Official title of the study:
Implementation of Knowledge-Based Palliative Care for Frail Older Persons in Nursing Homes

NCT number: NCT02708498

Date of the document: April 8, 2015, updated October 6, 2020

Statistical Analysis Plan

Simple descriptive statistics:
Median-Mean, Quartiles, Percentage
These are applied to all quantitative studies.

Baseline study:
Target population: Older people
Instruments: Characteristics of older participants, WHOQOL-BREF & WHOQOL-OLD

The data were analysed in one descriptive and one analytical section. In the descriptive section, numbers and percentages were calculated on a group level. A mean score was calculated for each domain using both the raw domain score, a standardized domain score (SDS) with the raw domain score divided by the number of items in the domain (1-5) and a transformed domain mean score (TDS 0-100) according to the instrument manuals. The different calculations were performed to enable comparison of our results with those from other studies.

In the analytical section, the participants were divided into two age groups (66-88 years and 89-102 years) by the population median age and gender (female and male). To examine differences in quality of life, a comparison of medians was conducted using the Mann-Whitney U test.

Comparative studies (Baseline and Follow-up):
Target population: Older people, next-of-kin, staff
Instruments: PCAT, PCQ-S, PCQ-P, WHOQOL-BREF, WHOQOL-OLD, Your experience of palliative care, Next-of-Kin Participation in Care (NoK-PiC)

The selection of the statistical tests to analyze data other than the descriptive statistics was based on whether the data were distributed normally and the scale level of the instruments. The Mann-Whitney U-test and the Kruskal-Wallis test was applied to compare the baseline characteristics of the two groups. The quantitative data collected from questionnaires were analyzed using within-group comparisons (the data from after the intervention period were compared with the data from before the intervention period in the intervention group and the control group) by Wilcoxon signed rank test.

Either Pearson's chi-square test or Fisher's exact test was used when the expected value was less than 5.

A two-tailed p-value of <0.05 was regarded as statistically significant.

In order to investigate if the drop-outs consisted a threaten of external validity we performed sub-groups analysis through comparing the drop-outs in the intervention group with the participants in the intervention group, and comparing the drop-outs within the control group with the participants in the control group.

Analyses were performed using IBM SPSS Statistics version 24.

Logistic regression studies (Baseline and Follow-up):

Target population: Staff
Instruments: P-CAT, PCQ-S

A univariate logistic regression analysis was performed with the independent variable improvement/non-improvement in P-CAT and PCQ-S and the dependent variable group (control/implementation).

Psychometric studies

Target population: Next-of-kin
Instrument: Next-of-Kin Participation in Care (NoK-PiC)

Step 1. Dimensionality
The dimensionality of the 37 items was explored by means of Rasch Model Analysis (RMA) and Explorative Factor Analysis (EFA). EFA needs complete data, i.e. no missing responses in any items. RMA can process missing data. Estimation methods merely summarize the non-missing observations that are relevant, and compare them with their expectations.

NoK-PiC data was analysed according to Rasch Measurement Theory (RMT), using the polythomous Rasch model in RUMM2030 (Professional Edition 5.4). A principal component analysis of the residuals was conducted in RUMM2030 to iexplore dimensionality of the NoK-PiC items. Person location estimates were derived from subsets of items that loaded positively and negatively on the first principal component of residuals. The overall proportion of persons with significantly different measures from the two item subsets should be <5% to support unidimensionality.

EFA was conducted based on recommendations including parallel analysis to decide for decision about how many factors to retain for rotation, Minimum Rank Factor Analysis (MRFA, that is able to estimate the percentage of common variance explained by the EFA model), and on a promin rotation that allows factors to correlate (oblique). In addition, since the items are polythomous, and since MRFA assume only minor departures from normality, we based the analysis on a polychoric correlation matrix. The software FACTOR (Lorenzo-Seva, Ferrando, release 10.5.03, June 2017) was used to conduct EFA.

Step 2. Item reduction
Based on conceptual reasoning among the researchers and based on RMA item reduction was conducted. The main statistical reasons for item reduction was based on item fit. In general, individual item fit residuals should range between -2.5 and 2.5, with the ideal being 0. Signs of multidimensionality/under discriminating items (fit residuals >2.5) and local dependency/item redundancy/over discriminating items (fit residuals > -2.5) led to further exploration of the possible reasons for misfit. This exploration included inspection of item residual correlations, and the item characteristic curves to get a graphical understanding of the deviations from Rasch model expectations. The nature of observed misfit (local response dependence or multidimensionality) often needs to be determined based on conceptual reasoning rather than statistics.

Step 3. Dimensionality after item reduction
Dimensionality of the remaining 30 items (after step 2) was explored using EFA, following the same procedures as described above (Step 1). The appropriateness of performing EFA was, in this step, checked according to quality criteria, i.e. by means of: the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (should be KMO 0.50 or above); and Bartlett's test (should be p-value <0.05).

Step 4. RMA of final and separate factors
According to RMA, a set of items, supposed to be unidimensional, rated (dichotomous or polytomous scoring) in a sample, are compared to the "Perfect Rasch Model" that is based on fundamental measurement principles from the physical sciences. The model separately locates persons and items on a common interval level logit metric, ranging from minus to plus infinity, with mean location set at zero. The extent to which a successful measurement has been achieved is determined by examining the fit between observed data and model expectations. If data accord sufficiently with the model, linear measurement and invariant comparisons are possible. However, almost no scale is perfect, thus "almost perfect" can be a goal to achieve. Here the RMA addresses targeting, reliability, model fit, Differential Item Functioning (DIF, by gender and age), and hierarchical item ordering.

Step 5. Group comparisons
The IBM SPSS Statistics 23 software was used for comparisons between independent groups. Data were assessed regarding underpinning assumptions, described, and analysed using Mann Whitney U-tests. Significance was set to $P<0.05$.

Target population: Next-of-kin
Instrument: WHOQOL-BREF

Data were analyzed using IBM SPSS Statistics 20 and IBM SPSS Amos, version 25. In a first step, WHOQOL-BREF's negatively phrased items (Q3, Q4, and Q25) were reversed (1=5, 2=4, 3=3, 4=2, 5=1). Two items – question 1, about overall QOL, and question 2, about the overall perception of health – were analyzed separately.

Validity
Internal validity in terms of data quality and targeting was assessed by missing values, item response, and floor and ceiling effects. Floor and ceiling effects were calculated and should be less than 20% in order to ensure that the scale is capturing the full range of potential responses in the population and that changes over time can be detected. The Kolmogorov-Smirnov test was used to assess normality at the item level. A statistically non-significant result ($p\geq0.05$) indicates normality. Construct validity was estimated by correlating item/question 1: "How would you rate your QOL?" and item/question 2: "How satisfied are you with your health?" with the four domains in the WHOQOL-BREF using Spearman's rank correlation.

Reliability
Internal consistency was assessed using Cronbach's alpha. To determine to which degree each item correlates with the total score, Corrected Item-Total Correlations (CIT) were calculated. The limit for satisfactory item correlation was set to >0.30. Values less than 0.30 indicate that the item is measuring something different from the scale as a whole.

Factor structure
The appropriateness of performing CFA was checked according to quality criteria. These criteria were fulfilled by means of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (0.90, should be 0.50 or above); Bartlett's test (0.01, should be <0.05); and determinant (0.008, should be >0.00001). The number of cases per item was also calculated. Recommendations range from 2 to 20 subjects per item, with an absolute minimum of 100 to 250 subjects. Confirmatory Factor Analysis (CFA) was applied to assess goodness of fit by means of various descriptive fit indices. Specifically, the normed fit index (NFI), the comparative fit index (CFI) and the root mean squared error of approximation (RMSEA) were used.

The NFI equals the difference between the chi-square of the two models divided by the chi-square of the null model. An NFI over 0.90 is preferable. An NFI of .90 shows that the model of interest improves the fit by 90% in relation to the null model. CFI assesses fit relative to a null model and ranges from 0 to 1, where values exceeding 0.95 are regarded as acceptable. The RMSEA test assesses the lack of fit per degree of freedom of the model. Zero represents a perfect fit, ≤0.05 indicates very good fit, >0.05–0.08 good fit, and ≥0.10 poor fit. IBM SPSS Amos only accepts data files with no missing values, and therefore, 14 participants were excluded from the CFA analysis (n=240).