Statistical Analysis Plan

BEVOR

Effects of an advance care planning program to improve patient-centred care: a cluster-randomised intervention trial focusing on nursing homes (BEVOR)

NCT04333303

Version: V01-00

Date: 04.03.2023

Authors: Susanne Lezius and Amra Hot

Reviewers: Antonia Zapf, Jürgen in der Schmitten, Kornelia Götze, Michaela Schunk, Gabriele Meyer

Institution: UKE

Document:	Statistical Analysis Plan for Final Analysis
Sponsor Project Code:	BEVOR
Study No.	NCT04333303
Study short title:	Effects of an advance care planning program to improve patient-centred care: a cluster-randomised intervention trial focusing on nursing homes (BEVOR)
Version / date:	**V01-00** / **March 4, 2023** based on study protocol version 1.0, dated May 22, 2022
Author:	Susanne Lezius and Amra Hot
Sponsor:	Heinrich-Heine University, Duesseldorf
Statistician:	Prof. Dr. Antonia Zapf Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg

Table of Contents

Abbreviations

ACP	Advance Care Planning
AD	Advance directives
ADBI	After Death Bereavement Interview
AE	Adverse Event
CCCP	Care consistency with care preferences
CI	Confidence interval
CPR	Cardiopulmonary resuscitation
cRCT	Cluster randomized trial
CRF	Case Record Form
DC1	Data collection 1
DC2	Data collection 2
EFS	Evalued for Safety Set
HADS	Hospital Anxiety and Depression Score
IEC	Independent Ethics Committee
IQR	Interquartile range
IRB	Institutional Review Board
IRR	Incidence rate ratio
ITT	Intention-to-treat
NH	Nursing home
OR	Odds ratio
POLST-E	Physician or patient/proxy order for life sustaining treatment in emergencies
PP	Per Protocol population
SD	Standard deviation
SAP	Statistical Analysis Plan
SDM	Shared decision making
3CP	Care consistency with care preferences

1. Introduction

This Statistical Analysis Plan (SAP) is based on the study protocol version 1.0 of May 22, 2022 and follows the guideline for statistical analysis plans [1].
Some elements of the statistical methods and of the study design have already been described in the study protocol [2]. This Statistical Analysis Plan (SAP) aims to further specify the procedures and statistical methods to be applied during the final analysis of the study data.
This SAP is finalized and published before unblinding of the outcome assessors (Biometry, Hamburg) took place.

1.1 Background and rationale

Advance Care Planning (ACP) for residents of nursing homes (NHs) has recently been covered by German sickness funds. However, clinical effects of ACP have not been studied in Germany yet. The purpose of this study is to investigate whether implementing a comprehensive ACP program improves care consistency with care preferences of NH residents. The aim is to increase consistency between medical treatments provided and the respective patient preferences, and to reduce the corresponding emotional burden for health care proxies and staff. In addition, place of death should more often be the NH, i.e. at home, rather than the hospital, consistent with the residents' wishes. The complex ACP intervention comprises training of professional ACP facilitators, supporting organizational development in NHs and other relevant institutions, and moderating regional change management among a network of all related health care institutions and players.

1.2 Objectives

The goal of the multicenter BEVOR study is to evaluate the effectiveness of a structured advance care planning (ACP) program in NH residents regarding the reduction of unwanted medical interventions, notably hospital admissions, as compared to a control group. Conversely, the rate of residents whose treatment preferences are known and honored should be increased.
In the following, the time point of baseline measurement is defined as t0, of the end of the run-in phase of the intervention as t1, and of succeeding 12-months follow-up as t2. For more details, see section 2.7.
Researchers will take two independent approaches for data collection:
Firstly, participating NHs will provide anonymous pooled data on all their residents which include long-term as well as respite care residents of a defined time period (data collection 1 = DC1), including the hospitalization rate (primary outcome).
If not stated otherwise, data of both groups are used, long-term and respite care residents.
Secondly, a cohort of all long-term residents who have given informed consent (or for whom their health care proxy has given informed consent), their health care proxies and/or relatives, and the

respective nursing staff will be repeatedly surveyed until t2 (or until residents die earlier) (data collection 2 = DC2).
The following primary hypothesis will be tested:

Primary hypothesis: Compared to the control group, the structured ACP program reduces the 12-month hospitalization rate of residents of NHs at t2.

1.2.1 Primary Outcome

The primary outcome is part of the data collection DC1 and is defined as rate of hospitalizations (long-term care only) per 100 residents per year during the 12-months observation period, compared between intervention group and control group at t2. The rate of hospitalizations per 100 residents per year is defined as the quotient of the number of all hospitalizations in the last 12 months observation period divided by the number of occupied bed days multiplied by 365 (days) and 100 (residents).

1.2.2 Secondary Outcomes

The secondary outcomes are as follows, analyzed at t2:
At NH level (DC1):

1. number of hospital days (long-term) per 100 residents per past 12 months
2. number of hospital stays (long-term) per 100 residents during observation period (without consideration of occupied bed days)
3. number of patients with at least one hospitalization per 100 residents per past 12 months
4. number of residents who died in the hospital per 100 residents per past 12 months
5. number of pre-specified unwanted invasive treatments per 100 residents measured:
    a. number of residents with feeding tube (PEG) in place (cross-sectional)
    b. number of residents with PEG in use (cross-sectional)
    c. number of attempts of cardiopulmonary resuscitation (CPR) per past 12 months
    d. number of attempts of successful CPR per past 12 months
6. number of legal guardians per 100 residents (cross-sectional)
7. number of durable powers of attorney (DPOA) per 100 residents (cross-sectional)
8. advance directives (ADs)
    a. number of all advance directives per 100 residents (cross-sectional)
    b. number of advance directives signed by resident per 100 ADs (cross-sectional)
    c. number of advance directives last signed in the last 6m per 100 ADs (cross-sectional)
    d. number of advance directives in long-term care only per 100 ADs (cross-sectional)

9. physician or patient/proxy orders for life sustaining treatment in emergencies (POLST-E)
    a. number of POLST-E per 100 residents (cross-sectional)
    b. number of POLST-E not signed by resident per 100 POLST-E (cross-sectional)
    c. number of POLST-E in the last 6m per 100 POLST-E (cross-sectional)
    d. number of POLST-E with a do-not-hospitalize order per 100 POLST-E (cross-sectional)
    e. number of POLST-E per 100 residents (long-term care only; cross-sectional)

At resident level (DC2):

10. proportion of treatment decisions in the face of potentially life-threatening events in which the resident's preferences were
    a. known…
        i. … ostensibly, i.e. mere consent or refusal to treatment could be elicited regardless of any level of shared decision making (SDM)) (yes vs. no)
        ii. …. in the sense of "informed", i.e. consent or refusal to treatment could be elicited and was based on at least the first two of four predefined steps of an SDM process (step 2: "...was informed about available treatment options with benefits and risks") (yes vs. no)
        iii. … in the sense of "informed" and "enabled", i.e. additionally at least the 3rd step of the predefined 4-step SDM process was reached ("...was given opportunity to deliberate on the available treatment option") (yes vs. no)
        iv. … in the sense of "informed", "enabled" and "empowered", i.e. additionally the 4th step of the predefined 4-step SDM process was reached ("...was encouraged to decide between the available options based on the resident's preferences") (yes vs. no)
    b. AND honored:
        i. "ostensible" preferences (independent of any underlying SDM process) were honored (yes vs. no)
        ii. "informed" preferences (at least the 2nd step of the SDM process was reached) were honored (yes vs. no)
        iii. "enabled" preferences (at least the 3rd step of the predefined 4-step SDM process was reached) were honored (yes vs. no)
        iv. "empowered" preferences (at least the 4th step of the predefined 4-step SDM process was reached) were honored (yes vs. no),

    continuously over the 12 months before t2.

11. Perception of care consistency in residents deceased during the observation period measured as judgement (Likert Scale 0-10) of
    a. proxies and

b. nurses, respectively,

whether medical care (F2a) and nursing care (F2b) delivered in the past 3 months before death was consistent with the residents' then care preferences, measured by key items F2a/F2b selected from the After Death Bereaved Family Member Interview (ADBI) and documented continuously over the 12 months before t2.

12. Shared decision making and Advance Care Planning in residents deceased during the observation period, observed in the 3 months prior to death, and measured by means of the corresponding sections of an adapted German version of the After Death Bereaved Family Member Interview (ADBI):
    a. Inform and promote shared decision making (problem score #2)
    b. Encourage advance care planning (problem score #3)
    c. Overall Rating Scale for patient focused, family centered care
13. number of hospitalizations per 100 residents per last 12 months at t2
14. number of total hospital days per 100 residents per last 12 months at t2
15. number of visits at hospital outpatient services per 100 residents per last 12 months at t2
16. number of attempted CPRs in hospital or EMS per 100 residents per last 12 months at t2
17. number of intensive care unit days per 100 residents per last 12 months at t2
18. proportion of residents with artificial ventilation in hospital or EMS at t2
19. number of invasive ventilation hours in hospital per 100 residents per last 12 months at t2
20. proportion of residents with PEG in place at t2
21. proportion of residents with newly inserted PEG at t2
22. proportion of residents with PEG in use at t2
23. mortality rate during the observation period before t2
24. place of death (nursing home vs. others) during the observation period before t2
25. number of EMS transports per 100 residents per last 12 months at t2
26. number of general practitioner visits per 100 residents per last 12 months at t2
27. number of referrals to specialists per 100 residents per last 12 months at t2
28. proportion of residents who received at least 1 antibiotic at t2
29. proportion of residents with legal guardian at t2
30. quality of life: residents capable of self-rating use WHO-QoL German Version; residents incapable of self-rating use QUALIDEM at t2
31. Process parameters measuring the effect of ACP at t2:
    a. proportion of residents with advance care plans signed by facilitators, and/or physicians and/or representatives
    b. proportion of residents with durable power of attorney
    c. proportion of residents with POLST-E
    d. proportion of residents with any ADs

e. proportion of ADs signed by resident
f. proportion of residents with AD-by proxy
g. proportion of residents with POLST-E

32. Anxiety, depression and trauma among bereaved relatives at t2:
    a. Scoring on Revised Impact of Event Scale (IES-R): min. 0 pts, max. 88 pts., higher score indicates higher impact (trauma)
    b. Scoring in German version of Hospital Anxiety and Depression Score (HADS-D): min. 0, max. 14, higher score indicates higher degrees of anxiety and depression

Key secondary endpoints are 9e, 10a ii and 10b ii. This choice is informed by blinded baseline analyses.

1.2.3 Specific outcome definitions

In the following, the calculation rules for specific outcomes are described on the basis of the corresponding questionnaires:

- ADBI

  Referring to outcome 12, for the calculation of problem score #2 (12a), the items C1a, C1b, C1c, C1c(a), D19, D26a, D27a, D28a and E1 are used. The calculation of problem score #3 (12b) is done from a medical (items D2a, D3a, D4a, D4b) and nursing (D2b, D3b, D4c, D4d) perspective. The Overall Rating Scale (12c) is calculated using items F1-F6.

  In each case, the score is calculated as sum of the items. The category "I don't know" is considered a missing value. For problem score #2, up to 2 missing values are allowed and replaced by the mean of the resulting items. For problem score #3, up to 1 missing value per item in each perspective is allowed and replaced by the mean of the resulting items. For the overall rating scale, up to 1 missing value is allowed and replaced by the mean value of the resulting items. In case of more missing items the respective score is missing.

- IES-R

  Referring to the questionnaire in endpoint 32a, we calculate two different types of endpoints: continuous and binary.

  Three different subscales are calculated from the questionnaire:

  - Intrusion (item no. 1,3,6,9,14,16,20)
  - Avoidance (item no. 5,7,8,11,12,13,17,22)
  - Hyperarousal (2, 4,10,15,18,19,21)

  The subscale values result from the addition of corresponding item values. The total score is then the sum of the subscales. This is a continuous endpoint.

  For the diagnostic interpretation, X is calculated with the following formula:

  X = (-0.02 x intrusion) + (0.07 x avoidance) + (0.15 x hyperarousal) - 4.36

A score X > 0 is then interpreted as suspected post-traumatic stress disorder, resulting in a binary variable.

For the IES-R questionnaire no missing items are allowed.

- WHO-QoL German Version

  The WHO-QoL questionnaire (30) is divided into six different subscales (sensibility, autonomy, past, social part, death and intimacy). The raw scores of the six subscales are formed by calculating the mean value of the items and multiplying by the factor 4. Each subscale value is then transformed to the range 0-100. The total score value is formed by calculating the mean of all subscale values. Only a maximum of one missing value per subscale is allowed (replaced by the mean of the resulting items), otherwise a missing is given.

- QUALIDEM (Dichter et al. 2016)

  The QUALIDEM questionnaire (endpoint 30) is divided into different subscales. In our study, we evaluate the total score. This results from the addition of the item values. It must be taken into account that the positively directed items are scored inversely to the negatively directed items. This means that the answer option "Never" corresponds to a score of zero for a positively directed item and a score of six for a negatively directed item. No missing items are allowed.

- HADS-D

  The HADS-D questionnaire (32b) is divided into two different subscales (anxiety and depression). The raw scores of the two subscales are formed by summing up the items. The total score is formed by the sum of the two subscale scores. A maximum of one missing item per subscale can be replaced by the rounded-up mean of the 6 available answers. In case of more missing items the score is missing.

2 Study Methods

2.1 Trial design

BEVOR is a two-arm, parallel group, prospective, multicenter, cluster randomized controlled trial (cRCT). Nursing homes and related health care providers form the clusters and they themselves (DC1) or, respectively, their residents (DC2) the observation units. The allocation ratio of the clusters is 1:1 for the two groups:

- ACP program (experimental treatment group):
  Each nursing home and related care providers will receive an institutional and regional implementation of the complex German ACP program including a ACP facilitator .
- Care as usual (control group): For each nursing home and health care provider all common care options within the German health care system are available. The control group will receive a delayed ACP intervention after final data collection.

The trial is observer blinded (outcome assessors (until finalization of the SAP) and data collectors are blinded).
44 nursing homes with an average of 96 residents each, from the 4 study centres in Düsseldorf (2 regions), Munich (2 regions), Halle (2 regions) and Göttingen (2 regions) are participating.

2.2 Randomization

A 1:1 cluster-randomization at the level of NHs stratified by study center and region for experimental and control group was carried out using a computer-based code with variable block sizes with length 2 to 4 (to ensure intervention NHs in every region since it has been promised to the NHs) within each stratum. The randomization list was created by a research associate of the Institute for Medical Biometry and Epidemiology of the University Medical Center Hamburg-Eppendorf, who is not directly involved in the implementation of the research project using R version 3.6.2.

2.3 Sample size

To be allocated to trial: 44 NHs (22 per group) including on average n = 100 residents each, resulting in a target sample size of n = 4.400 (2.200 per group) enrolled patients in total.
To be analysed: n = 1760 residents in 22 NHs per group and n = 3520 residents in 44 NHs in total.
Full details of sample size calculation can be found in the study protocol [2].

2.4 Framework

BEVOR is a cluster-randomized, prospective, parallel-group, superiority trial testing whether NHs involved in the structured ACP program will have a lower 12-month hospitalization rate than NHs in the control group.

2.5 Statistical interim analyses and stopping guidance

No interim analyses are planned.

2.6 Timing of final analysis

All endpoints (primary and secondary) are evaluated after collection of all data and the data cleaning process (i.e. database has been reviewed for completeness and accuracy), presumably December 2022 to March 2023.
The results will be reported according to the CONSORT 2010 Guideline [3].
Pooled descriptive analyses of data of both trial arms can be performed earlier, as soon as all data for this data collection is available, e.g. pooled baseline data of nursing homes (DC1) will be evaluated regardless of the intervention group after completion of the baseline DC1 data collection and the

corresponding data cleaning.

2.7 Timing of outcome assessments

The timeline of the BEVOR study is as follows (see Fig. 2 of the study protocol): at the start of the study, data collection is initiated, which corresponds with the baseline measurement (t0). At the end of the run-in phase of the intervention, the 12-month observation phase begins (t1). The end of the observation phase corresponds with the measurement at 12-months follow-up (t2). The run-in phase of the intervention describes the time in which the intervention is established and reaches the residents, the NHs and the regions. It lasts from the time of randomization until t1. The intervention itself continues during the observation period.

The 12 months observation period's data will be collected in one follow up collection (DC1) or, respectively, for DC2 continuously over the entire period, including

- interviews with relatives and nursing home staff of deceased residents,
- four three-monthly sessions to identify all potentially life-threatening events and related treatment decisions for the respective preceding 3 months (3CP data collection including interviews), as appropriate.

The start of the observation phase (t1) is fixed for all study centres at 01.09.2021.

NH level (DC1): All outcomes on NH level will be measured at t0, t1 and t2 for the preceding12 months.

Resident's level (DC2): Data collection 2 is performed with residents, their surrogates and/or relatives, and the respective nursing staff and contains three different sections with different time points:

1. Framework survey including quality of care survey is conducted at t0 and t2 for the 12 months observation period.
2. Care consistency with care preferences (3CP) is assessed at t0 for the preceding 3 months and at t2. Data collection at t2 will be splitted into 4 phases of the 12-months observation period (each covering the preceding 3 months), beginning 3 months after t1, to collect data over the twelve months of observation as accurately as possible. It is ensured that there is no overlap in the survey periods.
3. The survey of relatives and staff regarding deceased residents is conducted at t0 and continuously over the 12 months before t2.

3 Statistical Principles

3.1 Confidence intervals and *P* values

All applicable statistical tests will be two-sided and at a 5% significance level. Analyses of secondary outcomes will be performed without adjustment for multiplicity. All confidence intervals presented will be 95% and two-sided.

3.2 Adherence and protocol deviations

Major protocol violation includes ineligible NHs and residents who were included in the trial by mistake, or for whom the intervention or other procedure differed from that outlined in the protocol, or failure of consent process.

Protocol violations are defined as deviations from the planned procedure according to the study protocol for which the study team is responsible. Major protocol deviations occur if

- the ACP facilitator does not attend the NH as scheduled
- data collection could not be performed (e.g. due to covid-19 pandemic)
- in- or exclusion criteria are violated (for both, residents and NHs)
- residents move out of the participating NH

Absence of sufficient adherence from the intervention according to the study protocol is defined at the NH and residents level. A major adherence deviation occurs if

- NHs of the intervention group do not utilize a minimum of the intervention elements (e.g. meetings of institutional steering group, staff qualification, ACP-facilitation, information of residents, relatives and surrogates), the minimum being defined as an intervention load amount below the respective median of all NHs in the intervention group

3.3 Analysis populations

3.3.1 Intention to treat Population (ITT)

The primary analysis population is the ITT (intention to treat) population. The ITT population on NH level consists of all NHs randomized to the study.
The ITT population on resident level consists of all enrolled residents who gave informed consent to participate in the study.

3.3.2 Per Protocol population (PP)

The Per Protocol population includes all NHs randomized (on NH) or all residents enrolled (on resident level) who have no major protocol violation (see 3.2).

More precisely, the per protocol population of NHs in the intervention group includes all NHs that meet the following criteria:

- Qualified interview facilitators were assigned to the NHs for the time of the intervention period
- Certification offers were made to the primary care physicians who cared for residents in the NHs belonging to the intervention group
- NHs were offered the intervention elements specified in the intervention concept

The per protocol population of NHs in the control group includes all NHs that were included and randomized according to protocol and did not violate any in- or exclusion criteria.

On resident level, the per-protocol population includes all residents of the PP NHs who did not move out of the participating NHs to another place of living before t2.

3.3.3 Adherence population

The adherence population includes all NHs randomized and all residents enrolled who are part of the respective PP populations and have no major adherence violation (see 3.2).
On NH level the intervention group of the adherence population includes all NHs with an intervention load as defined in 3.2 above the respective median of all NHs in the intervention group. A list of such NHs is provided to the Biometry by the interventionists after unblinding.
The adherence population of residents consists of all residents being part of the PP population who lived in adherent NHs.

4 Trial Population

4.1 Screening data

The following summaries will be presented overall and by study centre:
recruiting period, number of NHs screened, number of NHs and residents recruited, number of screened NHs, and residents not recruited.

4.2 Eligibility

The number of ineligible NHs will be reported, with reasons for ineligibility.

4.3 Recruitment

All recruitment information will be included and presented in a flow diagram according to the CONSORT Statement (extended version for cRCTs) (see Appendix).
The flow diagram summarizes the number of residents and/or NHs who

- meet the inclusion / exclusion criteria

- do not meet the inclusion/exclusion criteria
- post-hoc violate the inclusion / exclusion criteria
- are lost to-follow up
- have major protocol deviations
- died between t0 and t1
- are part of the ITT/PP population.

4.4 Withdrawal/follow-up

The number of residents/NHs dropping out of the study in the follow-up is documented as well as tabulated and will be given in the flow-chart. Dropout reasons will be given if known.

4.5 Characteristics of participating NHs and residents

The following descriptive characteristics are measured at t0 and, at NH level, additionally at t1:

NH level:

- Type of owner
- Number of beds
- Occupied bed days (in the past 12 months)
- Number of residents (in past 12 months and on day of data collection) (respite & long term care)
- Sex ratio of residents
- Mean age of residents
- Rate of all deceased residents (in the past 12 months)
- Rate of registered/assistant/hired nurses per 100 residents
- Rate of full nurse equivalents per 100 residents
- Rate of palliative care qualified nurses per 100 residents
- Total sick leave days per nursing staff (i.e. number of registered/assistant/hired nurses) (in the past 12 months)
- Palliative care concept in place (yes vs. no)
- Active collaboration with palliative care services (yes vs. no)
- Collaboration with hospice care services (yes vs. no)
- Collaboration with ethic consultations (yes vs. no)
- Rate ethics consultations (in the past 12 months) per 100 residents
- Rate of general practitioners per 100 residents
- Defined private co-payment for residents (relative deviation from the regional mean)

Residents level:

- Age
- Sex
- Country of birth
- Migration (yes vs. no)
- Professional degree
- Days since moving in at start of survey
- Insurance
- Marital status
- Religion
- Care level

5 Analysis

5.1 Analysis methods

5.1.1 Descriptive analysis of baseline characteristics

For baseline characteristics (as defined in 4.5) descriptive statistics will be presented separately for the two treatment groups, i.e. intervention and control, on NH level as well as on resident level. Categorical data are summarized by absolute and relative frequencies. Continuous data are summarized by mean and standard deviation (SD) as well as median, quartiles and/or interquartile range (IQR), as appropriate. The number of available observations or the number of missing observations are reported separately for treatment groups. On resident level, statistical comparisons using Mann-Whitney U-test, Chi^2-test or t-test (according to the data type of the characteristics) will be used to identify any imbalances between the two study groups. On NH level, statistical comparisons are not performed due to randomization.

On resident level, responders and non-responders will be compared descriptively regarding care level, age and sex.

5.1.2 Primary Outcome

The primary analysis is performed in the ITT population at NH level. For rate comparison a Poisson regression model is calculated taking into account group (intervention vs control) as independent variable, realized follow-up time as offset, region of NHs and the respective baseline hospitalization rate as covariates. The treatment effect is presented as incidence rate ratio (IRR) with corresponding 95%-Confidence interval (CI) and p-value. Marginal estimators (numbers per 100 residents per past 12 -months observation period (from t1 to t2)) with corresponding 95%-CI are given per treatment group.

5.1.3 Secondary Outcomes

The evaluation of secondary outcomes is performed in an explorative manner without adjustment for multiplicity and evaluated in the ITT population.

At NH level:

All secondary endpoints are count data.

For the endpoints 1,2, 4 and 5 c. & d., a Poisson Regression model is calculated with the respective number per 100 residents per past 12 months' observation period (from t1 to t2) as dependent variable, group (intervention vs control) as independent variable, number of occupied bed days as offset and the respective baseline rate and region of NHs as covariates.

The treatment effect is presented as incidence rate ratio (IRR) with corresponding 95%-CI and p-value. Marginal estimators (numbers per 100 residents per past 12 -months observation period (from t1 to t2)) with corresponding 95%-CI are given per treatment group.

For the secondary endpoint 3, the number of residents during the observation period before t2 is used as offset variable. The resulting marginal estimators are numbers per 100 residents.

The secondary endpoints 5 a. & b. and 6 – 7, 8 a. and 9 a. are measured cross-sectional. For these endpoints, a Poisson Regression model is calculated with the respective number as dependent variable, group (intervention vs. control) as independent variable, number of residents as offset and the respective baseline rate and region of NHs as covariates.

The treatment effect is presented as IRR with corresponding 95%-CI and p-value. Marginal estimators (numbers per 100 residents per past 12 -months observation period (from t1 to t2)) with corresponding 95%-CI are given per treatment group.

The secondary endpoints 8 b. – d. and 9 b. – e. are measured cross-sectional, too, but refer to a predefined subgroup.

For these endpoints, a Poisson Regression model is calculated with the respective number as dependent variable, group (intervention vs control) as independent variable and the respective baseline rate and region of NHs as covariates.

For the endpoints 8 b. – d. the number of ADs and for the endpoints 9 b. – e. the number of POLST-Es are set as offset, respectively.

The treatment effect is presented as IRR with corresponding 95%-CI and p-value. Marginal estimators (number per 100 ADs or 100 POLST-Es per past 12 -months observation period (from t1 to t2)) with corresponding 95%-CI are given per treatment group.

At level of treatment decisions:

The endpoints 10 a i. & ii. and b i. & ii. (3CP) are outcome variables with levels 1 = “yes” and 0 = “no”, i.e. binary outcomes. For these endpoints, a mixed logistic regression model will be calculated with the specific binary endpoint as dependent variable, group (intervention vs. control) as fixed effect, region of NHs, respective NH, resident and potential life threatening event (PLE) as nested random intercepts.
The treatment effect is presented as OR with corresponding 95%-CI and p-value. Marginal estimators (proportions of residents at t2) with corresponding 95%-CI are given per treatment group.

The endpoints 10 a iii. & iv. and b iii. & iv. will be analysed descriptively. Absolute and relative frequencies will be reported for the treatment groups.

At resident level:

All analyses relating to the residents of NHs are performed within a framework of multi-level models.

The secondary endpoints 13. – 17., 19. and 25. – 27. are count data. For these endpoints, a mixed Poisson regression model will be calculated with the corresponding number as dependent variable, realized follow-up time as offset, group (intervention vs. control) as fixed effects, the particular baseline measurement as covariate and region NHs and respective NH as nested random intercepts.
The treatment effect is presented as IRR with corresponding 95%-CI and p-value. Marginal estimators (numbers per 100 residents per past 12 -months observation period (from t1 to t2)) with corresponding 95%-CI are given per treatment group.

The secondary endpoints 18., 20. – 24., 28. – 29. and 31 are outcome variables with levels 1 = “yes” and 0 = “no”, i.e. binary outcomes. The secondary endpoint place of death (nursing homes vs. others) is also binary.
For binary outcomes, a mixed logistic regression model will be calculated, taking into account the respective binary outcome as dependent variable, group (intervention vs. control) as fixed effect, the particular baseline measurement as covariate and region of NHs and NH as nested random intercept.
The treatment effect is presented as odds ratio (OR) with corresponding 95%-CI and p-value.
Marginal estimators (proportions of residents at t2) with corresponding 95%-CI are given per treatment group.

The secondary endpoints 11 a. & b. and 12 are continuous and compared between intervention and control group at t2.
A linear mixed model will be calculated including the respective outcome at t2 as dependent variable, group (intervention vs. control) as fixed effects and region of NHs and NH as nested random intercept.

The treatment effect is presented as mean difference with corresponding 95%-CI and p-value.
Marginal estimators with corresponding 95%-CI are given per treatment group.

The secondary outcomes 30. and 32 b. are continuous and compared between intervention and control group as change from baseline at t2.
A linear mixed model will be calculated including change from baseline (t2-t0) as dependent variable, group (intervention vs. control) as fixed effects, the respective baseline value (t0) as covariate and region of NHs and NH as nested random intercept.
The treatment effect is presented as mean difference with corresponding 95%-CI and p-value.
Marginal estimators (mean change from baseline at t2) with corresponding 95%-CI are given per treatment group.
Outcome 30 is analyzed stratified by the residents' self-rating capability:

- for residents capable of self-rating the WHO-QoL German Version is used
- for residents incapable of self-rating the QUALIDEM score is used

The secondary outcome 32 a. is evaluated in two different ways. The resulting total score is analysed within the framework of a linear mixed model with the change from baseline as dependent variable, as described above. Additionally, a mixed logistic regression model based on the binary variable suspected post-traumatic stress disorder (yes vs. no) determined by $X > 0$ vs. $X \leq 0$ is calculated, as described above.

5.2 Sensitivity analyses

For the primary endpoint analysis, an evaluation with the negative binomial distribution to detect a possible but not expected overdispersion will be performed.
Further, for both primary and key secondary endpoints, per protocol analyses as well as adherence analyses based on NH and resident level will be performed.
Finally, an adjusted analysis is performed taking into account remarkable ($p < 0.05$) differences at baseline and t1, regarding the following potential confounders:

NH level:

- Occupied bed days (in the past 12 months)
- Sex ratio of residents
- Mean age of residents
- Rate of all deceased residents (in the past 12 months)
- Rate of full nurse equivalents per 100 residents
- Rate of palliative care qualified nurses per 100 residents
- Rate of general practitioners per 100 residents

- Defined private co-payment for residents (relative deviation from the regional mean)
- Total sick leave days per nursing staff (i.e. number of registered/assistant/hired nurses) (in the past 12 months)

Residents level:

- Age
- Sex
- Marital status
- Care level
- Professional degree

5.3 Missing data

The number of available observations and the number of missing observations of each variable will be presented for the intervention and control group separately, as well as for the total sample. All data will be analysed in an ITT approach.

The main analysis strategy is using all available data of all randomized NHs (full analysis set, FAS). For the primary endpoint (and all secondary endpoints related to DC1) no missing data is expected due to the nature of aggregated data collection.

For the unlikely situation of a substantial rate of missing values (>5%) in the primary outcome, a multiple imputation using chained equations will be applied in the ITT population for a sensitivity analysis. We will follow the recommendations of White, Royston and Wood (2011) in order to set up a proper imputation model assuming missing data to be missing at random (MAR). The imputation model includes simultaneously all above described variables (primary and secondary endpoints) of the DC1 at all measured time points.

The number of imputations (M) will be defined by the percentage of missing values in at least one of those defined variables at all time points. If the assumption of M=100*FMI (Largest Fraction of Missing Information) is fulfilled, at least 20 imputations will be performed.

5.4 Additional analyses

5.4.1 Subgroup analyses

To assess the effect of the structural ACP program within specific types of NHs or groups of residents, subgroups according to the following characteristics will be analysed regarding the primary endpoint and key secondary endpoints:

On NH level:

- Size of NHs (small (< 100 residents) vs. large (≥ 100 residents))
- Study centre (Göttingen, München, Düsseldorf, Halle)

- Ownership (non for-profit vs. municipal)
- Palliative care structure (at least two palliative care nurses (40h or 160h) and palliative care concept vs. none)
- Staffing (care hours per resident > median vs. ≤ median)
- Death rate per 100 residents (> median vs. ≤ median)

On resident level:

- Sex (male vs. female)
- Age (very old aged (> 90 years) vs. old (≤ 90 years))
- Resident capable of decision making (as expressed in who signed his or her consent, the resident or the proxy)
- Years since moving in (> 3,5 vs. < 3,5 Jahre)
- Care level (1,2,3 vs. 4,5)
- Level of education (no degree, vocational or school training or apprenticeship vs. training at a master craftsman/technical school, vocational or technical academy, technical college degree/university degree)

For each of the subgroups the above defined models for the primary endpoint and the key secondary endpoints are supplemented by the subgroup variable and its interaction term with the treatment group. In forest-plots (for each variable) marginal effect measures of the treatment effect and corresponding 95% CIs will be presented for the subgroups and the interaction p-value will be given.

5.4.2 Covid-19-Analysis

To assess the effect of the Covid-19 pandemic on the primary outcome, the following analysis is planned:

A mixed Poisson regression model is calculated taking into account group (intervention vs. control), time (t0, t1, t2) as well as the interaction between group and time as fixed effect, realized follow-up time as offset, the respective baseline hospitalization rate as covariate, and region of NHs and the respective NH as random intercept. The aim is to evaluate the effect of Covid-19 as external confounder on the primary outcome over time.

5.4.3 Possible pre-defined analyses not to be presented in the study report

All secondary endpoints may be analysed as defined in the different sensitivity or adjusted analyses, as well as regarding the pre-defined subgroups and the possible impact of the Covid-19 pandemic, if it is of interest.

Furthermore, sensitivity and subgroup analyses of the other (not key) secondary outcomes might be performed.

These analyses are pre-defined but will not be part of the study report.

5.5 Statistical software

- STATA 14 or newer
- R 3.4.1 or newer
- SPSS 26.0 or newer

References

[1] Gamble, C., Krishan, A., Stocken, D. et al. (2017). Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA, 318(23), 2337. https://doi.org/10.1001/jama.2017.18556

[2] Götze, K., Bausewein, C., Feddersen, B. et al. (2022). Effectiveness of a complex regional advance care planning intervention to improve care consistency with care preferences: study protocol for a multi-center, cluster-randomized controlled trial focusing on nursing home residents (BEVOR trial). Trials 23, 770. https://doi.org/10.1186/s13063-022-06576-3

[3] Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004). CONSORT statement: extension to cluster randomised trials. Bmj, *328*(7441), 702-708.

[4] Dichter, Martin & Ettema, Teake & Sorg, Christian & Meyer, Gabriele & Bartholomeyczik, Sabine & Halek, Margareta. (2016). QUALIDEM Benutzerhandbuch für die deutschsprachige QUALIDEM Version 2.0.