plan, conducted research, assisted with analysis, developed publication and facilitated drafts and revisions.

Jill Welton: Organizer of validation study and creator of Alcohol Withdrawal Protocol In use at Saint Francis using the Alcohol Withdrawal Clinical Assessment (AWCA) tool.

Kevin Miller: Assisted in development of Alcohol withdrawal protocol at Saint Francis with oversight by Jill Welton managed the validation process of the AWCA tool at Saint Francis.

Cathy Chiu and Tricia Cavallero: Validator of AWCA tool at Saint Francis and participated in a team of validators performing real time, assessments of patients in both Medical/Surgical and Critical Care areas.

Regina Truong, Odette Horsthius and MaehleFFE

Hallsy: Validator of AWCA tool at Saint Francis and participated in a team of validators performing real time assessments of patients In both Medical/Surgical and Critical Care areas.

Herschel Knapp: Research design consultation, statistical analysis, authored Results section, coauthored Methods section.

Michael J. Sieczka: Manuscript editing and revision.

Frances L. Patmon: Review of publication.

Funding and Conflict of Interest

No funding or conflict of interest noted by any of the participants.

References


NIAAA, 2013. Alcohol-related emergency department visits and hospitalizations and their co-occurring drug-related, mental health and injury conditions in the United States: Findings from the 2006-2010 Nationwide Emergency Department Sample (NEDS) and the Nationwide Inpatient Sample (NIS).


