Creation of complexity assessment tool for patients receiving home care

Elaboração de instrumento de classificação de complexidade assistencial de pacientes em atenção domiciliar

Diseño de instrumento de clasificación de complejidad asistencial de pacientes en atención domiciliaría

Maria Leopoldína de Castro Villas Bôas¹, Helena Eri Shimizu², Mauro Niskier Sanchez²

How to cite this article:

ABSTRACT
Objective: To create and validate a complexity assessment tool for patients receiving home care from a public health service. Method: A diagnostic accuracy study, with estimates for the tool’s validity and reliability. Measurements of sensitivity and specificity were considered when producing validity estimates. The resulting tool was used for testing. Assessment by a specialized team of home care professionals was used as the gold standard. In the tool’s reliability study, the authors used the Kappa statistic. The tool’s sensitivity and specificity were analyzed using various cut-off points. Results: On the best cut-off point—21—with the gold standard, a sensitivity of 75.5% was obtained, with the limits of confidence interval (95%) at 68.3% and 82.8% and specificity of 53.2%, with the limits of confidence interval (95%) at 43.8% and 62.7%. Conclusion: The tool presented evidence of validity and reliability, possibly helping in service organization at patient admission, care type change, or support during the creation of care plans.

DESCRIPTORS
Home Care Services; Home Nursing; Validation Studies; Reproducibility of Results; Sensitivity and Specificity.
INTRODUCTION

Home care (HC) in Brazil’s public health system is still a very recent type of care, with great potential for implementation and growth, because after studies in trials for home care services were already in place in the country(1), the Ministry of Health launched the Programa Melhor em Casa (Better at Home Program), home care in the realm of the Unified Health System (SUS), in 2011, which is regulated under Ordinance no. 963, of 2013(2). It reaffirms HC as a technological incorporation that can substitute for or complement hospital interventions of low and medium complexity and care that start at urgency and emergency care services. It can also complement basic care(3).

This ministry ordinance defines three types of HC according to patients’ clinical profiles, the frequency of home visits, and the composition of the health team responsible for administering care:

HC 1: Users that require low-intensity care, with regular visits from the basic health team.

HC 2: Users that require higher intensity care, with at least weekly visits.

HC 3: Users that fall under HC 2’s criteria with the addition of noninvasive ventilation support, paracentesis, or peritoneal dialysis.

HC2 and HC3 patients are the target demographics of Home Care Services under SUS.

Therefore, HC in SUS requires the incorporation of new care technology(4), which places the challenges of comprehensiveness and continuity of care—both in relation to patients, caregivers, and families and to their care network—in a new noninstitutionalized practice field: the home(5,6).

In parallel to this context, current epidemiological tendencies show that the Brazilian population is quickly aging(7). Increasing operational costs related to medicine’s technological development and the need for changes in the current model of technological care (defined by low rationality and efficiency, among other factors), show that HC can be a very consistent alternative of care, especially in relation to hospital stays(8).

It is very important to consider possibilities for increasing coverage at lower costs and higher effectiveness is very important when it comes to patients, caregivers, and families, as well as in relation to public expenditure and resource management(9).

The creation of a tool that makes it possible for services to objectively and easily assess the complexity of each case will give health teams better ways to estimate resource use; therefore, it offers a clearer view of admission capacity and better conditions for prioritization.

Analysis of the literature addressing patient assessment tools showed that various tools have been developed in hospital units and, later, in intensive care units and clinics for allocation of nursing personnel, that is, the Patient Assessment System (SPC). This system evaluates the extension of the patients’ dependency on nursing care and the assessment of workload, considering the severity variable as directly related to the number of therapeutic interventions and hours of care required(10-13).

When analyzing instruments for the assessment of patients receiving home care in other countries as well as in Brazilian trial runs(14-20), the authors found considerable diversity of models and organizations in home care in the last few years, in addition to various study designs and sample characteristics. However, this analysis does not give a complete understanding of the main aspects of home care in those countries and in those services, especially public HC services, that are defined mainly by their humanized and inclusive nature, with user-centered care plans and multidisciplinary health teams.

The aims of this study were to elaborate and validate a patient assessment tool according to type of care (HC1 or HC2/HC3) for home care services.

METHOD

This is a diagnostic accuracy study(21), conducted in the Home Care Program of the State Health Secretariat of the Federal District (PID–DF).

TOOL CREATION

Monthly workshops were conducted in the period between 2009 to 2011, by professionals who are specialists in the area and who have vast experience with the definition of the main procedures conducted by multi-professional teams. These professionals helped the author in an analysis of time spent on home care by multi-professional teams as a basis for the allocation of personnel(22).

Following this, patients who were registered in the program were profiled according to parameters considered essential for complexity assessment, such as: cardiorespiratory pattern; nutritional status; hemorrhage risk; infections; professionals involved; types of procedures conducted; number of checkups; and degree of efficiency of caregiver/family. These parameters were based not only on procedures, but on patients’ clinical status, on the attributes of the multi-professional health teams (defined here as physician, nursing personnel, nutritionist, and physical therapist), and on the performance of caregivers/families (Chart 1).

Chart 1 – Profile of moderate (HC2) and severe (HC3) patients of the PID–DF, according to the selected criteria – Brasilia, Federal District, 2011.
Analyzing the list of the main procedures, the time spent by the health team to perform them, and the selected criteria for patient severity according to the particular aspects of the home care service, the authors found in the Nursing Activities Score (NAS) the necessary outline to format the tool.

The NAS is an internationally validated scale to assess the workload of nurses in quality care in intensive care units with seven domains and 23 activity categories: basic activities (monitoring and controls, hygiene, mobilization, support for patients and family members, and managerial and administrative activities); supports (ventilatory, cardiovascular, renal, neurologic, metabolic); and specific interventions\(^{23}\).

The biggest difference between NAS and previous tools was the inclusion of activities related to patient status. It was not limited to applied therapeutics and to activities that were excessively time-consuming for professionals; instead, it defined fewer activities in order to be wide-ranging and easily applied to services.

The NAS was adapted by expanding its activities to accommodate a multi-professional team, keeping them in the domain of “basic activities,” which was renamed “home activities.” The authors eliminated the following activities: “hygiene,” because it is conducted by caregivers and/or family members; and “supports” and “specific interventions,” because they are part of the ICU environment. The following activities were added: “therapeutic procedures”; “laboratory investigations;” and “rehabilitation.”

Thus, the instrument had two domains: “Home activities” and “Managerial and administrative activities” and a total of 20 activities.

Each activity was rewritten and operationally defined according to the main activities selected from the home care service, as well as their profiles; the score was calculated according to measurements of time spent in procedures conducted at home\(^{21}\) and the estimated time for administrative and managerial activities. A score of “1” was given to procedures that lasted from 0 to 9 minutes, a score “2” for activities that lasted from 10 to 19 minutes, and so on.

After the group of professionals/specialists reached a consensus on the tool, its first version was finished. It was pre-tested by being applied to 18 medical records of patients registered at the PID-DF, chosen at random. With some adjustments, the final version of the tool was achieved.

### Data Collection

Inclusion criteria for study participants were the same as admission criteria for the program: patients with chronic degenerative diseases that became acute; patients with sequelae and comorbidities; patients receiving palliative care; patients with functional disabilities for activities of daily living, temporary or permanent; and patients in clinical stability.

Exclusion criteria were: patients requiring invasive mechanical ventilation; those under continuous monitoring; those with intensive nursing and using complex medication with potentially severe collateral effects or with difficult administration.
Thus, the research's reference population consisted of 826 active patients of the PID-DF, Primary Care, Home Care Service, State Secretariat of Health of the Federal District, in the periods of July and August 2012 and June to July 2013, who were empirically defined as HC1, HC2, and HC3—in other words, without scales, based only on the experience and consensus of at least three PID-DF professionals.

For statistical calculation of sample size, the authors initially applied the assessment tool to a pilot sample of 96 patients (2012), aiming to estimate the proportion (sensitivity) who were empirically categorized as HC1, HC2, or HC3.

Considering a level of significance of 5% and a margin of error of 5%, the authors estimated a final random sample of 252 patients, who were categorized by empirical care type and by the coverage area of the PID-DF; the pilot sample was added to the final sample size.

Two PID-DF examiners (Examiner 1 and Examiner 2) were chosen and trained to apply the tool to patients who were selected inside their coverage area, during their work shifts, and without interfering in their service routine.

The criteria chosen to select examiners were: PID-DF professionals only, with the appropriate profile, and available to apply the tool in the homes of patients chosen inside their coverage area, who were either volunteers or appointed and who formally accepted by signing an informed consent.

The following standardization of procedures was applied during the examiners' training, with the goal of correctly applying the tool for statistical validity and reliability:

1. Each examiner scored tool items only when completing the tool in the selected patients' homes and did not categorize them because they did not know the cut off points of the tool (single blind assessment).

2. The first application of the tool was conducted by two examiners independently and separately, preferably during the same visit; after completing the tool, each examiner placed their sheet in an envelope without commenting on or seeing what the other examiner did.

3. The second application was conducted only by Examiner 1, with an interval of between four and seven days from the first application, because it was empirically observed by the team that the evolution of patient status does not initially change enough to alter their tool-generated category.

Sample randomization and masking of tool application were incorporated into the study design with the goal of eliminating external variations and minimizing sampling bias.

Supervision was performed remotely, with an agreement for weekly email exchange with the researcher. This communication displayed the number of visits that were scheduled, and completed or not completed because of refusals; discharge; death; hospital admission; change of address; and patient or residence not found after three scheduling attempts. In these cases, it was expected that the original patients would be replaced by other patients with the same profile until reaching a situation of “loss,” that is, when there would be no patients left in the area with the same profile in the existing databases.

Excel spreadsheets were employed for data processing; printed tools were typed and checked after delivery.

**Validity and reliability**

Sensitivity and specificity measures were considered when estimating validity; the resulting tool with scores was used for testing and the gold standard was the categorization of HC types performed by the specialized health team (HC1: 13 to 24 points; HC2: 25 to 30 points; HC3: 31 to 36 points).

Sensitivity was defined as the percentage of patients categorized as HC2/HC3 (the target demographic of HC) by the tool score and by the reference test, that is, the gold standard (true positive). Specificity was defined as the percentage of patients not categorized as HC2/HC3 by the tool score and by the reference test, that is, the gold standard (true negative).

For each tool score, researchers measured sensitivity and specificity in order to draw a receiver operating characteristic (ROC) curve with the respective 95% confidence intervals. This was done to find the optimal cut-off point.

The ROC curve is a type of graphic that estimates the best balance between sensitivity and specificity, based on the cut-off points of the instrument assessed. The curve makes it possible to evaluate the discriminatory power of the test. For tests with good discriminatory power, as sensitivity increases, specificity slightly decreases or remains unchanged until high levels of sensitivity are reached.

In order to employ the tool in service practice, researchers opted to take into consideration only two categories: HC1 and HC2/HC3, because most of the demand is centered on the two profiles that outline the specific action of primary care teams and home care teams, respectively.

When studying tool reliability, the Kappa statistic was employed to measure result agreement among examiners and among the same examiner's results, observing the proportions obtained in the first and second applications of the instrument. Researchers opted for the weighted Kappa calculation. The chosen weight system was square weighted deviation. For interpretation, the Landis and Koch agreement scale was chosen.

The software SAS 9.3 was used for statistical analyses, with a significance level of 5%.

The research project was approved the Research Ethics Committee of the Foundation for Learning and Research in Health Sciences of the State Health Secretariat of the Federal District, Project No. 392/2008. All examined patients and/or caregivers and family members signed a free and informed consent form.

**RESULTS**

In the first stage of study (2012), the tool was applied to the 96 patients in the random sample; in the second stage (2013), it was applied to 157 patients, with a total of 253 patients. Of these, 10 patients were excluded during the validity study (3%); 29 patients were excluded during the reliability assessment for the same examiner's results at two different times (11.4%); and 13 patients were excluded during reliability assessment among different examiners at the same time (5%).

The tool’s validity was supported by Examiner 1’s score in the first assessment, in which it was found that of the 123
patients empirically classified as HC1, 36 (29%) scored less than 13 points and, of the 33 patients empirically classified as HC3, three scored above 36 (10%), both in relation to the established gold standard.

The ROC curve was constructed based on measurements of sensitivity and specificity calculated for each score given by the tool (Figure 1), with the following cut-off points (the area under the curve was equal to 0.694 with CI [95%] limits between 0.627 and 0.760):

![ROC Curve](image)

**Figure 1** – ROC Curve with some cut-off points marked—Brasília, Federal District, Brazil, 2013.

The sensitivity and specificity of the tool were analyzed by considering different cut-off points. The best balance, as provided by the ROC curve, between these two classifications and the gold standard at a cut-off score of 21 (Table 1).

**Table 1** – Best cut-off point with gold standard defined by the ROC Curve – Brasilia, Federal District, Brazil, 2013.

<table>
<thead>
<tr>
<th>Care type</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (IC 95%)</td>
</tr>
<tr>
<td>HC2/HC3 (&gt;21)</td>
<td>75.56</td>
</tr>
<tr>
<td>HC1 (≤21)</td>
<td>46.73</td>
</tr>
</tbody>
</table>

The following results were found in the reliability study (Table 2).

**Table 2** – Weighted Kappa test for assessment of the tool’s reliability – Brasilia, Federal District, Brazil, 2013.

<table>
<thead>
<tr>
<th>Agreement</th>
<th>KAPPA index</th>
<th>Variation (IC 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same examiner</td>
<td>0.81</td>
<td>0.74–0.88</td>
</tr>
<tr>
<td>Different examiners</td>
<td>0.52</td>
<td>0.43–0.61</td>
</tr>
</tbody>
</table>

**DISCUSSION**

The results of this study showed that the tool developed to assess patient complexity in HC presented evidence of both validity and reliability.

Thus, the tool may help in service organization, either on patient admission or when changing care type and in supporting the creation of therapeutic plans, contributing to define the frequency of visitations from health team professionals, transport logistics, and necessary appointments.

The authors observed that 29% of patients assessed as HC1 scored below 13 points, which is the minimum score as defined by the gold standard. This fact may indicate patients who do not have the due care complexity to be monitored at home. A relevant portion of patients undergoing oxygen therapy who use the equipment only in the evening or for some hours during the day may be included in this profile.

There were also scores that exceeded the maximum limit of the HC3 category (10%), which suggested above average care complexity, represented by patients with sequelae and/or more severe comorbidities who require more frequent and longer care, especially in the presence of multiple pressure ulcers of various degrees, palliative oncological care, and difficulties with caregivers/families, among other factors.

In relation to the validity study, the identification of the best cut-off point for the tool—21—took into consideration the maximization of sensitivity for HC2/HC3, because that is the target demographic of HC. Its sensitivity was 75.5%, with CI [95%] limits at 68.3% and 82.8% and specificity of 53.2%, with CI [95%] limits at 43.8% and 62.7%), which can be considered an adequate set of parameters for the objective of this evaluation.

This assessment considered that it is better for the home care service to care for more patients in the HC2/HC3 profile and fewer patients in the HC1 profile. Caring for more of the HC2/HC3 profiles patients would result in absorbing the most severe cases—who would truly require attention from a multi-professional home care team, than those who would not present this care complexity and who could be absorbed by primary health teams.

Therefore, cut-off point 21 was chosen to differentiate between HC1 patients, who reached that score, and HC2/HC3 patients, who scored above 21.

The main limitation to the use of these estimates for assessing validity through sensitivity and specificity is the existence of a “gold standard” to which test results are compared, with the possibility of the true status of the patient being, if the information is available, a collection of exams considered more adequate, or other types of diagnoses that serve as references, as was the case with this research.

Despite this limitation, the literature includes many validity studies for scores based on gold standards that use, for example, clinical descriptions for early detection of pneumonia caused by *Mycoplasma Pneumoniae* in order to initiate antibiotic therapy, or tools that help parents to detect pain in their children in the postoperative period at home so that they can decide whether analgesic medication should be administered and even investigation of the contacts of individuals exposed to tuberculosis to validate a predictive score to develop active tuberculosis, among others.

Reliability analysis had satisfactory results for same-examiner agreement (Kappa coefficient 0.81), which is a...
general trend already observed, because it is a comparison among the same examiner’s scores, which were obtained by visiting the same patient with an interval of between four to seven days. For different examiner agreement, the authors estimated a Kappa coefficient of 0.52 (moderate), which may suggest differing interpretations among examiners when assessing the tool’s items. This may indicate the need for more thorough training and more time to discuss and address doubts about the tool’s application.

CONCLUSION
The HC complexity assessment tool has been shown to be a useful tool, suitable for the routines of multi-professional HC teams, because it enables validated, reliable, and self-explanatory categorization of care type for HC patients, which makes it possible to improve service organization and care quality. The authors believe that continuing training, monitoring, and assessment of the tool are crucial for its improvement in the practice of HC services.


