Development of a medical error scale for nurses

Havva Öztürk 1 and İknur Kahriman 2

1Karadeniz Teknik Universitesi, Trabzon, Turkey. (Correspondence to: Havva Öztürk: ozturkhavva@gmail.com)

Abstract

Background: Medical errors have a negative impact on patients, healthcare providers and healthcare organizations. Determining these errors with a valid and reliable tool is important to ensure correct and effective solutions.

Aims: The aim of this study was to develop a medical error scale for nurses.

Methods: The methodological study was conducted with 298 nurses at a university hospital in Trabzon, Turkey. Data were collected with a questionnaire on demographic characteristics and a draft medical error scale.

Results: The content validity index of the scale was 0.82, Cronbach’s alpha was 0.89, and the item-total correlation values ranged from 0.31 to 0.54. Kaiser-Meyer-Olkin was 0.81, Bartlett’s test $\chi^2=5909.75$ and $P = 0.000$, and the anti-image correlations ranged between 0.63 and 0.90. In the four rotations done with Varimax Rotation, 42 items were excluded because their factor loadings were less than 0.45. The scale was formed with 43 items and 6 subscales. The scale had six subscales: falls, blood transfusions, medication practices, care practices, communication, and other controlled practices. The six-subscale structure was confirmed by confirmatory factor analysis ($X^2/df= 2.52$, RMSEA= 0.072, $\chi^2=2143.65$, CFI= 0.90, and IFI= 0.90), and there was a good fit between the scale and its subscales.

Conclusion: The scale was a valid and reliable tool to collect and consistent data on whether nurses acted carefully regarding medical errors during their patient-related practices.

Keywords: Scale, medical error, nurse, nursing, Turkey

Citation: Öztürk H; Kahriman İ. Development of a medical error scale for nurses. East Mediterr Health J. 2019;25(x):xxx–xxx. https://doi.org/10.26719/emhj.19.025

Received: 03/12/17; accepted: 18/09/18

Copyright © World Health Organization (WHO) 2019. Some rights reserved. This work is available under the CC BY-NC-SA 3.0 IGO license https://creativecommons.org/licenses/by-nc-sa/3.0/igo

Introduction

Nurses are primarily responsible for identifying unexpected situations and adverse effects in patient care, and they play a key role in the early identification and prevention of risks as well as in the diagnosis and disclosure of errors regarding patient care (1). However, it is reported that nurses also make medical errors due to the insufficient number of nurses in medical institutions and the resulting fatigue, burnout (2–6), long working hours (4), extensive workload, high number of night shifts per month (2,5,6), ineffective communication among healthcare staff (3,7,8), work with critically ill patients, heavy job stress, unfavourable working conditions, and shift work (3,9,10).
Similarly, a study reported that staff are likely to make medical errors due to a lack of knowledge or professional experience, carelessness or incautiousness, negligence, fatigue or burnout (11). Another study explained other reasons for medical errors, such as careless work, lack of training and communication, indifference to their job, lack of motivation, and hectic working schedule (12). The most common medical errors performed by nurses involve errors about medications, infections, falls, communication errors, and use of incorrect or inappropriate materials (2,13).

Therefore, medical errors should be diagnosed and reported in a timely manner before they pose a serious threat to human health; in addition, their causes should be elucidated, solutions should be offered, and lessons should be learned from such experiences (14). Moreover, it is significant that valid and reliable measurement tools are developed and used periodically to help identify the areas in which nurses are more likely to make errors, and for precautions to be taken to address these risk areas.

A review of the literature in Turkey indicated that there are already several such scales, one of which is the “Malpractice Trends Scale in Nursing”(2). Although it provided a discussion of some aspects of medical errors, it neglected care practices, which is an important main role of nurses. Another published scale in Turkey is the “Medical Errors Attitudes Scale” (15). However, this scale did not focus on the nursing practices that lead to medical errors. Instead, it attempts to identify nurses’ attitudes toward medical errors. In a different study in the Islamic Republic of Iran, a data collection instrument was used to identify the types and causes of medical errors. Nevertheless, the instrument had unsatisfactory validity tests (16). Another survey developed in the United States of America (USA), examined the causes of medication errors (17). In addition, There are also national and international scales about the Patient Safety Culture (18,19). In these scales, however, medical errors were discussed either under a single dimension or as a subscale of other dimensions.

The aim of this study was to develop a valid and reliable scale regarding medical errors. Based on this scale, it will be possible to identify the areas in which errors are more likely to occur, take precautions against the causes of these errors, and therefore minimize the likelihood of their occurrence.

Methods
This study was methodological in its approach to develop a scale called the “Medical Errors Scale for Nurses (MESN)”.

Population and sample
The population consisted of 560 nurses working at a university hospital in Trabzon, Turkey, whereas the sample contained 298 nurses. The study was conducted with nurses who were volunteers for participation in the study and who were not on leave. In the factor analysis, sufficient sample size is indicated as “50 very weak, 100 weak, 200 medium, 300 good, 500 very good and 1000 excellent” (20). According to this explanation, the sample of our study was very close to a good sample. To assess test–retest reliability, a different group of 50 nurses from the same university hospital were included in the sample. As indicated by Sims (21), a minimum of 50 nurses was included in the test–retest analysis. In practice, it is sufficient that the test–retest analysis is carried out with a sample of 30 participants.
Ethical considerations
Written permission was obtained from the management of the university hospital on July 25, 2013 to carry out the study by volunteering nurses whose informed consent had already been obtained. Ethical approval was granted by the Ethics Committee of the Medical Faculty of Karadeniz Technical University on December 2, 2013.

Instruments
The data were collected using an Information form and a draft version of the MESN. The information form contained 13 questions about socio-demographic characteristics: age, marital status, educational status, position, years of working profession, and years of working at the hospital, whereas the draft scale was developed by the researchers based on certain primary references (1–5,9–11,17,19) and opinions about medical errors and patient safety of experts specializing in nursing, nursing management, and nursing ethics.

The draft scale consisted of 94 items, and the components was focused on nursing care practices, medication practices, blood and blood products transfusion, falls prevention, infection control, and communication, using a 5 point Likert scale rated as “always” (5), “often” (4), “sometimes” (3), “rarely” (2), and “never” (1).

Data collection
Data on the validity and reliability of the draft scale were collected June 17–30, 2014. The test–retest data were collected July 15–31, 2014. The scale development process occurred in four stages: assessed face validity, content validity, construct validity with exploratory and confirmatory factor analyses, and reliability. However test–retest reliability was assessed with the 43 item-scale after the other validity and reliability analyses had been performed. This test was administered to a different group of 50 nurses working at the same university hospital on two occasions with a two-week interval in between, and they were asked to complete the scale under a pseudonym.

Data analysis
For analyses of the MESN, the normality of the distribution tests with one sample, Kolmogrov–Smirnov tests were first done and did not show a normal distribution. The validity was analyzed with percentages, averages, and exploratory (EFA) and confirmatory factor analyses (CFA). The EFA was performed with KMO (Kaiser-Meyer-Olkin) and Bartlett’s test, anti-image correlation, Principal Components Analysis, and Varimax Rotation, whereas the CFA was tested with X², root mean square error of approximation (RMSEA), comparative fit index (CFI), and incremental fit index (IFI) etc. t-tests and regression analyses were among the other considerations. The reliability was tested with Cronbach’s alpha, item-total correlations, and test–retest reliability.

Results
Sixty-five percent of the nurses were married, 88.9% had a clinical nurse position, and 70.5% had graduated from a bachelor programme. The mean age was 32.11±7.6 years, with an average of 10.24±7.2 years of professional working and their years of working at the university hospital was an average of 9.10±7.4 years.
Face validity and content validity
For face validity, the scale was submitted to three nurses, who were asked to assess the comprehensibility and length of its items. In addition, when the group of experts evaluated the content validity, face validity was also assessed whether the items were expressed properly, accurately and clearly. Nine items were revised to improve the comprehensibility.

Content validity was tested using Lawshe’s technique. The draft scale was submitted to 15 experts specialized in nursing who were asked to rate each item as “essential”, “useful but not essential” and “not necessary”. Based on their opinions, nine items with a content validity ratio (CVR) less than 0.49 were excluded from the draft scale, which initially contained 94 items. Afterwards, the scale comprised 85 items with a mean content validity index (CVI) of 0.82.

Construct validity
Construct validity was tested with exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). For EFA, Principal Components Analysis and Varimax Rotation were used. The 85-item draft scale had a Kaiser-Meyer-Olkin (KMO) value of 0.81. Bartlett’s test resulted in $\chi^2=5909.75$ and $P < 0.0001$, and the anti-image correlation coefficients ranged between 0.63 and 0.90. For EFA, four rotations were performed with a rotated component matrix, and 42 items were excluded because their loadings were less than 0.45. The finalized version showed no overlap and contained 43 items in six subscales (Table 1).

The scree plot graph also showed that the slope plateaued after the sixth point, supporting the finding that the scale consisted of six factors (20). These factors accounted for 51.58% of the total variance (Table 1). The next stage aimed to name the factors (20,22).

CFA was conducted to confirm the factor structure of the 43-item MESN. The resulting fit indexes were as follows: $\chi^2=2143.65$, $\chi^2/sd= 2.52$, RMSEA= 0.072, CFI= 0.90, and IFI= 0.90. In addition, the independent t-test was performed for the upper and lower 27% of the sample. The results of the t-test were as follows: $t = -24.703$ for the overall scale, $t = -17.887$ for F1, $t = -6.428$ for F2, $t = -6.829$ for F3, $t = -12.069$ for F4, $t = -8.246$ for F5, and $t = -11.582$ for F6, all with significant $P$ values ($P < 0.0001$).

Reliability analysis
The internal consistency of the draft scale was tested using Cronbach’s alpha, Spearman-Brown’s coefficient, and finally Guttman’s coefficient. For the overall scale, Cronbach’s alpha value was 0.89, Spearman-Brown’s coefficient was 0.71, and Guttman’s coefficient was 0.70. Item–total correlation values, which are used not only as a test of reliability, validity, and internal consistency but also as an item analysis or item-discrimination analysis, ranged between 0.31 and 0.54 for all scales, which was significant ($P < 0.0001$). The test–retest correlation values for the overall scale were $t = 0.197$ and $P = 0.845$, whereas $r = 0.562$ and $P < 0.0001$.

Scoring of the scale
The final version of the scale consisted of 43 items in six subscales: 12 items in Factor 1-Falls (F1), six items in Factor 2- Blood and Blood Products Transfusion (F2), six items in Factor 3-Medication Practices (F3), eight items in Factor 4-Care Practices (F4), five items in Factor 5-Communication, and
six items in Factor 6-Other Controlled Practices (F6) (Tables 1). A score close to 215 suggested the nurse was disciplined or cautious about medical errors, whereas one close to 43 indicated that he/she was not careful about medical errors or was at risk of making medical errors. Scores were divided by the number of items to facilitate comparisons, which yielded a value between 1 and 5 for the overall scale and the subscales.

**Discussion**

The development of valid and reliable measurement tools for many topics is crucial to achieving consistent and accurate data. For this reason, a valid and reliable instrument in the study was tried to develop to determine whether nurses have made medical errors.

For the scale, the face validity and content validity were first studied. Face validity refers to the extent to which a scale appears to assess the notion being studied (23). It also involves an analysis of the scale’s legibility, comprehensibility of the terminology, and length of the statement (22). In this study, three nurses who were within the immediate circle of the researchers were asked to test the comprehensibility and length of the items. The next step was to assess the content validity of the scale, in which a group of experts were asked to test both the content validity and face validity of the scale.

In the content validity tested with Lawshe’s technique, nine items were excluded from the 94-item draft scale, and the 85-item scale had a content validity index (CVI) of 0.82; this finding suggested acceptable content validity, because a CVI of 0.80 and higher is considered acceptable (23,24).

Afterwards, the construct validity of the MESN was tested with exploratory factor analysis and confirmatory factor analysis. The objective of factor analysis is to determine the dimensions that account for specific constructs (25). A prerequisite for factor analysis is having a certain amount of correlation between the variables (26). Bartlett’s test is used to determine whether variables are sufficiently related to each other. If the $P$ value of this test is less than the level of significance 0.05, the correlation between the variables is sufficient for factor analysis (26). The factor analysis of the 85-item scale showed that the KMO, a value used to determine whether the selected sample is adequate for factor analysis, was above the acceptable limit, and Bartlett’s test was highly significant. A KMO higher than 0.50 suggests that one can conduct factor analysis; therefore, the acceptable lower limit of the KMO is 0.50 (22,25,27). The KMO measures whether the overall group of questions is suitable for factor analysis. On the other hand, anti-image correlation is a coefficient that tests whether each item/question is suitable for factor analysis, and this coefficient should not be less than 0.50. If the value of any of the items is less than 0.50, it is recommended that it should be excluded from the analysis (22,26,27). The anti-image correlation coefficient of the draft scale was greater than 0.62. These results indicated that factor analysis could be conducted. The next step was to perform Varimax rotations; as a result, 42 items were excluded from the scale because their factor loadings were too low (less than 0.45). Ideally, factor loadings should be between 0.45 and 0.50 or higher (25).

In brief, 42 items were excluded from the draft scale, and the finalized version consisted of 43 items in six subscales: falls, blood and blood product transfusions, medication practices, care practices, communication, and other controlled practices. The six-factor structure was confirmed by the
explained variance and by the fact that the scree plot line plateaued after the sixth point (20). In brief, the six subscales accounted for more than 0.50 of the total variance, which was acceptable in practice (27).

The exploratory factor analysis was followed by confirmatory factor analysis, which was conducted to confirm the constructs that emerged. Confirmatory factor analysis usually follows conventional factor analyses or exploratory factor analysis, and it tests whether a previously defined or restricted construct can be confirmed as a model or not (27). Alternatively, it tests the accuracy of a correlation that was previously determined by the researcher (24). All items with significant t-test values were confirmed by the CFA to be significant. Then, the fit indices were used to test the validity of the model. One of the most common fit is the chi-squared goodness-of-fit ($\chi^2$) (20). This study had an index of 2.5, suggesting that the items fit the subscales well. In studies with large samples, a chi-squared index less than three represents an excellent fit (20). The RMSEA of the path scheme was 0.72, suggesting that the scale had good fit. An RMSEA less than 0.5 represents excellent fit, whereas one less than 0.8 indicates good fit (20,26). The CFI and IFI are two other fit indices. The CFI and IFI are known to produce very reliable and impartial predictions when the assumption of a normal distribution is not violated. A CFI and IFI higher than 0.95 indicates excellent fit, whereas one higher than 0.90 represents good or acceptable fit (28). In this study, the CFI and IFI were 0.90, which suggests good or acceptable fit. In brief, all the observed variables in the model indicating the factor structure of the scale and the coefficients of the correlations between the factors were sufficient. Considering the fit indices calculated with the CFA, the construct of the scale was generally consistent with the data.

The factor analyses were followed by internal consistency and item-discrimination analyses. The difference in the mean scores between the upper and lower 27% of the sample showed that the items had significant discriminating power and could appropriately distinguish between the two groups. In other words, the items were highly valid, could appropriately distinguish between nurses’ erroneous medical practices, and measured the same behaviour.

The next step was to measure the internal consistency of the 43-item scale to test its validity and determine its homogeneity. Using a single measurement instrument in one session, internal consistency analyses attempt to determine whether items can consistently measure a certain construct (22). In the present study, Cronbach’s alpha, Spearman-Brown’s coefficient, and Guttman’s coefficient were all 0.70 or higher for the overall scale. The higher these values, the more consistent the items are and the better they can measure the same property (22). An internal consistency coefficient of 0.70 and higher is considered sufficient for the reliability of the test scores (22,25).

Item–total correlation refers to the correlation between the score of an individual item and the overall score of the test (22). In this study, the item–total correlation coefficients of the 43-item scale were above 0.31. Certain limit values are accepted to represent standards for interpreting item–total correlation analysis. It is reported that items should have an item–total correlation coefficient of 0.30 and higher, as these items can discriminate well between individual items (22). As the coefficients were high in this study, the items belonged to the same construct, and the overall scale was reliable.
According to the results of the t-test, which determines whether a property measured by a test changes over time and how consistently the test measures a construct or how similar answers it obtains independent of time (21), the scale yielded consistent and reliable results when administered at different times and was thus reliable in terms of the coefficient of continuity.

**Conclusion**

The results of the reliability and validity analyses suggest that the scale had both face validity and content validity. The 43-item the final scale is reliable according to the results of reliable tests. In addition, the scale consisted of 43 items and in six subscales, namely falls, blood and blood product transfusions, medication practices, care practices, communication, and other controlled practices. Furthermore, the six-factor structure revealed by the exploratory factor analysis was supported by the confirmatory factor analysis. In fact, the DFA fit indices suggested that the scale had good fit overall. All these suggest that the scale is valid. This scale can thus accurately and consistently measure whether nurses are cautious about medical errors, in which areas they are more likely to have problems, and in which areas they need to improve. In addition, this study can be used as a guide or reference for future studies on scale development and can provide opportunities for comparison with previous studies on scale development.

**Funding:** None.

**Competing interests:** None declared.

**References**


Table 1. Distribution of the items in the MESN by mean values and factor loadings

<table>
<thead>
<tr>
<th>Factor</th>
<th>Variance %</th>
<th>Items</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
<th>Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1-Falls</td>
<td>11.71</td>
<td>49-51, 55,59-66</td>
<td>2.17</td>
<td>5.0</td>
<td>4.40</td>
<td>0.52</td>
<td>0.67-0.51</td>
</tr>
<tr>
<td>F2- Blood and Blood Product Transfusions</td>
<td>9.19</td>
<td>42-47</td>
<td>3.0</td>
<td>5.0</td>
<td>4.83</td>
<td>0.35</td>
<td>0.85-0.59</td>
</tr>
<tr>
<td>F3- Medication Practices</td>
<td>8.66</td>
<td>21-26</td>
<td>2.33</td>
<td>5.0</td>
<td>4.79</td>
<td>0.37</td>
<td>0.80-0.60</td>
</tr>
<tr>
<td>F4- Care Practices</td>
<td>8.46</td>
<td>1-5,7, 11,12</td>
<td>1.75</td>
<td>5.0</td>
<td>4.57</td>
<td>0.44</td>
<td>0.78-0.49</td>
</tr>
<tr>
<td>F5-Communication</td>
<td>6.87</td>
<td>76,78-81</td>
<td>2.60</td>
<td>5.0</td>
<td>4.76</td>
<td>0.40</td>
<td>0.74-0.53</td>
</tr>
<tr>
<td>F6- Other Controlled Practices</td>
<td>6.67</td>
<td>15-17, 29,30,39</td>
<td>1.67</td>
<td>5.0</td>
<td>4.62</td>
<td>0.28</td>
<td>0.79-0.47</td>
</tr>
<tr>
<td>Total</td>
<td>51.58</td>
<td></td>
<td>2.99</td>
<td>5.0</td>
<td>4.49</td>
<td>2.83</td>
<td>0.85-0.47</td>
</tr>
</tbody>
</table>