6th European Conference of ACENDIO, Amsterdam 2007
Paper presentation 2007 April 21st

Clinical testing of ePA©-AC, a screening instrument to assess relevant clinical indidators of care dependency in acute care clinics

Dirk Hunstein, Madlen Fiebig, Birgit Sippel, Yvonne Dintelmann

Since several years there has been a consensus on the need for standardised data in nursing in order to prove its success and to calculate and to explore it. At this nursing diagnoses are the decisive factor.
However, valid nursing diagnostics is ineffective without the use of assessment or screening instruments (e.g. Bartholomeyczik, 2003; Ehrenberg & Ehnfors, 1999; Gordon, 1994). Unfortunately, appropriate instruments developed for acute care are still missing to date.
Therefore, the Department for Nursing Research and Development at the HSK, Dr. Horst Schmidt Klinik in Wiesbaden, Germany, has developed a screening instrument referred to as ePA-AC. The acronym ePA-AC stands for the German *ergebnisorientiertes PflegeAssessment AcuteCare*, which can be translated to outcome-oriented nursing assessment Acute Care.
The ePA-AC has been developed to record the essential aspects of nursing care requirements in the acute care setting and to measure their degree of severity.
In its function as a screening instrument it collects attributes or clinical indicators of nursing diagnoses, such as signs and symptoms. These attributes are prestructured and clustered by means of triggerpoints of the ePA. Supplemented by etiology the differentiated diagnosis is made. Thus the ePA-AC is the first step in the diagnostic process – whatever diagnosis classification system is used.

Patient abilities are quantified in the ePA-AC screening procedure by means of numeric values so that nursing sensitive outcomes can be measured and the subsequent processes can be controlled.
Furthermore, ePA-AC data can be used in terms of a Nursing Minimum Data Set as economic and epidemiological operating figures.
Now, regarding a few words about the structure:
As shown in this slide, the ePA-AC consists of a total of 50 items in 10 categories. This involves measuring primarily the scope of functional and cognitive abilities, but it also

includes bodily functions such as continence or context information such as the presence of a urinary catheter.
These different levels of the ePA-AC can be readily mapped using the International Classification of Functioning, Disability and Health (ICF).
Seventeen 17 of the Items contain data on activities and participation, 17 items are related to body functions and body structures.
These informations are supplemented by 16 items from the domain of context structures.

The examples I have selected here from the category "elimination" should make these relationships clear.
The self-care ability to perform toileting activities has been selected from the domain Activities and Participation. The degree of ability is measured along a four-point scale, point 4 standing for complete independence and 1 for complete dependence. Points 2 and 3 are correspondingly in between.
At this juncture it is important to mention that the term "self-care abilities" (abbreviated in the slide as SCA) does NOT refer to Orem's self-care deficit theory but generally to the degree of ability to perform activities on one's own. This decision was made to permit the ePA-AC to be used in other theoretical contexts as well.
The degree of urinary continence which is in the Body Functions domain is also documented on a 4-point scale.
The context information "presence of an urinary catheter" is documented dichotomously.

To enable you to get an impression of how the ePA-AC can look in a software application I have brought along a screenshot. Here you can see the first three categories along with their items and the integrated risk indicators for pressure sore, falls, pneumonia and post-discharge care deficit.

The curve you see in the upper right reflects the course of the functional abilities of the patient. This visualization is viewed by caregivers as especially helpful in their work because it enables them very easily to get a rapid overview of the successes of the care.
Moreover, each item of the ePA-AC and it's score is linked with the corresponding interventions from the workload measurement system LEP Nursing 3 so that the appropriate interventions are directly recommended. This is experienced as a major aid in care planning.

But what good is an instrument that is seen by caregivers in practice as practicable but does not even meet the most basic scientific quality criteria?
Hence, the ePA-AC was scrutinized to determine how well it met the following quality criteria: interrater reliability, construct validity, and sensitivity to change.

Let us start with the interrater reliability.
For the testing of interrater reliability more than 230 ratings were carried out, during which it was possible to include 115 valid rating pairs into the analysis.
The degree of the agreement between the ratings was measured using Cohen's weighted kappa respectively Cohen's quadratically weighted kappa as well as in observed agreement.
The sampling was purposive because a uniform distribution of the severity of the patients' nursing care dependency was the intended goal. This occurred under the assumption that the degree of severity of nursing care dependency has an influence on the level of interrater reliability.
The analysis of the data in fact did reveal indications of this sort of influence. But this remains to be verified.
Collecting the data proved to be quite difficult because the application of the ePA-AC requires that the patient is known and a physical examination has been conducted. But at the same time there must also be an assurance that the paired rating by the two raters take place within a short period of time to avoid any influences caused by a change in the condition of the patient.
Hence, the following approach was set down: during the first shift, one caregiver (rater 1) completes the ePA-AC-form shortly before the shift change. Rater 2 then rates the patient shortly after the shift change. In order for rater 2 to have sufficient knowledge about the patient, she must have cared for the patient on the previous day so that she need only visit the room briefly to see whether the patient's condition has changed or not.

The analysis merely studied the agreement or non-agreement between the rating pairs. There was no scrutiny of the technical quality of the rating.

I would like to show you a few selected results:
Overall, a high degree of agreement was found, only four of the scaled items showed an observed rate of agreement of less than 75%, one of them being pain intensity which is rated by the patients themselves using a numeric rating scale.

The kappa values range from 1.0 for obvious circumstances such as the presence of a tracheostoma .267 for the assessment of an altered sleep-wake cycle.
For 32 items the kappa value was over.7, for 23 it was even over .8.

I would like to single out two items:
Firstly: the item "awareness and vigilance". The value of the quadratically weighted kappa is at .577 in the middle. If one looks at the observed agreement, the rate of observed agreement is over 92%.
This phenomenon reveals one of the major difficulties in the interpretation of Cohen's kappa. For if one observes the underlying data more carefully, one finds that for 106 of the total of 115 rating pairs there was agreement in the recording of no disabilities, which corresponds precisely to the 92.2% observed agreement. None of those nine ratings for which at least one of the raters coded a change in the level of consciousness agree. For the non-agreeing ratings only slight disabilities were recorded with one exception.
This clear ceiling effect impairs the validity of the test values for this item because its possible variance width was not at all exhausted. In a more heterogeneous sample other results would come to light.
But, for us, this also means that in the coding manual the borderline between the conditions "awake" and "somnolent" must be more sharply drawn.
Secondly: sleep. The assessment of sleep was especially difficult for caregivers as is revealed both by the low kappa values and by the generally low observed agreement. For example, an impaired ability to fall asleep was surveyed in more than one third of cases, specifically 41. Of these 41 cases of impairment, however, only 16 of them were rated in agreement.
These poor values are possibly influenced by the time of day of the survey, around lunch time.
In my opinion, however, the result is more an indication of how the subject of sleep in acute care hospitals assumes a generally subordinate role and only becomes an issue if, for example, an altered wake-sleep cycle disturbs work routines. However, our sample does not permit any statements about this.
I would like to briefly discuss an interesting incidental finding of the analysis. The tested ePA-Beta version differs in some items from the ePA-Alpha introduced at the beginning of the instrument development. Although the new items were taught [to staff], this subsequent training was not as intensive as the original training.

During the analysis of the agreements it turned out that the new items performed somewhat more poorly than the old items, and the difference is significant. This indicates the importance of training, which is something frequently neglected during the introduction of standardized instruments.
After it was shown that most ePA-AC items are rated similarly by different observers, the point now is to test whether the ePA-AC measures at all what it is supposed to measure, specifically the signs and symptoms of nursing care dependency and, thus, in the end, the triggers for nursing care interventions.
This brings us to studying aspects of validity.

The study considered aspects of construct validity, sensitivity to change and prognostic validity. In this presentation I will consider the first two aspects of validity in greater detail.
The design of the validity testing was quantitative, prospective, non-experimental and multicentred. During March and June 2006, a comprehensive set of data was collected in a full sample of all patients in four wards of our Hospital.
Added to these were two interdisciplinary wards at the Cantonal Hospital Uri (Switzerland).
The data stem from both the anonymised ePA-AC documentation as well as other data records of the electronic patient documentation. The sample comprises more than 1.400 cases for the validity tests.

Unfortunately, to date, the construct "nursing care dependency" has not been clearly defined, hence there is no gold standard against which one could measure it. For this reason, the construct validity of the ePA-AC ought to be tested by means of hypotheses as is recommended, for example, by Streiner & Norman (2003).
To begin with, I have brought two examples of convergent validity. What is tested here is a conjectured relationship between theoretical groups, which are assumed to be related to care dependency, and ePA-AC data.
As I showed you earlier using the example of urinary elimination, various dimensions are measured in the ePA. For the analyses shown here only those items were included which have a close relationship with functional and cognitive self-care abilities. These items are compiled in a total score which we call the "CaseManagementScore", abbreviated as CMS. Individually, CMS-Items are related to: mobility/locomotion, eating and drinking, managing urination and defecation, washng and dressing oneself, as well as acquiring new knowledge.

The following hypothesis was tested:
Patients who have cared for themselves at home show significantly higher (this means "better") values in the ePA-AC CaseManagmentScore than patients who were cared for at home.
In the box plots you can see on the ordinate the height of the CaseManagementScore at the time of admission. In this chart the higher the value, the more independent the patient. Hence, a patient with 40 CMS points is completely independent, a patient with 10 CMS points is completely dependent on assistance from others. As you can see from the box plot, the four groups clearly differ from each other.
The Kruskal-Wallis Test for independent samples indicates with a probability of over 99% that the differences are factually significant.

A further hypothesis for testing the convergent validity was:
"There is a significant correlation between the ePA-AC score values and the time expenditure for nursing interventions triggered by nursing care dependency".
A linear regression analysis of ePA-AC and LEP data was conducted for the testing. For this purpose, the interventions of the LEP were divided into two groups:
Firstly, the group of interventions triggered by the impairments of self-care abilities such as: giving support during food intake, assistance with dressing, etc.
Secondly, the group of interventions triggered by medical diagnosis and therapy such as caring for an operation wound, replacing surgical dressings or infusions, etc.
The analyses which I would now like to show you are based on current data which were analyzed just last week.

As you can see from this chart, there is a clear connection between the extent of nursing care dependency – measured with the CMS – and the time spent on care interventions – measured via LEP – in response to nursing care dependency. The variance in the number of minutes spent on care can be explained by as much as 60% by the CMS. The beta coefficient, the value indicating by how many minutes the time is shortened in relation to lesser impairments, stands at nearly minus 6.5, i.e. for every increase of one CMS point, the time expenditure is reduced by round about six and a half minutes.

Regarding discriminant validity, the following hypothesis was tested: „There is little correlation between the ePA-AC score values and the time expenditure of the interventions triggered by medical diagnostics and therapy".

Here too, as in the previous slide, there is a relationship between the minutes, here spent on medical therapy and diagnostics, and the degree of nursing care dependency. But this relationship is negligible, the beta coefficient is down to merely just under a minute.
The CMS is not suitable in this context to explain the variance in the minutes, the value of R^2 is not even 6%.
One obvious interpretation: Nursing interventions which are provided due to medical diagnosis or therapy are related to other dimensions than care dependency. The occasions for these interventions can be charted, for example, with medical diagnosis classifications, i.e. the ICD-10.

Let us turn now to the third aspect of clinical testing, specifically to the study of sensitivity to change.
In contrast to the medical diagnoses which generally remain stable for a certain period of time, nursing diagnoses are designed to respond to change. They do not describe illnesses but rather changing problematic situations (Gordon & Bartholomeyczik, 2001; Hunstein, 2002). Hence, an instrument designed to quantify nursing care dependency and for use as a support for nursing diagnostics must be capable of measuring such changes.

The assumptions on which the test of the sensitivity to change was based were:
First: A surgical intervention following an accident ("external criterion") influences the self-care abilities. This change generally leads to lower post-operative values of the abilities which are compiled in the ePA-AC in the CaseManagementScore.
Second: Since the functional impairments which are caused by an emergency surgical intervention are generally reversible, the patient's abilities can be expected to improve again after the operation. Hence, the CMS values at discharge must be higher than at the time of the first post-operative rating.

In a first step, the assumption that a surgical intervention in fact improves the odds ratio of a change in a patient's abilities was tested.

To this end, the extent of the change in self-care abilities was quantified for all patients by determining the difference between the highest and lowest measured CMS.
Then, all the patient cases were assigned to one of two groups. Group one includes all those patients for whom an accident-related operation was performed—that is the intervention group. Patients who were not operated were assigned to Group two—the control group.
Then a logistical regression analysis was conducted to calculate by how much the odds of operated patients increase that, in contrast to non-operated patients, they will experience a change in functional abilities.

Result: patients in the intervention group had more than a seventeen times greater chance of a change in their abilities than patients in the control group.
It can be concluded from this that an accident-related surgical intervention does have an impact on patient abilities.

In the next step, then, a test was conducted to determine whether the changes follow the expected patterns:
As shown at this slide, the self care abilities, measured with the CMS, of the first postoperative days are significantly lower than the pre-operative values *(Wilcoxon-Ranksum test* $z = -8.57$, $p < .0001$, $N = 155$*).*
The values for the day of discharge are significantly higher than the postoperative values *(*$z = -9.63$. $p < .0001$, $N = 186$*).*
It can be concluded from this that changes in patient abilities do lead to changed ePA-AC scores.

Any study has its limitations, and this one is no exception.
Despite all the effort to study a sample in which the whole range of abilities and their impairments is more or less evenly distributed, the preponderant share of the patients are those who are largely or completely independent. But even if one excludes patients with little need of help from the analysis steps, a ceiling effect remains. Nursing care dependency is simply not normally distributed so that certain higher quality statistical tests cannot be carried out, or can only be carried out with reservations.
This naturally limits the generalizability of the study.

Furthermore, the quality of the data collection can only be assessed to a limited extent, because—apart from the testing of interrater reliability—this does not involve data which was collected for a study but data which originated from routine documentation. Although we regularly tested the plausibility of the data both on the wards as well as before the beginning of the analysis, this does not replace an independent confirmation of the accuracy of the data collected.

The influence of staffing, work loads and commitment during the collection cannot be ruled out.

But these limits apply to every study which is conducted in the field and not in the laboratory.

Conclusion

The test results show—with few exceptions—that the ePA-AC is a feasible and valid instrument which delivers reliable data for scientists and practitioners.

The intensive development phase of the ePA-AC has shown that it is reasonable to test a new instrument in practice as soon as possible.

This will ensure not only the scientific quality of an instrument but also its practicability and acceptance in nursing practice.

Bartholomeyczik, S. (2003). Pflegebedürftigkeit oder die Schwierigkeiten, einen komplexen Begriff zu operationalisieren. In M. Halek (Ed.), *Wie misst man Pflegebedürftigkeit? Eine Analyse der deutschsprachigen Assessmentverfahren zur Erhebung der Pflegebedürftigkeit* (pp. 7-14). Hannover: Schlütersche.

Ehrenberg, A., & Ehnfors, M. (1999). Patient problems, needs, and nursing diagnoses in Swedish nursing home records. *Nurs Diagn, 10*(2), 65-76.

Gordon, M. (1994). *Nursing Diagnosis: Process and application* (3 ed.). St. Louis: Mosby.

Streiner, D., & Norman, G. (2003). *Health Measurement Scales. A practical guide to their development and use* (3rd ed.). Oxford: Oxford University Press.

Contact:
HSK Dr. Horst Schmidt Klinik
Dept. of Nursing Research & Development
Dirk Hunstein
Ludwig-Erhard-Straße 100
65199 Wiesbaden
Germany
Phone: +49 611 / 43 – 31 79
mail: dirk.hunstein@hsk-wiesbaden.de
www.epa-online.info