

# The 24-h recall instrument for home nursing to measure the activity profile of home nurses: development and psychometric testing

Kristel De Vlieghe<sup>1,2</sup>, Bert Aertgeerts<sup>3</sup>, Anja Declercq<sup>4</sup>, Christiane Gosset<sup>5</sup>, Isabelle Heyden<sup>5</sup>, Michel Van Geert<sup>1</sup> and Philip Moons<sup>1,6,7</sup> for the ASSISTANT-study investigators

<sup>1</sup>Centre for Health Services and Nursing Research, University of Leuven, Leuven, Belgium

<sup>2</sup>Federation of the White and Yellow Cross of Flanders, Brussels, Belgium

<sup>3</sup>Academic Centre for General Practice, University of Leuven, Leuven, Belgium

<sup>4</sup>LUCAS, Research in Care and Consultancy, University of Leuven, Leuven, Belgium

<sup>5</sup>Department of Public Health, University of Liège, Liège, Belgium

<sup>6</sup>Department of Congenital Cardiology, University Hospitals of Leuven, Leuven, Belgium

<sup>7</sup>The Heart Centre, Copenhagen University Hospital, Copenhagen, Denmark

Home health care today is challenged by a shift from an acute to a chronic health-care model, moving the focus of care from the hospital to home-care setting. This increased focus on care at home emphasizes the need for an efficient, effective, and transparent management of home health care. However, it is not precisely known what home-care nurses do; what kind of care is received by patients; what the performance of home nurses is; and what the impact of the increasing need for home nursing is on the current and future role of home nurses. In this respect, it is necessary to gain a clear insight into the activity profile of home nurses, but there is no gold standard to measure their activities. This study reports on the development and psychometric testing of the '24-hour recall instrument for home nursing' to measure the activity profile of home nurses. Five home nurses in Belgium, simultaneously with the researcher, registered the performed activities in a total of 69 patients, using the 24-h recall instrument for home nursing. The validity and the interrater reliability of this instrument were high: the proportions that observed agreement were very high; the strength of kappa agreement was substantial to almost perfect; the prevalence index showed great variety; and the bias index was low. The findings in this study support the validity evidence based on test content and the interrater reliability of the 24-h recall instrument. This instrument can help to shape practice and policy by making the home nursing profession more transparent: a clear insight into the kind of care that is provided by home nurses and is received by the patients in primary care contributes to the development of a clear definition of the role of home nurses in health care.

**Key words:** activity profile; home nursing; interrater reliability; validation; 24-h recall

Received 14 January 2013; revised 26 November 2013; accepted 29 December 2013;  
first published online 31 January 2014

## Introduction

Demographic and economic changes, such as the ageing population; the increased number of

chronically ill patients with more intensive and technically complex nursing needs; the changing preferences of patients; and the financial constraints in health-care budgets nurture a shift of care from the hospital to the home-care setting (Jansen *et al.*, 1996; Mistiaen *et al.*, 1997; Blegen *et al.*, 1998; McCorkle *et al.*, 2000). This evolution increases the focus on health care at home.

Correspondence to: Philip Moons, Centre for Health Services and Nursing Research, University of Leuven, Kapucijnenvoer 35 PB 7001 – BE-3000 Leuven, Belgium. Email: philip.moons@med.kuleuven.be

© Cambridge University Press 2014

The role of home nurses, however, is not well defined, resulting in a large dispersion of role implementations in different countries (Keller *et al.*, 1998; Hughes *et al.*, 2002; Frich, 2003; Robinson *et al.*, 2006). Indeed, it is not precisely known what home-care nurses do; what kind of care is received by patients; what the performance of home nurses is; and what the impact of the increasing need for home nursing is on the current and future role of home nurses. In this respect, it is necessary to gain a clear insight into the activity profile of home nurses. Internationally, research on nursing interventions is mainly conducted in intramural settings. Research in the field of home nursing is much more limited. The few existing studies confirmed a wide diversity of job functions in home nursing (Keller *et al.*, 1998; Hughes *et al.*, 2002; Frich, 2003; Robinson *et al.*, 2006). However, these studies are hampered by the lack of a sound measurement. There is no gold standard in the assessment of home nurses' activities. The measurements that were used had different levels of abstraction in the description of the work of home nurses (Martin *et al.*, 1993; Schumacher and Marren, 2004; De Vlieghe *et al.*, 2005), and they used different classification systems, of which not all appropriately captured the dimensions of physiological, technical, social, and psychological activities of home nurses' practice. Furthermore, validity and reliability of these measurements are uncertain. Hence, to gain insight into the work of home nurses, valid and reliable instruments need to be developed to adequately measure the activities of these nurses. Therefore, we developed a tool to capture the activity profile of home nurses: 'the 24-h recall instrument for home nursing'. The present article describes (i) the development and pilot testing of this 24-h recall instrument and (ii) the evaluation of its content and face validity, and its interrater reliability.

## Methods

### Instrument development

The instrument was developed in three phases. First, to capture the actual everyday work of home nurses, the concept of a 24-h recall was chosen to develop the instrument under study. The 24-h recall is a method that is mainly used in the dietary and food setting to record a person's food and eating habit over a period of 24 h (Conway *et al.*, 2003; Park *et al.*, 2004; Subar *et al.*, 2007). Non-participatory

*Primary Health Care Research & Development* 2015; **16**: 79–86

observations of home nursing rounds and individual interviews at the end of these rounds allowed us to develop a first checklist of activities, which comprised 74 items. Second, this first checklist was evaluated by experts in home nursing and was compared with the activities listed in the nomenclature (reimbursement system in home nursing in Belgium) and the Nursing Interventions Classification (NIC). The comparison of the checklist with the nomenclature resulted in adding more technical activities, such as bladder irrigation, positioning, transfer, and replacement of a suprapubic probe. On the basis of the comparison with the NIC, we subdivided the list of activities into domains (basic physiological, complex physiological, psychosocial support, recording/administration, and communication) and categories (self-care facilitation, elimination management, nutrition support, immobility management, drug and perfusion management, skin/wound management, diabetes care, and other care). Altogether, these adaptations allowed us to make a preliminary version of the 24-h recall instrument for home nursing. In this version, we included a time recording per activity, to assess the time spent per intervention.

This version of the 24-h recall instrument contained three parts: (i) items related to the demographic data of the patient; (ii) items measuring the home nursing activities; and (iii) items related to the demographic data of the participant (home nurse). This preliminary version of the instrument was further evaluated in the present study.

### Psychometric properties

A selection of psychometric properties of the 24-h recall was tested, using the Standards for Educational and Psychological testing as framework (American Educational Research Association, American Psychological Association and National Council on Measurement in Education, 1999): validity on test content and interrater reliability. The terminology as used in the Standards was used in the present manuscript.

#### *Validity: evidence based on test content*

In testing psychometric properties, the validity on test content (content validity and face validity) is fundamental. It refers to the themes, wording, and format of the items, tasks, or questions in the test, as well as to the guidelines for procedures regarding the administration and scoring of the

test (American Educational Research Association, American Psychological Association and National Council on Measurement in Education, 1999). We evaluated the validity on test content by means of non-participatory observations of home nurses, in which we observed and recorded their activity profile. Each home nursing round was concluded with an individual interview to allow the home nurses to add activities that were not included in the instrument. Furthermore, comments on the layout, relevance, clarity, and feasibility of the instrument could be given. In this phase, both items that investigated the activities and the time spent at each activity during working rounds were evaluated. On the basis of the information obtained in this validity testing, adaptations to the instrument were made, and subsequently, reliability was tested.

#### *Interrater reliability*

The interrater reliability was tested to assess the degree to which two individuals would yield the same results. Home nurses and a researcher completed the 24-h recall, simultaneously and independently, for each patient on a same nursing round.

#### **Participants**

Home nurses were eligible to participate in this study if they (i) were Dutch-speaking; (ii) employed in an organization for home nursing in Flanders (Belgium); (iii) employed in the organization for at least five years; (iv) working on a nursing round that reflects the diversity of home nursing activities; (v) agreed to be observed during the nursing round and to be interviewed by the researcher; and (vi) had the permission of the patients to be accompanied by an observer.

Five home nurses who met the inclusion criteria were selected, and all were willing to be accompanied by a researcher on their round and to record the performed activities on one round (morning or afternoon). They were all women, had a mean age of 45 years (range between 31 and 51 years), worked part time (mean of 24 h per week), and had an average working experience of 24 years as a nurse and 13 years in the home nursing organization. During this validation study, a total of 69 individual patients received care from these five home nurses, 71% of which were female, and having a mean age of 79 years (range between 43 and 97 years).

#### **Procedure**

The selected home nurses received an envelope with an invitation letter, explaining the purpose of the study and the instrument, an informed consent form, and copies of the 24-h recall for each patient in care. During the nursing round, the researcher recorded all activities immediately after each patient's visit and this recording was set out to be the standard measurement. The five participating nurses completed the 24-h recall form independently from the researcher. Three of the participating nurses completed each form immediately after a patient's visit. The two other nurses completed the forms for all patients at the end of their nursing round. This strategy allowed us to investigate whether there is an impact of the moment of recording the performed activities on the reliability of the data. The home nurses were assured that the results would remain confidential. At the end of each round, the researcher collected the recorded forms for analysis.

#### **Data analysis**

To assess the interrater reliability of the 24-h recall instrument for home nursing, we calculated the proportion of observed agreement ( $P_0$ ), the Kappa coefficients ( $K$ ), the prevalence index (PI), and the bias index (BI). The proportion of observed agreement is the ratio of exact agreement between the raters per total number of assessments. To correct for agreement by chance, we also calculated Cohen's Kappa coefficient (Fleiss and Cohen, 1973). The 24-h recall instrument consists of binary items (performed or not performed). Therefore, the unweighted Cohen's Kappa was used. Kappa values are scaled to range from  $-1$  to  $+1$ : a negative value indicates poorer than chance agreement, zero indicates exactly chance agreement, a positive value indicates better than chance agreement, and a value of unity indicates perfect agreement (Fleiss and Cohen, 1973). Furthermore, the strength of agreement for the Kappa coefficient is considered to be poor for Kappa values below 0.40, moderate from 0.41 to 0.60, substantial from 0.61 to 0.80, and almost perfect above 0.80 (Landis and Koch, 1977). For binary items, paradoxes in agreement parameters can be owing to bias and prevalence effects (Byrt *et al.*, 1993; Hoehler, 2000; Sim and Wright, 2005; Vach, 2005). Therefore, the PI and BI are reported

*Primary Health Care Research & Development* 2015; **16**: 79–86

in addition to the Kappa to address these paradoxes of the Kappa statistic (Sim and Wright, 2005; Cunningham, 2009). The PI for binary items was calculated as follows: the absolute value of the difference between the number of cases rated as positive by the researcher and the nurse, and the number of cases rated as negative by the researcher and the nurse, divided by the total number of assessments (Sim and Wright, 2005; Cunningham, 2009; Wellens *et al.*, 2012). The BI calculates the extent to which the researcher and the nurse disagree on the proportion of cases in a specific category: the absolute value of the difference between the number of cases rated as positive by the researcher and as negative by the nurse, and the number of cases rated as negative by the researcher and as positive by the nurse; divided by the total number of assessments (Sim and Wright, 2005; Cunningham, 2009; Wellens *et al.*, 2012). The Kappa coefficient was calculated using SAS 9.2, and the PI and BI were calculated using SAS 9.2 and Excell.

## Results

### Validity: evidence based on test content

The home nurses, who revised the preliminary instrument during a nursing round, provided the following suggestions to improve the instrument: the addition of the interventions ‘communication with colleagues’, ‘communication with other health care workers’, ‘giving education and information to patients’, ‘providing emotional support to patients and their informal caregivers’, ‘aid with daily care’, and ‘home dialysis’; a columnar listing of the nursing activities; a removal of the time recording per activity (they forgot to record the time or they did not know when to start or end the recording or they performed a lot of activities simultaneously, which made it difficult to rate each activity separately, and there was a huge discrepancy between the nurses’ and the researcher’s time rating). These adaptations resulted in an instrument containing 137 items (cf. annex): six items with regard to patient characteristics; two items with regard to the time recording per visit (time entering and time leaving the home of the patient); and 129 binary items with regard to the performed activities (‘psychosocial support: subject’, ‘aid with daily care: specify’, and ‘Other Care – Specify clearly’ are not part of these 129 items).

*Primary Health Care Research & Development* 2015; **16**: 79–86

### Interrater reliability

The interrater reliability of the 24-h recall could be evaluated for 55 of the 129 binary items (Table 1). The remaining 74 activities were not performed during the test period. The mean observed proportion of agreement for the 55 items was 0.94. For all, but two items (observation overall health status; giving information/advice) the observed agreement was  $\geq 0.80$ . This suggests that the agreement between raters is high.

The 55 items had a mean kappa value of 0.71, indicating that the level of agreement was substantial. Applying the conventional cut-offs on these 55 items for interpreting kappa values, reliability was perfect to almost perfect ( $K = 1$  or  $K \geq 0.81$ ) for 24 items (44%), substantial ( $0.61 \leq K < 0.81$ ) for 15 items (27%), moderate ( $0.41 \leq K < 0.61$ ) for eight items (15%), and poor ( $K < 0.41$ ) for eight items (15%).

The PI ranged between 0.16 and 0.97, showing great variety. In general, prevalence indices reflect the homogeneity of a sample: if the PI is known to be high (eg, 0.97), the magnitude of kappa might be reduced (Sim and Wright, 2005; Burn and Weir, 2011). The BI was very low, ranging between 0 and 0.14. This finding suggests that there is no tendency between the raters to score the performed activities to a different pattern.

To investigate the potential impact of the moment of completing the 24-h recall form on the reliability of the data, we calculated the mean kappa value for each nurse/researcher pair. The kappa values for the three nurses who completed the forms after each patient visit were  $K = 0.86$ ;  $K = 0.90$ ; and  $K = 0.83$ . For the two nurses who completed the 24-h recall forms for all patients at the end of their nursing round, the mean kappa values were  $K = 0.61$  and  $K = 0.84$ . This indicates that the moment of recording of the performed activities did not influence the interrater reliability.

## Discussion

In the present study, we offered evidence to support the face validity, content validity, and the interrater reliability of the 24-h recall instrument for home nursing in measuring the activity profile of home nurses.

The iterative process of the developmental and testing phase resulted in an instrument that is valid

**Table 1** Observed proportion of agreement, kappa (CI of kappa), PI and BI for each item of the 24-h recall ( $n = 69$  patients)

Activity	$P_0$	Kappa (95% CI)	Strength of agreement	PI	BI
Physiological: basic					
Self-care facilitation					
Taking/cleaning up material to perform personal hygiene	0.88	0.76 (0.61–0.92)	Substantial	0.16	0
Self-care assistance: dressing	0.96	0.91 (0.81–1.00)	Almost perfect	0.20	0.04
Bathing	0.97	0.93 (0.85–1.00)	Almost perfect	0.33	0.03
Partial (excl legs/feet)	0.87	0.66 (0.45–0.86)	Substantial	0.49	0.01
Complete personal hygiene	0.87	0.79 (0.65–0.93)	Substantial	0.17	0.1
Washing hair	0.98	0.88 (0.65–1.00)	Almost perfect	0.87	0.01
Introducing lotion/deodorant	0.88	0.76 (0.61–0.91)	Substantial	0.19	0.06
Hair care	0.97	0.92 (0.82–1.00)	Almost perfect	0.48	0.03
Shaving the patient	1	1 (1.00–1.00)	Perfect	0.80	0.00
Nail care	1	1 (1.00–1.00)	Perfect	0.94	0.00
Brushing teeth/prosthesis	1	1 (1.00–1.00)	Perfect	0.97	0.00
Introducing/replacing incontinence material	0.91	0.74 (0.55–0.94)	Substantial	0.57	0.00
Elimination management					
Other care elimination	0.97	0.48 (–0.13–1.00)	Moderate	0.94	0.00
Immobility management					
Transfer	0.96	0.82 (0.62–1.00)	Almost perfect	0.72	0.01
Positioning	0.94	0.47 (0.03–0.91)	Moderate	0.88	0.03
Physiological: complex					
Drug and perfusion management					
Oral	0.97	0.86 (0.67–1.00)	Almost perfect	0.77	0.00
Administration	0.98	0.79 (0.40–1.00)	Substantial	0.93	0.01
Preparation	1	1 (1.00–1.00)	Perfect	0.83	0.00
Supervision administration	0.96	0.39 (–0.15–0.92)	Poor	0.93	0.04
Injection	0.98	0.92 (0.78–1.00)	Almost perfect	0.78	0.01
Subcutaneous (excl insulin)	0.98	0.92 (0.78–1.00)	Almost perfect	0.78	0.01
Other care medication/perfusion	0.97	0.73 (0.38–1.00)	Substantial	0.88	0.00
Skin and wound care					
Administration medication: eye	0.98	0.92 (0.78–1.00)	Almost perfect	0.78	0.01
Administration therapeutic creme	0.87	0.63 (0.41–0.85)	Substantial	0.55	0.04
Compression therapy	0.94	0.72 (0.45–0.98)	Substantial	0.77	0.00
Wound care: simple	0.98	0.93 (0.80–1.00)	Almost perfect	0.75	0.01
Wound care: complex	0.98	0.79 (0.40–1.00)	Substantial	0.93	0.01
< 60 cm <sup>2</sup>	0.98	0.66 (0.04–1.00)	Substantial	0.96	0.01
Gauze and/or irrigation	1	1 (1.00–1.00)	Perfect	0.97	0.00
Stoma care	1	1 (1.00–1.00)	Perfect	0.97	0.00
Ureterostomy	1	1 (1.00–1.00)	Perfect	0.97	0.00
Diabetes care					
Measuring blood glucose	1	1 (1.00–1.00)	Perfect	0.74	0.00
Insulin preparation	1	1 (1.00–1.00)	Perfect	0.77	0.00
Administration insulin	1	1 (1.00–1.00)	Perfect	0.77	0.00
Other care					
Measuring parameters (blood pressure, respiration, ...)	1	1 (1.00–1.00)	Perfect	0.86	0.00
Observation overall health status	0.62	0.15 (–0.07–0.38)	Poor	0.36	0.14
Washing/disinfecting hands before/after care	0.84	0.67 (0.49–0.85)	Substantial	0.20	0.07
Psychosocial support					
Education	0.97	–0.02 (–0.03–0.01)	Very poor	0.97	0.00
Information/advice	0.78	0.31 (0.04–0.58)	Poor	0.61	0.04
Sensibilization	0.97	0.48 (–0.13–1.00)	Moderate	0.94	0.00
Emotional support	0.97	0.41 (0.11–0.71)	Moderate	0.75	0.10
Patient	0.87	0.41 (0.07–0.76)	Moderate	0.81	0.07
Informal caregiver/family patient	0.9	0.55 (0.10–1.00)	Moderate	0.90	0.01
Aid with daily care	0.98	0.79 (0.40–1.00)	Substantial	0.93	0.01

**Table 1:** (Continued)

Activity	$P_0$	Kappa (95% CI)	Strength of agreement	PI	BI
Administration					
Registration (global)	0.87	0.68 (0.49–0.87)	Substantial	0.43	0.07
Registration in patient record	0.93	0.74 (0.52–0.96)	Substantial	0.67	0.01
Registration in diabetes record	0.98	0.94 (0.82–1.00)	Almost perfect	0.72	0.01
Consultation patient record	0.88	0.18 (–0.13–0.48)	Poor	0.86	0.12
Consultation diabetes record	0.91	0.37 (–0.00–0.74)	Poor	0.86	0.09
Communication					
Consultation/information exchange with	0.8	0.46 (0.22–0.69)	Moderate	0.51	0.09
Nurse	1	1 (1.00–1.00)	Perfect	0.97	0.00
General practitioner	1	1 (1.00–1.00)	Perfect	0.91	0.00
Patient	0.83	0.39 (0.11–0.67)	Moderate	0.65	0.00
Family	0.9	–0.02 (–0.07–0.02)	Very poor	0.90	0.07
Other	0.96	0.39 (–0.15–0.92)	Poor	0.93	0.04

PI = prevalence index, BI = bias index.

on test content. Twelve categories of activities are withheld that are relevant to measure the activity profile of home nurses. In each category, respondents have the possibility to record additional activities that are not included in the list. This has both positive and negative consequences. It is positive that the full spectrum of activities can be reported, even if they are not listed in the instrument. A negative consequence is that it can result in an ‘easy-way recording’: participants write down activities that are in fact listed in the category or that are listed in another category. Therefore, it is important to thoroughly explain the purpose of the free-writing space. The time recording per activity was demonstrated to fail in terms of validity, and therefore was replaced by a time recording per patient visit. Although the instrument captures the whole spectrum of home nursing activities, it fails to reflect the process of care. Further research is needed to refine the ‘psychosocial support’ section in a way that the instrument shows the shift to more complex homecare in which technical activities are performed in combination with more intellectual activities. Despite the fact that psychosocial support items are less directly observable nursing care activities, 7 of the 55 items for which the interrater reliability could be measured were psychosocial support items. All but one item (information/advice) had an observed proportion of agreement higher than 0.90, indicating that the level of agreement was almost perfect and that both the home nurse and the researcher observed

and recorded these less directly observable activities. However, it is important to indicate that working with a predetermined list of activities tends to influence nurses’ description of the care they provide. This influences them to record activities that they did not actually provide and can also help them to focus on the less observable intellectual activities. The instrument was developed to record the actual work of home nurses, but it does not give an indication of the quality of the performed activities.

In general, the proportions observed agreement were very high, and the strength of kappa agreement was substantial to almost perfect. Only two items had low Kappa values in combination with observed agreement values below the threshold of 0.80 (observation of overall health status and giving information/advice). This suggests potential inequality between raters. A total of 14 items had kappa values lower than 0.60, despite high-observed agreement values ( $\geq 0.80$ ). This reflects homogeneity of the sample on the particular items, which reduces the kappa value (Wellens *et al.*, 2012). This finding is supported by the fact that the PI was  $>0.80$  for 11 of these 14 items. For binary items, a high percentage of agreement along with a poor kappa agreement for the same item can be because of bias and prevalence effects (Byrt *et al.*, 1993; Vach, 2005). The prevalence of item scores affects the stability of kappa. If the ratings of the sample of patients lack variability (ie, are homogeneous), it is unlikely that kappa will be

close to 1. This phenomenon is independent of the sample size (Wellens *et al.*, 2012). Therefore, for binary items, the PI should be calculated: the absolute value of the difference between the number of cases rated as positive by both raters, and the number of cases rated as negative by both raters, divided by the total number of assessments (Sim and Wright, 2005). Furthermore, the BI was very low, indicating that there is no tendency between the raters to record the performed activities according to a different pattern. Consequently, it can be concluded that bias did not play a role in the magnitude of kappa. Taken together, we can conclude that the interrater reliability of our 24-h recall instrument is good.

Hence, the 24-h recall instrument for home nursing can be considered to be a new and unique instrument in the landscape of home health care: it is an important first step in capturing the everyday activities of home nurses; it does not take too much of home nurses' time (no more than 5 min per patient); and it has shown evidence on face and content, and interrater reliability. Because of these characteristics, this instrument is used in a nationwide study on the activity profile of home nurses in Belgium, in which 81 493 patients were included. The results of that study will be reported in a related article.

The 24-h recall instrument for home nursing can be applied for different purposes. It can be used to investigate the current activity profile and job characteristics of home nurses. However, the instrument can also be used to study the activities performed by auxiliary nurses or health-care assistants in home nursing. This would provide important information for managers and policy-makers. Profiling the activities of home nurses and health-care assistants would allow managers to proactively anticipate to the fast evolving and challenging change in needs towards home nursing. Furthermore, the instrument is an important tool for policymakers in terms of adequate decision making and future policy planning, for example, determining the future need for and competences of home nurses. From a research perspective, the newly developed 24-h recall instrument for home nursing could be a step towards the creation of a gold standard to measure the activity of nurses and auxiliary services.

This study had some methodological limitations that need to be considered in the interpretation of the findings. First, with regard to the development of the instrument the non-participating observations

might have influenced the nurses in a way that they did not perform their activities as they usually do. Second, a small number of home nurses participated in this study. Hence, the distribution of activities performed by these nurses is unlikely to be representative, and therefore was not reported in the present article. The results of a nationwide study including more than 80 000 patients will provide more accurate data on home nursing activities. They will be reported in a related article. Third, this study focused only on the validity evidence based on test content and on the interrater reliability to evaluate the psychometric properties of the 24-h recall instrument for home nursing. Other aspects of validity and reliability of this instrument should be evaluated when using the instrument in large-scale projects. Fourth, this study was conducted with home nurses in Belgium. It might be possible that the replication of this study with other health-care professionals, in other settings and in another health-care system would result in different findings. Hence, we need further research on other settings and countries to see to what extent our findings are transferable.

## Conclusion

We developed a 24 h recall instrument to measure the activities of home nurses. The findings in this study support the validity evidence based on test content and the interrater reliability of this instrument. This instrument can help to shape practice and policy by making the home nursing profession more transparent: a clear insight into the kind of care that is provided by home nurses and is received by the patients in primary care contributes to the development of a clear definition of the role of home nurses in health care. However, further research is necessary to examine other aspects of validity and reliability of the instrument.

## References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1999: *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 9–36.
- Blegen, M.A., Goode, C.J. and Reed, L. 1998: Nurse staffing and patient outcomes. *Nursing Research* 47, 43–50.
- Primary Health Care Research & Development* 2015; 16: 79–86

- Burn, C.C. and Weir, A.A.S.** 2011: Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. *The Veterinary Journal* 188, 166–70.
- Byrt, T., Bishop, J. and Carlin, J.B.** 1993: Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 46, 423–29.
- Conway, J.M., Ingwersen, L.A., Vinyard, B.T. and Moshfegh, A.J.** 2003: Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *American Journal of Clinical Nutrition* 77, 1171–178.
- Cunningham, M.** 2009: More than just kappa coefficient: a program to fully characterize interrater reliability between two raters. *Statistics and Data Analysis (SAS Global Forum 2009)* 242, 1–7.
- De Vliegher, K., Paquay, L., Grypdonck, M., Wouters, R., Debaillie, R. and Geys, L.** 2005: A study of core interventions in home nursing. *International Journal of Nursing Studies* 42, 513–20.
- Fleiss, J.L. and Cohen, J.** 1973: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 33, 613–19.
- Frich, L.M.** 2003: Integrative literature reviews and meta-analysis: nursing interventions for patients with chronic conditions. *Journal of Advanced Nursing* 44, 137–53.
- Hoechler, F.K.** 2000: Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology* 53, 499–503.
- Hughes, L.C., Robinson, L.A., Cooley, M.E., Nuamah, I., Grobe, S.J. and McCorkle, R.** 2002: Describing an episode of home nursing care for elderly postsurgical cancer patients. *Nursing Research* 51, 110–18.
- Jansen, P.G.M., Kerstra, A., Abu-Saad, H.H. and Van der Zee, J.** 1996: The effects of job characteristics and individual characteristics on job satisfaction and burnout in community nursing. *International Journal of Nursing Studies* 33, 407–21.
- Keller, L.O., Strohschein, S., Lia-Hoagberg, B. and Schaffer, M.** 1998: Population-based public health nursing interventions: a model from practice. *Public Health Nursing* 15, 207–15.
- Landis, J.R. and Koch, G.G.** 1977: The measurement of observer agreement for categorical data. *Biometrics* 33, 159–74.
- Martin, K.S., Scheet, N.J. and Stegman, M.R.** 1993: Home health clients: characteristics, outcomes of care, and nursing interventions. *American Journal of Public Health* 83, 1730–734.
- McCorkle, R., Strumpf, N.E., Nuamah, I.F., Adler, D.C., Cooley, M.E., Jepson, C., Lusk, E.J. and Torosian, M.** 2000: A specialized home care intervention improves survival among older post-surgical cancer patients. *JAGS* 48, 1707–713.
- Mistiaen, P., Duijnhouwer, E., Wijkel, D., de Bont, M., Veeger, A.** 1997: The problems of elderly people at home one week after discharge from an acute care setting. *Journal of advanced nursing* 25, 1233–240.
- Park, S.Y., Paik, H.Y., Skinner, J.D., Spindler, A.A. and Park, H.R.** 2004: Nutrient intake of Korean-American, Korean, and American adolescents. *Journal of the American Dietetic Association* 104, 242–45.
- Robinson, L., Dugger, K., Fong, G., Heintzman, T., Hnizdo, S., Libby, J., McDaniels, J., Miles, J., Rempel, H., Taylor, A. and Wrashaw, M.** 2006: Palliative home nursing interventions for people with HIV/aids: a pilot study. *Journal of the Association of Nurses in Aids Care* 17, 37–46.
- Schumacher, K.L. and Marren, J.** 2004: Home care nursing for older adults: state of the science. *Nursing Clinics of North America* 39, 443–71.
- Sim, J. and Wright, C.C.** 2005: The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy* 85, 257–68.
- Subar, A.F., Thompson, F.E., Potischman, N., Forsyth, B.H., Buday, R., Richards, D., McNutt, S., Hull, S. G., Guenther, P.M., Schatzkin, A. and Baranowski, T.** 2007: Formative research of a quick list for an automated self-administered 24-hour dietary recall. *Journal of the American Dietetic Association* 107, 1002–1007.
- Vach, W.** 2005: The dependence of Cohen's kappa on the prevalence does not matter. *Journal of Clinical Epidemiology* 58, 655–61.
- Wellens, N.I.H., Van Lancker, A., Flamaing, J., Gray, L., Moons, P., Verbeke, G., Boonen, S. and Milisen, K.** 2012: Interrater reliability of the interRAI Acute Care (interRAI AC). *Archives of Gerontology and Geriatrics* 55, 165–72.