Predicting future unwell-being of hospice patient using real world patient reported outcome measures, evaluation of the practical application of a frequentist and Bayesian approach.

Student: Everlien de Graaf 9446427

Examiner: Anne May

Second reviewer: Hans Reitsma

Daily supervisors: René Eijkemans † and Rebecca Stellato

Number of words: Abstract:354 ; Summary:427 ; Main text =5089 (excl. tables)

Number of appendices: 4

Predicting future unwell-being, the practical application of two methods

Abstract

Background

In palliative care, Patient-Reported Outcome Measures (PROMs) gauge symptom severity. These PROMs help develop predictive models supporting proactive care. Longitudinal data prediction models in palliative care need to incorporate survival, and joint models, though appropriate, have not been applied in this context.

Objective

Evaluate the practical application of frequentist and Bayesian joint modeling for predicting future hospice patient unwell-being using real-world PROMs.

Methods

Design: a mixed-methods approach, combining prospective cohort design for practical application and cross-sectional design for evaluation. Utilize the Sympal cohort from Dutch hospices (August 2015 – May 2023). Data collected in clinical practice were entered into the research database including, patient characteristics, illness characteristics and symptoms and concerns assessed by means of the Utrecht Symptom Diary – four dimensional (USD-4D) clinical practice.

Outcomes: Assess processing, congeniality, software implementation, practical application, and estimate comparisons. Predictors include pain, sleep disturbance, dry mouth, dysphagia, anorexia, constipation, nausea, dyspnea, fatigue, anxiety, depressed mood, time for oneself, bearing life events, letting go of loved ones, feeling harmony in life, being at peace with the end of life, unwell-being and value of life, assessed by means of the USD-4D. In addition, patient characteristics and illness characteristics were added to the models.

Analysis: Conduct descriptive analysis for the evaluation and employ frequentist and Bayesian joint modeling in three steps: 1)multiple imputation, 2)linear mixed modeling for model optimization, and 3) joint modeling combining linear mixed model and Cox proportional hazards model.

Results

Joint models using frequentist and Bayesian approaches predicted future unwell-being, life value, pain, dry mouth, anorexia, fatigue, depressed mood, and life balance with comparable estimates. Bayesian approach required ten times more computation time. Frequentist approach analysis lacked congeniality,

requiring additional programming during multiple imputation, linear mixed models, and joint modeling. Bayesian approach necessitated a specific statement on imputation covariates. Both methods are not fully implemented in statistical software, limiting validation and use for future patients.

Discussion / Conclusion

Future unwell-being prediction, accounting for survival, involves physical, psychological symptoms, and sociospiritual concern. Despite longer computational times, the Bayesian approach fits the analysis better. Additional implementation is essential for applying Bayesian models in developing predictive models for future palliative care.

Summary

In taking care of very sick patients, doctors and nurses use feedback from patients to understand how bad their symptoms are. This helps them talk about and treat the symptoms and worries of the patients in the best way. The information from this patient feedback can also help predict how well the patient will be in the future and which symptoms might get worse. These prediction models can help caregivers give proactive care and inform patients and families about what to expect. Statisticians and epidemiologists can support clinical care by creating and validating prediction models using data from patients in past years to support decision making for future patients.

To create prediction models in palliative care, the time to death needs to be considered, as making predictions beyond death is not possible. Joint modeling approaches allow the development of prediction models using real-world data from patients collected during their illness, considering the time to death. Two approaches, a frequentist and a Bayesian approach, are described in literature, but their practical use in real-world symptom severity data is not yet described.

We evaluated the practical application of these methods on real patient data, looking at symptoms like pain and fatigue, as well as social and existential concerns of patients admitted to hospices in the Netherlands. Since these data were collected in clinical settings, they were not complete. Missing data were studied and seemed to be linked to illness and other unmeasured causes. Therefore, missing data had to be handled as well, to prevent flaws in the prediction models. This resulted in an analysis plan for missing data using multiple imputation and joint modeling using the frequentist approach and the Bayesian approach.

Results showed that both methods gave similar prediction models, but the application differed. The Bayesian method took ten times longer for the calculations. The frequentist method had its own issues, like not being able to handle missing data properly, leading to lower quality prediction models. Handling missing data and analyzing them following the frequentist approach needed additional assumptions and extra computer work. Finally, both methods were not fully ready for use in the software doctors currently use.

The study found that a person's future well-being is connected to their physical health, mental well-being, and social/spiritual concerns, considering how long they might live. Even though the slower method seemed to work better, both methods need more improvements to be truly helpful for doctors and nurses in the future. The study highlights the importance of refining these methods so that

prediction models can be developed and tested to genuinely support the prediction of future well-being of patients and optimize care and treatment for patients nearing the end of life.

Predicting future unwell-being, the practical application of two methods

Background

Hospice care aims to optimize the quality of life of patients in the last months of life. At admission, patients suffer from, on average, seven symptoms concurrently.[1] Symptom burden can be assessed in daily practice using patient reported outcome measures (PROM).[2, 3]

PROMs are self-reported measures, and provide insight into the severity of the patient's symptom burden, support clinical decision making and support communication between the patients, families and the multi-professional team.[2] Monitoring symptoms as a foundation of palliative care optimizes quality of life.[4-6] Supporting clinicians to early identify patients at risk for increased symptom burden or decreased quality of life, could optimize proactive palliative care and treatment.[5]

Research aiming to use PROM data mainly focused on overall symptom burden, prevalence and intensity and symptom burden in specific subgroups.[3] Prediction studies, aiming to reveal the underlying mechanisms of multidimensional symptoms have been increasingly performed. The analyses proved to be difficult, mainly due to non-normal distributions, attrition, irregular follow up times, and missing values. [7-11] Different solutions have been used by researchers, e.g., dichotomizing data at different cut-offs, categorizing data, or using cross sectional analyses for longitudinal data.[9] These studies contributed to the body of knowledge of symptom burden in palliative care and to some extent unravelled the underlying mechanisms. Models to predict future symptom burden or quality of life would support proactive palliative clinical care, but have not yet been developed.

The prediction of future well-being in hospice patients has to take survival into account, as the prediction of well-being beyond death is not feasible nor desirable. The statistical joint modelling method combines a time to event analysis with a mixed model analysis.[12, 13] Joint models are predominantly used to perform a time to event analysis while correcting for time varying covariates. However, this method is also suitable to analyze a longitudinal outcome, tacking survival into account. Two statistical approaches for joint model analysis are described in literature, a frequentist approach[13] and a Bayesian approach.[14] Both methods are suitable for the longitudinal analysis of PROMs.[15, 16]

PROMs, specifically in palliative care, are known to be problematic with respect to attrition and missing items.[17] Data collected in hospice care are prone to complex patterns of missing data. PROMs of patients with a limited life expectancy, collected during daily care, are hypothesized to be even more challenging. Although joint models can handle missing outcome measurements, missing values of predictors is still problematic and could lead to biased results. Analyzing complete data would 1) decrease the amount of

data and 2) increase the risk of biased estimates as the missing data are likely related to specific patients or patient characteristics. Dealing with missing data is essential to ensure validity of the estimates and increase the efficiency of the analysis by using all available data.[15, 18]

Three mechanisms of missing data can be identified, missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). When data are missing at random, no additional actions need be undertaken as the missingness will not affect the estimates. If data follow MNAR and MAR mechanisms, estimates will be affected and less valid, therefore multiple imputation strategies are preferred to minimize this effect.[19] Multiple imputation is a valid approach for dealing with missing data, introducing additional uncertainty to the estimates to correct for the imputation. Which imputation methods are chosen depends on the data analysis planned. Imputation methods should be congenial to the analysis methods, to ensure alignment of the data closely to the true data structure.[20] For multiple imputation, both frequentist and a Bayesian approaches are available.[21, 22]

This work is motivated by the research question: how can symptoms and needs be used to predict future well-being of hospice patients? The aim of this project is to evaluate the practical application of both a frequentist and a Bayesian approach to predict future well-being using real world PROM data of hospice patients and to compare the estimates of the two methods applied.

The ultimate aim of this study is to pave a foundation of future dynamic prediction models to support proactive palliative care for all patients.

In congruence with the dual aims of this study, two occurrence relations under study are:

1) Practical application → F(frequentist approach joint model, Bayesian approach joint model)
2) Future unwell-being → F(survival, unwell-being, value of life, symptoms and concerns, patient characteristics, time|patient)

Study Design

A descriptive mixed method study was performed using both a longitudinal cohort study design to predict future well being of hospice patients by means of their symptoms and concerns and patient characteristics, taking survival into account, and a cross sectional observational study to evaluate the application of the frequentist and the Bayesian approach.

Description of the SYMPAL cohort, setting and population

The domain is patients with an estimated life expectancy of less than three months. The symptoms in palliative care (SYMPAL) study cohort consists of hospice inpatients, admitted from August 2015 – May 2023 to one of the 15 hospices participating hospices (all members of the Association of Hospice care in the Netherlands). Hospices all used a specific electronic patient record, Madenco, and implemented the assessment of the Utrecht Symptom Diary at least twice a week as part of standard care. All adult patients who completed Utrecht Symptom Diaries at least twice during their admission in a hospice were eligible for inclusion. At admission, patients received information about research studies performed by means of anonymized clinical and administrative data. Broad consent was requested for the use the data for research purposes. If consent was obtained, the data collected during hospice care were electronically anonymized and entered the study database SYMPAL.

Outcomes

The main outcome of this study is practical application, which is defined as:

1) Processing time in minutes and seconds per imputation and analysis.
2) Congeniality defined as the appropriate fit between imputation methods and analysis methods.
3) Implementation of software necessary to perform the frequentist and Bayesian approaches.
4) Practical application and comparison of the estimates of the joint models using the two approaches on the data from the SYMPAL cohort.

Description of the outcome and predictors of the SYMPAL cohort data, used for the practical application.

The outcome of the prediction model is future well-being, defined as well-being experienced by a patient assessed at a given point in time (T+1), predicted by the covariates assessed at the previous point in time (T).

The determinants are at patient level (level 2) and at measurement level (level 1). Level 1 determinants are collected during the admission at multiple timepoints during the hospice admission.

- Symptom intensity is the severity of symptoms and concerns as experienced by patients at T
- Unwell-being is the unwell-being experienced by patients at T.
- Value of life is the value of life experienced by the patient at T.
- Time is the number of days between admission and T.
- Lead length is the number of days between the assessment of future unwell-being at T+1 and symptom intensity at T.

Level 2 determinants are collected once per patient at admission.

Demographics:

- Age in years.
- Gender: male, female, other.
- Living situation: home, hospital, nursing home, elderly care, other, unknown
- Philosophy of life: none, religious, not active, spiritual, unknown.

Illness characteristics:

- Diagnosis: cancer, organ failure, combined, other
- Performance status defined as the patients level of functioning.

The longitudinal model will be adjusted for all-cause mortality:

- All-cause mortality is dichotomized where 1 is death and 0 is censored.
- Survival time is defined as the number of days between admission and death. When a patient is censored, survival time is defined as the last moment the patients was known to be alive, and calculated as the number of days between admission and the date of release from hospice or the last assessment of a patient is applicable.

Measurements

The Utrecht Symptom Diary- four dimensional (USD-4D) is a Patient Reported Outcome Measure (PROM) used to assess the severity of physical and psychological symptoms, social and existential concerns and unwell-being and perceived value of life in clinical practice (Appendix 1). The USD-4D is a valid, Dutch-adapted translation of the Edmonton Symptom Assessment System.[23] The USD contains nine physical symptoms (pain, sleeping disturbance, dry mouth, dysphagia, anorexia, constipation, nausea, dyspnoea and fatigue), two psychological symptoms (anxiety and depressed mood) and five social and existential concerns (time for yourself, bearing what happens, letting loved ones go, feeling harmony of life, being at peace with the end of life). In addition, the USD-4D entails a 1 item unwell-being measure and a 1 item perceived value of life measure.[24] All symptoms are assessed using an 11-point numerical scale (0=no symptom, best possible to 10=worst intensity, worst possible). The content validity of the USD4D has been established from a patient and clinician perspective, as was the construct validity based on hypotheses testing per item.[24] All patients are invited to complete the USD-4D twice a week as standard hospice care.

Performance status was assessed by means of the ECOG performance status scale. The ECOG grades the patient's level of functioning in terms of self care, daily activity and physical ability, from 0 (= fully active) to 5 (= dead).[25]

Data analysis

Analysis of primary outcome

For the analysis of the practical application of the two approaches qualitative descriptive analysis was performed.

Analysis of the SYMPAL cohort data, the practical application

The joint model analyses consisted of two parts: 1) a linear mixed effects model analysis and 2) a joint model was fitted to the data. Due to missingness in covariates, the missing data patterns were first analyzed. Results indicated that most covariates were either missing at random or missing not at random.[18] Therefore, a multiple imputation approach was applied before model building. Data used in the imputation were: all symptoms included in the USD-4D, age, performance status, diagnosis, living situation and philosophy of life. To ensure a fit between the multiple imputation method and the data analysis methods, two methods for imputation and analyses were performed: a frequentist [26] and a Bayesian approach.[15, 22]

To compare the frequentist and Bayesian approach to joint models, the data from the Sympal cohort were used. In both approaches, optimization of the joint model was performed in two steps. First the linear mixed effect model was fitted to the data using backward selection. Next the joint model was fitted using both the linear mixed effects model and the Cox proportional hazards model. All analysis were performed in R.

FREQUENTIST JOINT MODELLING APPROACH

Step 1. Perform multiple imputation

Multivariate Imputation by Chained Equations (MICE) was used to enable multilevel imputation.[27] The imputation method must fit the analysis method used for the final analyses to ensure congeniality. This can be ensured for the mixed method part of the analyses. To optimally fit the data structure, predictive mean matching is used to fit the level 1 variables, USD-4D symptoms and concerns and level 2 variables, patient-level data, as the data are not normally distributed. In addition, the grouping variable is set to

patientID, which is the patient identification variable. Since a theoretical rationale for the selection of interactions was lacking, no interactions were included in the imputation process.

The IntraClass Correlation (ICC) of the original data was compared to the ICC of the imputed dataset to select the most appropriate imputation structure and number of iterations.[26] In total, five imputed datasets were used for this exploratory analysis and 50 iterations. Iteration and convergence were checked using density plots and iteration plots.

After the multiple imputation procedure was performed, the data were transformed to a long format to enable the procedure to include the outcome future unwell-being, using unwell-being assessed at T+1 and copy this value to the assessment at T creating the outcome variable future unwell-being.

Step 2. Building the linear mixed effects models

Next, to select predictors for the mixed model part of the joint model a pooled linear mixed effects model (LME) was built using the multiply imputated data set. The pooled full model contained all demographics, illness characteristics, and symptom burden as fixed effects in the model with a random intercept and a random slope for time. Using backward selection, the model was refined, excluding the predictor with the highest p-value. The new pooled model was tested against the former model using the Wald test based on Rubin's rules, which is recommended for variable selection in multiply imputed data, preserving type 1 error.[28] This continued until only significant predictors ($p < 0.1$) were left in the pooled model.

Step 3. Building the joint model

To fit a pooled joint model the following steps were performed for each imputed data set:

1. The final LME model was fitted to the five separate multiply imputation datasets.
2. The Cox Proportional Hazards regression model was fitted to the five separate multiply imputation datasets. This Cox Proportional Hazard regression model used baseline performance status as a covariate, since it is known to be associated to survival and was deleted from the mixed model.
3. A joint model was fitted to each imputed dataset, using the linear mixed effects models and Cox Proportional Hazards model. Per joint model, the target estimated coefficients and their variance/covariance matrix was saved.

4. To enable a pooled joint model, two lists were created in R, one for all estimated coefficients and the second for the variance-covariance matrix.
5. Finally, the pooled joint model was estimated by pooling the coefficients and covariances. Only the estimates of the longitudinal part of the model are presented as the CoxPH was only used to correct these estimates for survival.

BAYESIAN JOINT MODELLING APPROACH

In a Bayesian approach, using the package Joint AI, the imputation and analysis are fitted in one model, ensuring congeniality.[29] Therefore, to fit a joint model, two steps are described: optimizing the linear mixed effect model; and subsequently fitting the joint model.

Before the Bayesian approach was applied, unwell-being at T+1 was copied to the previous assessment (T) as the outcome, as it was not possible to perform this procedure between imputation and analysis.

Step 1. Multiple imputation and building of the linear mixed effects model

To optimally fit an LME model, backward selection was applied. Starting with the full model, using all demographic, illness characteristics and symptom burden predictors as fixed effects and random intercept and random slope for time. At every step, the predictor with the largest p value was deleted from the fixed effects of the model. If a covariate was deleted from the model, this covariate was entered in the model as an auxiliary variable if less than 20% was missing. Backward selection continued until all predictors had a p value < 0.1.

Step 2. Multiple imputation and building of the joint model

The final linear mixed effects model fitted to the data was used in the joint model. In addition, a Cox Proportional Hazards Ratio model was fitted to the data, using ECOG performance status and unwell-being as covariates and patient as random effect. Age, gender and length of lead were added to the model as auxiliary variables.

In line with the frequentist approach, the results of the survival analysis are not presented as it was only used to remove bias in the longitudinal process and prevent prediction beyond death.

The Markov Chain Monte Carlo settings were set on 100 adaptations, 1000 iterations, and three chains. Two model checks were performed 1) the Delman – Rubin criterion for convergence should be approximately 1.0 and the Monte Carlo Error (MCE/SD) should be close to zero.[29]

Results

Between August 2015 and April 2023, 12220 USD measurements of 1835 unique patients entered the Sympal database of which 3819 USD were self-reported of 1237 unique patients and were eligible for the frequentist approach and 739 had completed outcomes and were eligible for the Bayesian approach.

Patients were 75 years, 675(54.5%) women ,cancer was the primary diagnosis of 634 patients (51.2%) and most of them stayed at home before admission. During admission 1048 (84.7%) died. In total 739 patients have a complete outcome measure. Since the frequentist approach handles missing outcomes but the Bayesian approach does not, the characteristics of the patients with and without completed outcome measure are presented (table 1). There are no differences between both groups.

Table 1. Patient characteristics

Item		**Data used for frequentist approach All eligible patients**		**Data used for Bayesian approach Patients with complete outcomes**	
Patients		**1237**	100.0	**739**	100.0
Age	**Median [IQR]**	76 [68-84]		75 [67-84]	
Gender	**Man**	561	45.4	335	45.3
	Woman	675	54.6	403	54.5
	Other	1	0.1	1	0.1
Place of stay	**Home**	348	28.1	227	30.7
	Hospital	203	16.4	126	17.1
	Nursing home	17	1.4	5	0.7
	Elderly care	67	5.4	45	6.1
	Other	24	1.9	13	1.8
	NA	586	47.4	323	43.7
Diagnosis	**Cancer**	634	51.3	409	55.3
	Organ failure	29	2.3	18	2.4
	Cancer and organ failure	59	4.8	36	4.9
	NA	515	41.6	276	37.3
Philosophy of life	**None**	53	4.3	31	4.2
	Religious	148	12.0	96	13.0
	Spiritual	47	3.8	35	4.7
	Not practicing	218	17.6	151	20.4
	NA	771	62.3	426	57.6
Functional status	**Ambulant**	80	6.5	56	7.6
	Less than 50% in bed	249	20.1	180	24.4
	More than 50% in bed	330	26.7	203	27.5
	Bedbound	87	7.0	39	5.3
	NA	491	39.7	261	35.3
Event	**Yes**	1048	84.7	624	84.4
	No	189	15.3	115	15.6
Duration of stay	**Median [IQR]**	31 [18-52]		39 [5-61.5]	

Missing data

Analysing the missing data patterns (Appendix 2), it becomes apparent that more items are missing when the illness progresses and when the functional status declines. Furthermore, the sociospiritual items are more frequently missing, indicating MAR mechanisms. Due to a failure, measurements of the 'Balance in life' item were not stored correctly for the six months of setting up the database, thus MCAR. Furthermore, MNAR mechanisms cannot be excluded, due to the amount of missingness and limited data available.

Table 2 Missing data percentages

	Missing % **Frequentist approach**	**Missing%** **Bayesian approach**
Unwell being	15 %	7 %
Value of life	14 %	10 %
Pain	4 %	0 %
Dry mouth	5 %	1 %
Anorexia	3 %	2 %
Fatigue	8 %	1 %
Depressed mood	10 %	3 %
Balance in life	37 %	32 %
Days	15 %	0 %

ANALYSIS TIME

The frequentist method with five imputation sets took 6 minutes and 50 seconds to be performed. The analysis had to be performed manually and took 6 minutes and 1 second for five imputation sets. The Bayesian method took 7 hours, 8 minutes, and 46 seconds.

CONGENIALITY

The Bayesian method ensures congeniality since the analysis method used is used for the performance of the imputation method. In using the frequentist methods, congeniality is not completely ensured since the time to event data is not used in the imputation model.

IMPLEMENTATION OF SOFTWARE

Frequentist approach

The multiple imputation by chained equations software needs specification of the data structure and multilevel data to be imputed. The patient level USD variables were mostly zero inflated, however it was not a possible to set the methods for the imputation of multilevel data, therefor the methods were set at predictive mean matching using 2l.pmm for the measurement level variables and 2lonly.pmm for the patient level variables. The prediction matrix was set, using patient ID at -2 being the grouping variable. Based on the analysis of missing data, variables were manually added to the prediction matrix. Theoretically, interactions should be added for models with random intercepts and random slopes for incomplete level-1 and level-2 variable. Due to the number of variables, this would quickly inflate the number of of parameters in the model.

After imputation the leaps could be calculated. As a result, the imputed data was used optimally for the analysis.

For the joint model analysis of multiple imputation data, the analysis had to be performed for each imputed data set separately and only then could the model estimates be pooled. This required additional programming for which the procedures are not yet documented nor published.

Bayesian approach

The JointAI package is under development. The first analysis implemented and developed were linear mixed models, joint models were implemented but not fully operational. Obtaining the imputed datasets from the JointAI is possible, however it requires additional programming. Some bugs were identified, the main being that the mcmc.list was not identified to obtain a traceplot and denceplot although the MCMC list is available in the joint model output, but the naming is not identical to the underlying script. For regular LME the functions to obtain the traceplot, densplot and retrieving the multiple imputation data were implemented for checking performance and analysing the imputed datasets. Furthermore, for joint models the imputed data could not be obtained. In addition, application of the developed joint model for validation and the prediction of unwell-being to be used for future patients is not yet possible.

APPLICATION OF THE FREQUENTIST AND BAYESIAN JOINT MODELLING APPROACH

Frequentist joint modelling approach

Step 1. Perform multiple imputation

The imputation structure used was checked by comparing the ICC of the observed data and the imputation (Table 3). This was optimized by using predictive mean matching for two level data. The

symtoms and concerns were used except 'thought about the end of life' due to the percentage of missingness. The prediction matrix was set using ID-patient as the grouping variable and defined per variable using all variables with less than 20% missing. Although an increase in number of imputation sets would have improved the imputation, at 50 iterations convergence was optimized as the iterations plots intermingle and were free of any trends in later iterations (Appendix 3). The density plots of the imputed densities look very reasonable (Appendix 4).

Table 3 Intraclass correlation of observed and imputed data

Variables	Observed	MICE imputation
Pain	0.456	0.455
Sleeping problems	0.437	0.439
Dry Mouth	0.545	0.546
Dysphagia	0.577	0.581
Anorexia	0.537	0.539
Constipation	0.425	0.424
Nausea	0.421	0.423
Dyspnea	0.580	0.583
Fatigue	0.525	0.522
Feeling different	0.374	0.372
Anxiety	0.541	0.540
Depressed mood	0.526	0.521
Unwell-being	0.425	0.416
Value of life	0.521	0.507
Time for self	0.429	0.425
Bearing what happens	0.532	0.499
Letting loved ones go	0.589	0.585
Balance in life	0.505	0.455
Thought about the end of life	0.580	0.519

Step 2. Building the linear mixed effects models

A mixed model was developed starting with the full model using all symptoms and concerns and all patient level information. Using a stepwise backward approach the subsequent covariates were removed from the model: Thoughts about the end of life (p= 0.835), feeling different (p= 0.936), ECOG (p=0.580,

p=0.958, p=0.775), philosophy of life (p=0.581, p=0.722, p=0.498), swallowing problems (p= 0.691), bear what happens (p= 0.672), anxiety (p= 0.590), diagnosis (p=0.565, p=0.541), dyspnea (p=0.686), gender (p=0.430, p=0.455), Time for self (p=0.425), Letting lovd ones go (p=0.396), living situation (p=0.025, p=0.508, p=0.132, p=0.370), age (p=0.203), sleeping problems (p=0.207), nausea (p=0.170), constipation (p=0.125).

Step 3. Building the joint model

Pooled results of the Frequentist Joint model indicated that unwell-being at T had the highest predictive value of future unwell-being followed by anorexia, pain and depressed mood. The complete model is shown in table 4.

Table 4 Joint model estimates using Frequentist method

	Mean estimate	**SE**	**2.5%**	**97.5%**	**Missing%**
Intercept	1.597	0.152	1.300	1.900	
Unwell being	0.096	0.030	0.036	0.156	15 %
Value of life	0.056	0.019	0.019	0.094	14 %
Pain	0.076	0.021	0.034	0.117	4 %
Dry mouth	0.042	0.017	0.008	0.075	5 %
Anorexia	0.076	0.017	0.043	0.108	3 %
Fatigue	0.076	0.020	0.037	0.114	8 %
Depressed mood	0.040	0.024	0.006	0.087	10 %
Balance in life	0.050	0.021	0.008	0.091	37 %
Days	0.013	0.004	0.005	0.020	15 %

Bayesian joint modelling approach

Following the Bayesian method, the outcome future unwell-being cannot be computed after imputation, as a result,first unwell-being, assessed at T+1 is copied to T as future unwell-being, the outcome. Subsequently only USD-4D measures with a complete outcomes were selected for the Bayesian approach, resulting in 3167 USD-4D measures from 739 unique patients including 624 events (table1).

Step 1. Multiple imputation and building of the linear mixed effects model

Starting with a full linear model, using a stepwise backward approach, the subsequent covariates were removed from the model: Thoughts about the end of life (p= 0.896), feeling different (p= 0.856),

philosophy of life (p=0.615, p=0.92, p=0.184), diagnosis (p=0.789, p=0.589), dyspnea (p=0.856), ECOG (p=0.065, p=0.424, p=0.715), anxiety (p= 0. 985), swallowing problems (p= 0.512), sleeping problems (p=0.409), Letting lovd ones go (p=0.475), Time for self (p=0. 404), Bear what happens (p= 0.155), Living situation (p=0.004, p=0.007, p=0.134, p=0.383), gender (p=0.499, p=0.501), age (p=0.445), nausea (p=0.121), constipation (p=0.123).

The final linear mixed effects model with only p<0.1 covariates, consisted of unwell-being, value of life, pain, dry mouth, anorexia, fatigue, depressed mood and balance in life.

Step 2. Multiple imputation and building of the joint model

For the imputation, all covariates with less than 20% missings were added as auxillary variables. The estimates show that unwell-being is the largest predictor for future unwell-being. All estimates are positive which is expected.

Table 5: Joint model estimates using Bayesian method

	Mean estimate	SE	2.5%	97.5%	Missing %	GR-crit	MCE/SD
Intercept	1.431	0.144	1.150	1.714		1.08	0.047
Unwell being	0.126	0.032	0.064	0.191	7 %	1.11	0.060
Value of life	0.055	0.019	0.019	0.093	10 %	1.00	0.027
Pain	0.071	0.022	0.029	0.114	0 %	1.00	0.024
Dry mouth	0.041	0.017	0.008	0.076	1 %	1.00	0.026
Anorexia	0.080	0.017	0.046	0.113	2 %	1.00	0.031
Fatigue	0.076	0.020	0.037	0.114	1 %	1.00	0.030
Depressed mood	0.047	0.024	0.001	0.093	3 %	1.00	0.027
Balance in life	0.069	0.021	0.028	0.110	32 %	1.02	0.033
Days	0.011	0.004	0.004	0.020	0 %	1.08	0.119

The Gelman Rubin criterion differs 0.11 maximum from one, indicating convergence of the model.In addition, the Monte Carlo Error devided by the SD differs 0.12 maximum from zero, indicating a good precision of the MCMC sample.

MODEL COMPARISON

The mean estimates of the two methods differ slightly mean difference of -0.005. The largest differences are observed in Unwell being -0.03 and Balance in life -0.02. The mean estimates and their 95% confidence interval are displayed in Figure 1.

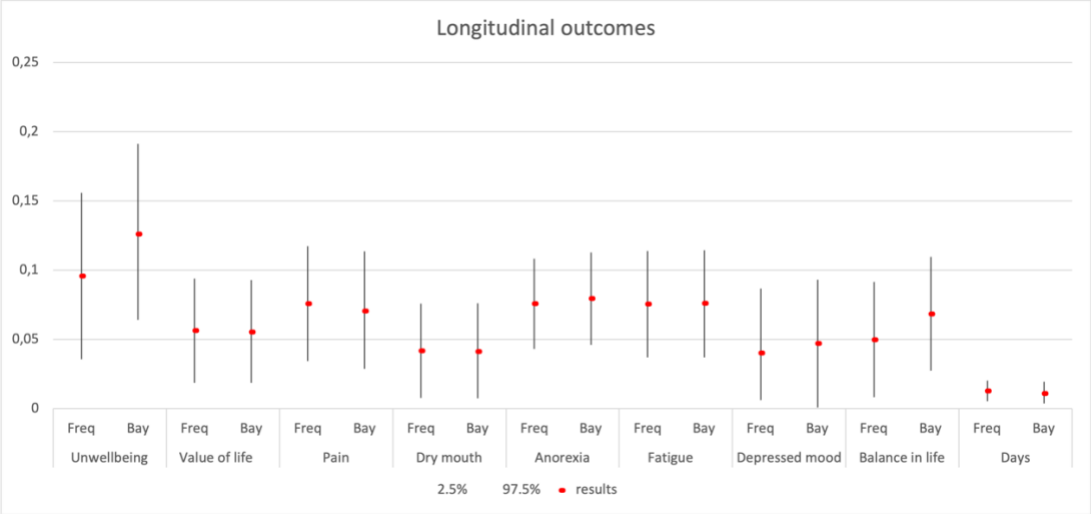


Figure 1. Visual comparison of mean estimates and CI of the Joint models using the frequentist and Bayesian approach

Discussion

The aim of this study was to compare the application of a frequentist method and a Bayesian method for the prediction of future unwell-being using PROMs of hospice patients, while taking survival into account, as we do not want to predict unwell-being after death. The application of both methods was possible, resulting in predominantly comparable results, except for unwell-being and balance in life. Unwell being as well as balance in life both have a stronger predictive value when the Bayesian approach is used, confidence intervals are comparable. This is probably due to slightly more biased estimates when the frequentist approach is followed, due to the larger proportion of missings in the original data.[15]

Frequentist or Bayesian approach?

The practical evaluation indicated pros and cons for both approaches. First, the Bayesian method ensured congeniality, was more time consuming and the application of the resulting model for future patients is not yet implemented in the software. The frequentist method was complicated due to a lack of possibilities to exactly match the imputation to the multilevel data structure and include survival ensuring congeniality between imputation model and analysis model. Second, both the multiple imputation and the application of the joint model using the results of the multiple imputation required additional programming in the frequentist approach. Ideally, interactions would have been added to the

imputation, inflating the numbers of predictors and the amount of programming. In addition, building the joint model had to be performed per imputed data set manually before pooling of the estimates was possible. For this application we limited the number of imputed datasets to five for practicality. This method is not described in literature or books and required expert consultation. Previous simulation studies comparing multiple imputation by chained equation and Bayesian methods concluded that Bayesian methods are more suitable to use in longitudinal data with missing covariates.[15] The structure of these real world data are comparable to the simulated data of this study, implying that the differences in estimates between the frequentist approach and the Bayesian approach found in this study could be caused by biased estimates of the frequentist approach. Thirdly, the computation time was ten times longer for the Bayesian approach. However, as stated before, if the time of the additional programming of the frequentist approach would be added to the total computation time, this time difference would be consideratly smaller.

According to the philosophy of Occam's razor, the most simple solutions should be preferred over the more complex solutions.[30-32] Applying Occams razor to these evaluation results, the Bayesian approach requires less additional programming, fewer assumptions on the part of the researcher and congeniality is guaranteed. Therefore, the Baysian approach should be preferred over the frequentist approach.

Strengths and weaknesses

The strength of this study is that it includes a large cohort of hospice patients, using data collected during clinical practice, which enabled us to study the use of the joint model approach in real world data. Many patients were able to assess symptom severity and the contribution of symptoms and needs in the physical, psychological, social, and existential dimension is in line with models of quality of life and well-being. Other predictors age, sex, diagnosis, and performance status were not significantly associated to future unwell-being. For diagnosis and performance status, this lack of association could be due to the limited documentation and the late stage of illness in hospice patients. The primary diagnosis was only categorized in cancer and organ failure. As most patients suffered from cancer, diagnosis did not discriminate enough to contribute to the prediction of future unwell-being. If data describing the specific diagnosis is available, it will need to be reevaluated for its association to future unwell-being. In addition, performance status is only assessed at admission. Since performance status is known to be associated to survival time, performance status at admission was entered into the Cox model as a predictor. Longitudinal assessments of performance status could have an added predictive value to the mixed model. Performance status has to be assessed multiple times during the illness trajectory to be

informative, especially when the objective is to dynamically predict future symptom burden and unwell being.[5] Finally, the current models did not apply penalization of the esimates, as this study aimed to evaluate the application of both methods using real world data. If joint models are further developed for future patients, penalization of estimates will need to be applied, followed by internal and external validation of models.[33] Furthermore, for the frequentist approach, the multiple imputation applied was a simplified method, omitting interactions and using a limited number of imputation. Although the ICC of the imputed data corresponds to the original data, these choices could have affected the accuracy of the imputed datasets and biased the estimates as a result. However, since the aim of the current study was the application and evaluation of the methods , this did not affect the results and conclusion of the evaluation.

The use of joint modelling for future patients

Although the implementation of the Bayesian method for joint models is not yet optimal and computation time is long, the approach is feasible for use in real world data. In addition, the models should incorporate penalized estimates to prevent overfitting and enable model testing and validation.[33] In comparison to traditional statistical sollutions, machine learning and artificial intelligence alternatives could also provide methods to perform prediction studies with longitudinal data.[34] Both the Bayersian approach as well as machine learning methods e.g., long short term memory neural networks are suitable for dynamic prediction.[35] Dynamic prediction model are continuously updated when new data become available and adapt the prediction based on this information.[36] Through dynamic prediction, learning from the patients' experience to predict future symptom burden, quality of life and survival seems promising to optimally support healthcare professionals. Patients and families also indicate a need for proactive support and thinking in future trajectories, to feel in control and safe.[37] The information from these dynamic models could be used to provide this information.[36] However, if healthcare professionals receive too many alerts, this will decrease the use of the models.[38] Furthermore, it is currently unknown what information would be beneficial for patients and their families.[39]

Therefore future research should focus on the development of these dynamic prediction models and the information and implementation needs of patients, families and health care professionals to support an optimal implementation dynamic prediction models and support proactive palliatice care for all patients.

Conclusions

The results of this study showed that both the frequentist and the Bayesina approach are feasible to be used for the prediction of future unwell-being of hospice patients. However, the Bayesian approach is the most simple approach, requiring fewer assumptions and less programming, and ensuring congeniality. As such, it should be recommended for use in the development of prediction models using real world PROMs in hospice care or other cases when prediction beyond death has to be prevented. Future research should study the Bayesian method for the development of dynamic prediction models and study the information and implementation needs of these models for patients, caregivers and healthcare professionals.

References

1. de Graaf E, Zweers D, de Graeff A, et al. Does age influence symptom prevalence and intensity in hospice patients, or not. *A retrospective cohort study J Geriatr Palliat Care* 2014; 20147.
2. Bausewein C, Simon ST, Benalia H, et al. Implementing patient reported outcome measures (PROMs) in palliative care-users' cry for help. *Health and quality of life outcomes* 2011; 9: 1-11.
3. van der Willik EM, Terwee CB, Bos WJW, et al. Patient-reported outcome measures (PROMs): making sense of individual PROM scores and changes in PROM scores over time. *Nephrology* 2021; 26: 391-399.
4. Coelho A, Rocha A, Cardoso D, et al. Monitoring and Management of the Palliative Care Patient Symptoms: A Best Practice Implementation Project. *Nurs Rep* 2022; 12: 365-370. 20220507. DOI: 10.3390/nursrep12020035.
5. García Abejas A, Serra Trullás A, Sobral MA, et al. Improving the Understanding and Managing of the Quality of Life of Patients With Lung Cancer With Electronic Patient-Reported Outcome Measures: Scoping Review. *J Med Internet Res* 2023; 25: e46259. 20230503. DOI: 10.2196/46259.
6. Koshy AO, Gallivan ER, McGinlay M, et al. Prioritizing symptom management in the treatment of chronic heart failure. *ESC Heart Fail* 2020; 7: 2193-2207. 20200805. DOI: 10.1002/ehf2.12875.
7. Gill TM, Han L, Leo-Summers L, et al. Distressing symptoms, disability, and hospice services at the end of life: prospective cohort study. *Journal of the American Geriatrics Society* 2018; 66: 41-47.
8. Hui D and Bruera E. The Edmonton Symptom Assessment System 25 years later: past, present, and future developments. *Journal of pain and symptom management* 2017; 53: 630-643.
9. Omran S, Khader Y and McMillan S. Symptom clusters and quality of life in hospice patients with cancer. *Asian Pacific journal of cancer prevention: APJCP* 2017; 18: 2387.
10. Tang ST, Liu LN, Lin K-C, et al. Trajectories of the multidimensional dying experience for terminally ill cancer patients. *Journal of pain and symptom management* 2014; 48: 863-874.
11. Zweers D, de Graaf E, de Graeff A, et al. The predictive value of symptoms for anxiety in hospice inpatients with advanced cancer. *Palliative & supportive care* 2018; 16: 602-607.
12. Chi YY and Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* 2006; 62: 432-445.
13. Rizopoulos D. *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press, 2012.
14. Rizopoulos D and Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine* 2011; 30: 1366-1380.
15. Erler NS, Rizopoulos D, Rosmalen J, et al. Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full Bayesian approach. *Stat Med* 2016; 35: 2955-2974. 20160404. DOI: 10.1002/sim.6944.
16. Enders CK, Mistler SA and Keller BT. Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological methods* 2016; 21: 222.
17. Rouanet A, Helmer C, Dartigues J-F, et al. Interpretation of mixed models and marginal models with cohort attrition due to death and drop-out. *Statistical methods in medical research* 2019; 28: 343-356.
18. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology* 2017: 157-166.
19. Groenwold RH and Dekkers OM. Missing data: the impact of what is not there. Oxford University Press, 2020, p. E7-E9.
20. White IR, Royston P and Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine* 2011; 30: 377-399.

21. Rizopoulos D, Verbeke G and Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* 2010; 66: 20-29.
22. Erler NS, Rizopoulos D, Jaddoe VW, et al. Bayesian imputation of time-varying covariates in linear mixed models. *Stat Methods Med Res* 2019; 28: 555-568. 20171025. DOI: 10.1177/0962280217730851.
23. van der Baan FH, Koldenhof JJ, de Nijs EJ, et al. Validation of the Dutch version of the Edmonton symptom assessment system. *Cancer medicine* 2020; 9: 6111-6121.
24. de Vries S, Lormans T, de Graaf E, et al. The content validity of the items related to the social and spiritual dimensions of the Utrecht symptom Diary-4 dimensional from a patient's perspective: a qualitative study. *Journal of Pain and Symptom Management* 2021; 61: 287-294. e282.
25. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American journal of clinical oncology* 1982; 5: 649-656.
26. Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research* 2007; 16: 219-242.
27. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med* 2016; 4: 30. DOI: 10.3978/j.issn.2305-5839.2015.12.63.
28. Wood AM, White IR and Royston P. How should variable selection be performed with multiply imputed data? *Statistics in medicine* 2008; 27: 3227-3246.
29. Erler NS, Rizopoulos D and Lesaffre EM. JointAI: joint analysis and imputation of incomplete data in R. *arXiv preprint arXiv:190710867* 2019.
30. Blumer A, Ehrenfeucht A, Haussler D, et al. Occam's razor. *Information processing letters* 1987; 24: 377-380.
31. Braithwaite JJ. Occam's razor: The principle of parsimony. *Academia edu* 2007.
32. Lycan WG. Occam's razor. *Metaphilosophy* 1975; 6: 223-237.
33. Moons K, Donders ART, Steyerberg E, et al. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. *Journal of clinical epidemiology* 2004; 57: 1262-1270.
34. Mandel F, Ghosh RP and Barnett I. Neural networks for clustered and longitudinal data using mixed effects models. *Biometrics* 2023; 79: 711-721.
35. Lin J and Luo S. Deep learning for the dynamic prediction of multivariate longitudinal and survival data. *Statistics in medicine* 2022; 41: 2894-2907.
36. Rizopoulos D, Molenberghs G and Lesaffre EM. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal* 2017; 59: 1261-1276.
37. de Vries S, Verhoef M-J, Vervoort SCJM, et al. Barriers and facilitators that hospital clinicians perceive to discuss the personal values, wishes, and needs of patients in palliative care: a mixed-methods systematic review. *Palliative Care and Social Practice* 2023; 17: 26323524231212510.
38. Warren M. Electronic Patient Reported Outcome Measures for Optimal Remote Symptom Management in Advanced Breast Cancer: Embracing Digital Health Technology. In: *Seminars in Oncology Nursing* 2023, p.151550. Elsevier.
39. Engel M, Kars MC, Teunissen SC, et al. Effective communication in palliative care from the perspectives of patients and relatives: A systematic review. *Palliative & Supportive Care*: 1-24.

Appendix 1. Utrecht Symptom Diary – four dimensional

Please circle the number that best reflects the intensity or severity of your symptoms or complaints.

The scores are provided by: 0 patient 0 loved one/ relative 0 caregiver

I have												
No pain	0	1	2	3	4	5	6	7	8	9	10	Worst possible pain
No sleeping problems	0	1	2	3	4	5	6	7	8	9	10	Worst possible sleeping problems
No dry mouth	0	1	2	3	4	5	6	7	8	9	10	Worst possible dry mouth
No dysphagia	0	1	2	3	4	5	6	7	8	9	10	Worst possible dysphagia
No lack of appetite	0	1	2	3	4	5	6	7	8	9	10	Worst possible lack of appetite
Normal stool	0	1	2	3	4	5	6	7	8	9	10	Worst possible abnormal stool
.....	0	1	2	3	4	5	6	7	8	9	10	.....
I feel												
No nausea	0	1	2	3	4	5	6	7	8	9	10	Worst possible nausea
No shortness of breath	0	1	2	3	4	5	6	7	8	9	10	Worst possible shortness of breath
No fatigue	0	1	2	3	4	5	6	7	8	9	10	Worst possible fatigue
Not different than usual	0	1	2	3	4	5	6	7	8	9	10	Very different than usual
Not anxiety	0	1	2	3	4	5	6	7	8	9	10	Worst possible anxiety
No depressed mood	0	1	2	3	4	5	6	7	8	9	10	Worst possible depressed mood
.....	0	1	2	3	4	5	6	7	8	9	10	.....
At this moment, I experience												
Best possible well-being	0	1	2	3	4	5	6	7	8	9	10	Worst possible well-being
At this moment, my life is												
Worthwhile	0	1	2	3	4	5	6	7	8	9	10	Not worthwhile

According to you, which symptoms have to be prioritized?

1.

2.

The scores are provided by: 0 patient 0 loved one/ relative 0 caregiver

The following question describe topics that might concern or worry you. Please circle the number that best reflects the intensity or amount of your worries or concerns.

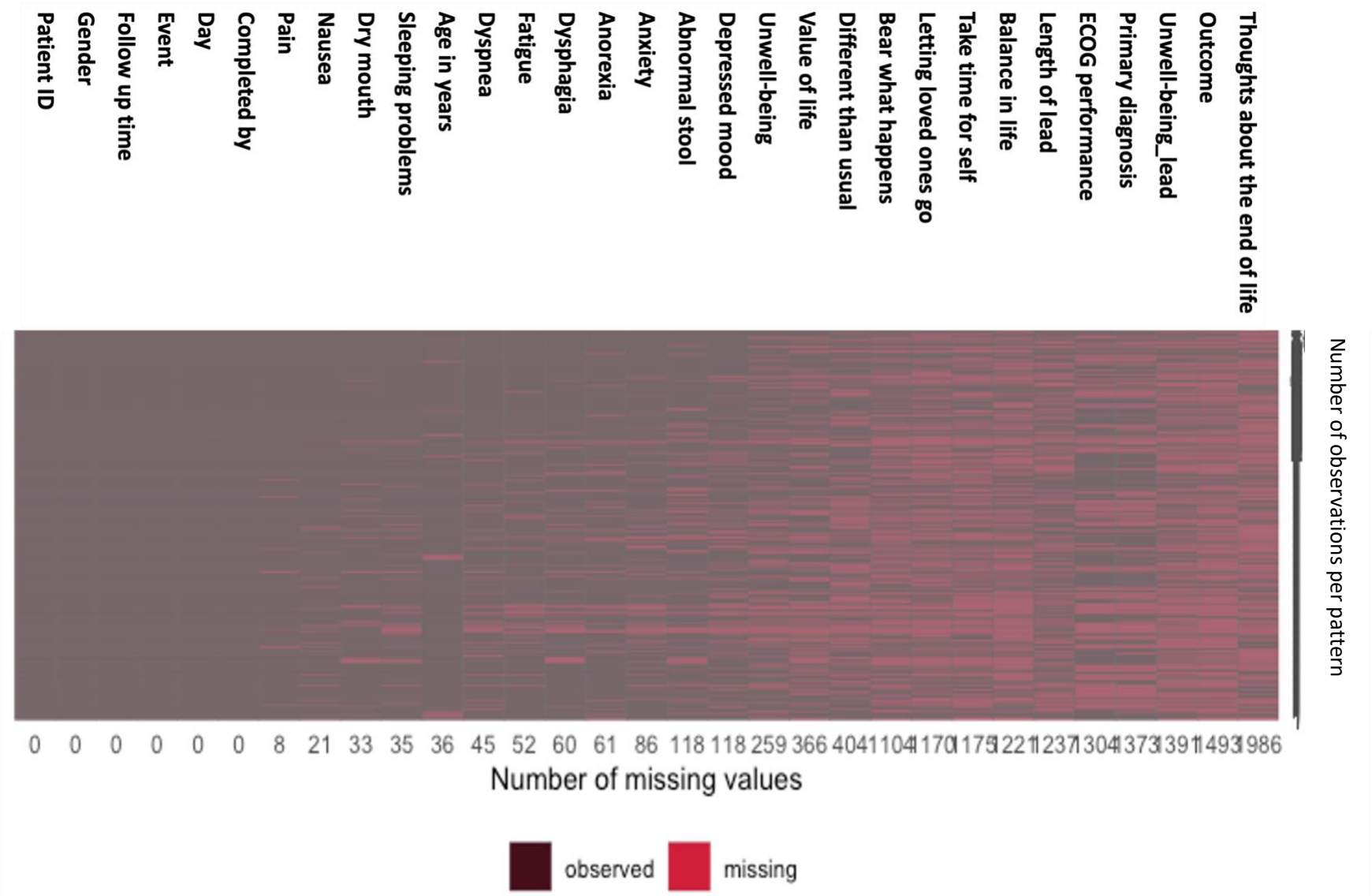
I take time for myself												
yes	0	1	2	3	4	5	6	7	8	9	10	Not at all
I can bear what happens to me												
yes	0	1	2	3	4	5	6	7	8	9	10	Not at all
I can let my loved ones go												
yes	0	1	2	3	4	5	6	7	8	9	10	Not at all
I feel a sense of balance in my life												
yes	0	1	2	3	4	5	6	7	8	9	10	Not at all
My thoughts about the end of life give me peace of mind												
yes	0	1	2	3	4	5	6	7	8	9	10	Not at all

Furthermore, I would like to let you know…..

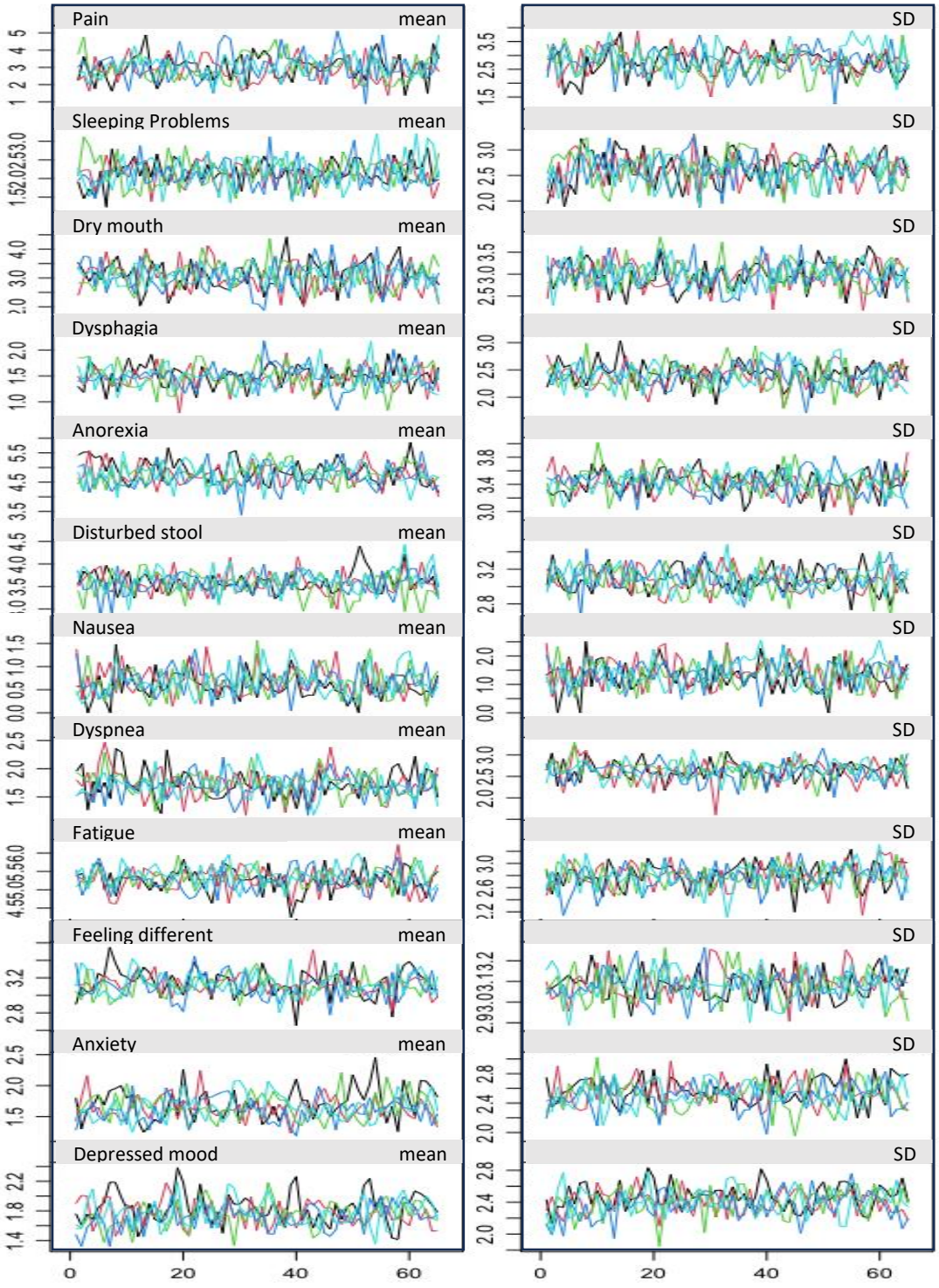
1.

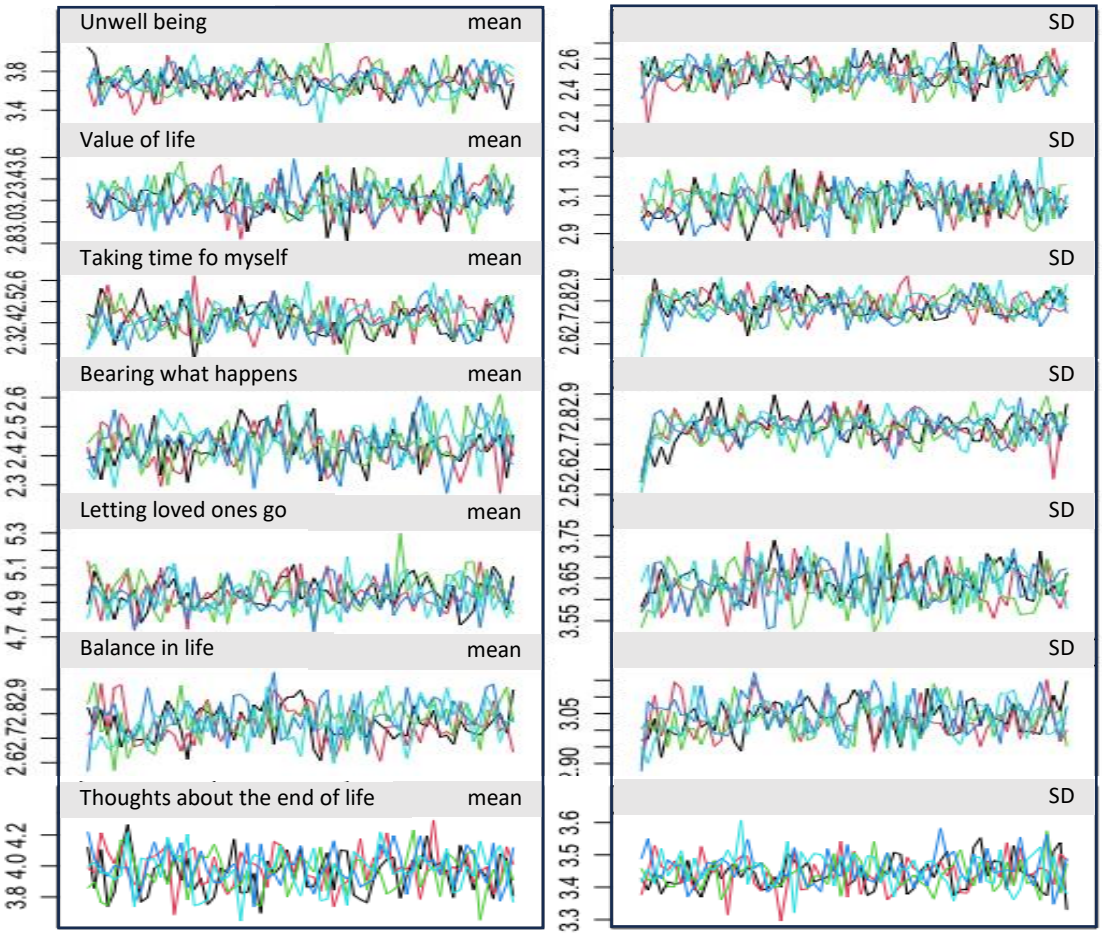
2.

Appendix 2. Missing data analysis

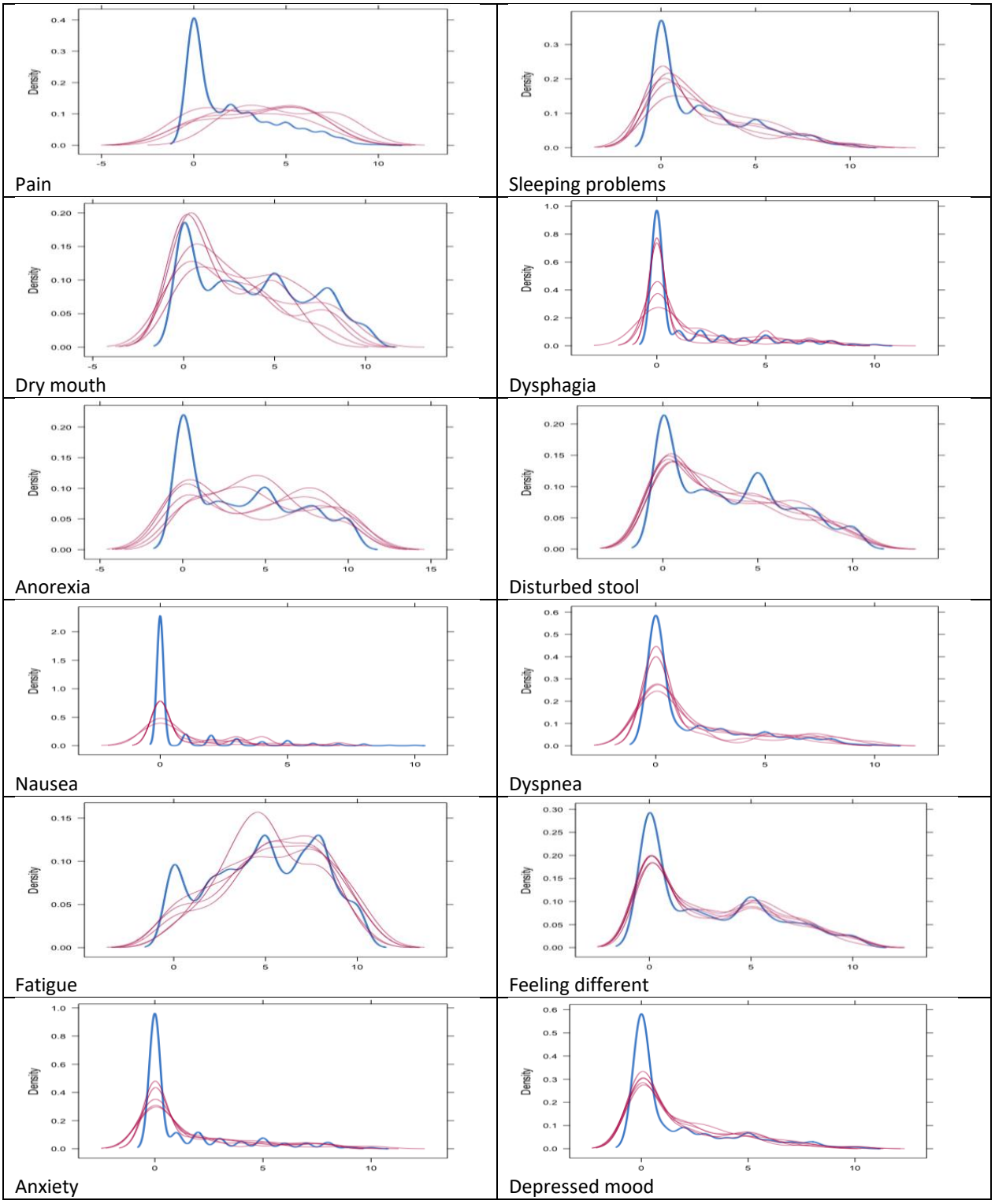


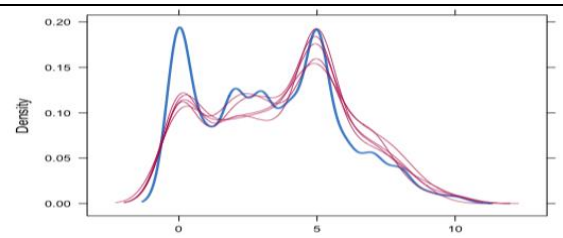
Appendix 3. Iteration plots of MICE multiple imputation



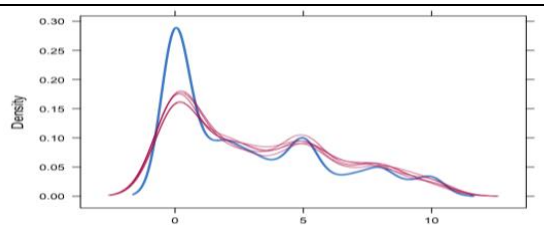

Unwell being
mean
SD
Value of life
mean
SD
Taking time fo myself
mean
SD
Bearing what happens
mean
SD
Letting loved ones go
mean
SD
Balance in life
mean
SD
Thoughts about the end of life
mean
SD

§Appendix 4. Density plots of MICE multiple imputation

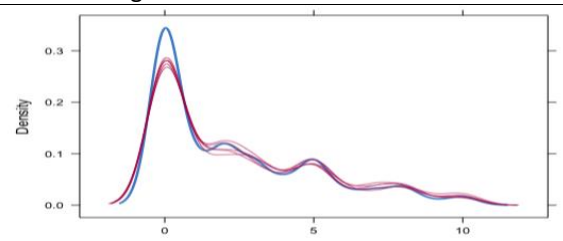




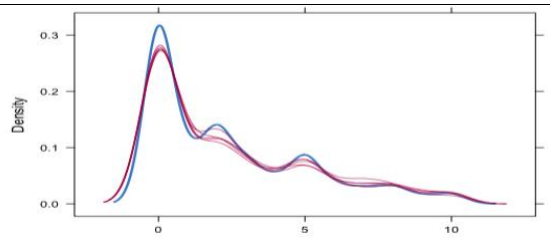
Unwell-being



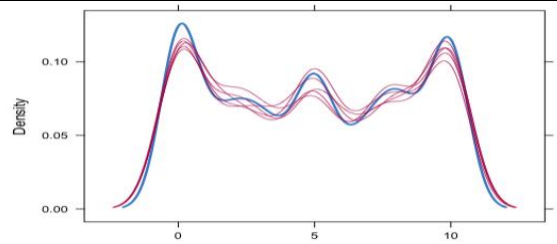
Value of life



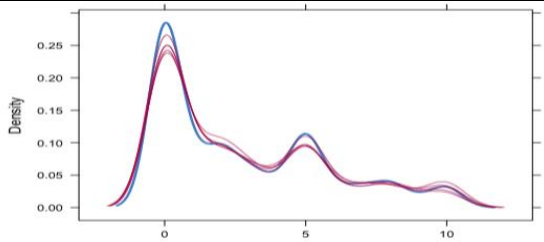
Taking time for myself



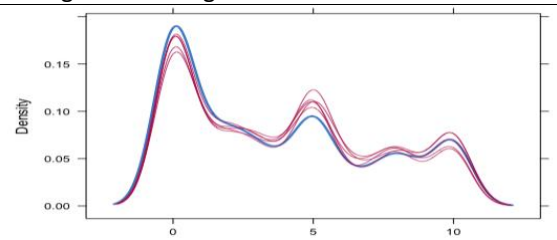
Bearing what happens



Letting loved ones go



Balance in life



Thoughts about the end of life