Health profiles among nonagenarians from Mugello district (Tuscany, Italy) and their socio-economic characteristics

Paper under review -
please do not circulate without written permit from the authors

Cosmo Strozza[1], Patrizio Pasqualetti[2], Viviana Egidi[1]

[1]Sapienza University of Rome, Italy
[2]Fatebenefratelli Foundation for Health Research and Education, Rome, Italy

Summary

Objectives

Studying the health conditions of the oldest-old is an important public health challenge. This study aims to investigate whether (1) it is possible to identify health profiles among the oldest-old, taking into account physical, emotional and psychological health; (2) there are demographic and socio-economic differences among the health profiles.

Methods

We applied a Latent-Class-Analysis to the Mugello Study data: 504 nonagenarians residing in Mugello (Tuscany, Italy).

Results

Four groups were identified: "healthy", "non-testable", "unhealthy", "semi-autonomous senile". Demographic and socio-economic differences were found among the groups: women and older-nonagenarians are more likely to be in worse health conditions; more educated nonagenarians are less likely to be in extremely-bad conditions; the lowest educated are more likely to be cognitively impaired; past-office-workers are less likely to have poor health than farmers.

Conclusions

Identifying health profiles and their socio-demographic differences could help finding groups to target with specific social policies and health care.

Introduction

Nowadays, world's population is aging and the burden of diseases is changing [1, 2]. As defined by the World Health Organization (WHO) in 1948, "health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [3]. This definition describes properly the complexity of the concept of health. The share of nonagenarians in Italy increased from around 0.06% in 1950 to 1.27% in 2015 and seems to keep growing, according to World Population Prospects [4], during the next years, reaching 4.34% in 2050. This phenomenon affects most of the developed countries. For this reason, identifying the oldest-old health profiles and characteristics is currently considered a big challenge [5]. The Latent Class Analysis (LCA) is perfect for dealing with multidimensional concepts as health is. Some scholars applied this approach on elderly (60+ years old and more) from different countries [6, 7, 8, 9] including in the analysis information about their physical, emotional and psychological status and obtaining, as a result, different health profiles. However, there is still not much evidence about health profiles among the oldest-old and the extremely-old [10]. Aiming to fill this gap in the literature we analyzed data from the Mugello Study [11], which included 504 nonagenarians from a rural area in Tuscany (Italy) called Mugello, to determine whether (a) it is possible to identify health profiles among the oldest-old and (b) if there are demographic and socio-economic differences among health profiles.

Methods

Study population and measures

The study population comes from the Mugello Study [11] aimed to evaluate the aging process among nonagenarians, living in 9 of the 11 municipalities of the Mugello area in Tuscany (Italy), focusing on different health aspects. It comprised 504 non selected individuals representing about 65% of the whole nonagenarians living in that geographical territory in 2012. Participation rate was 69% after excluding potential participants died before being interviewed or those who were not found. More information about the study design and survey methods are available in [12].
Many informations about the individual health conditions of the nonagenarians have been collected. Variables have been categorized according to the existing literature. Individuals who were not tested due to their (very) bad health conditions have been categorized as non-testable. Being non-testable is considered the worst health condition for each of the variables including this category. Cognitive function was measured according to the Mini-Mental State Examination (MMSE): the higher the score (0-30) the better the cognitive status [13]. It was divided into three categories in or-

der to distinguish people with severe (0-17), mild (18-23) and no cognitive impairment (24-30) [14]. Functional status was assessed according to the ability of performing five (eating, dressing, bathing, toileting, transferring) of the Activities of Daily Living (ADL) [15]. The number of ADLs that people could manage independently was used to distinguish between non- (0), semi- (1-4) and fully-autonomous (5) oldest-old [16]. Mugello's nonagenarians were divided into disease-free (0), single-disease (1) and co-morbid (2+) according to the number of chronic disease (cardiovascular, neurological, pulmonary, connective tissue, gastroenterological, endocrine, renal, oncological, immunodeficiency syndrome) reported. Geriatric Depression Scale (GDS) was used to evaluate depression status: the higher the score (0-15) the higher the level of depression [17]. It was divided into three categories in order to distinguish between non-depressed (0-4), depressed (5-15) and non-testable individuals [18]. Self-reported health (SRH) status was assessed using the Italian version of the Short-Form-12 questionnaire (SF12) from which it was possible to get the health perception of the individual according to the first item (in general, you would describe your health status as:) and two synthetic indicators combining the 12 items together: Physical and Mental Component Summary (PCS and MCS) [19]. The self-rated health was divided into three categories in order to distinguish between nonagenarians declaring excellent/very good/good health, acceptable/poor health and being non-testable. Also PCS and MCS have been divided into three categories: who scored more (or equal) than the average were considered in good health condition, who scored less than the average were considered in bad health condition and being non-testable.
Results are controlled for age (90-91, 92-94, 95+), gender, education (0-2, 3, 4-5, 6+ years of education) and main work did in the past defined according to Italian National Institute of Statistics (ISTAT) classification of jobs [20]: farmer, low (laborer or unskilled worker), medium (office or industry worker), high (entrepreneur, intellectual or technical worker) skilled work.

Statistical analysis

Health is a complex state involving different aspects or dimensions. To capture the heterogeneity of the health status among the oldest-old, we supposed that Mugello's nonagenarians could belong to an unobserved or latent class according to their health characteristics. For this purpose, we chose the Latent Class Analysis (LCA) that aims to group individuals into classes according to their indicator patterns. Each class includes individuals with similar characteristics, but different from those who are in other classes.
LCA was used to identify different health profiles according to the health condition through the variables described in the previous paragraph and controlling for demographic and socio-economic characteristics. LCA with covariates is an extension of the basic LCA permitting the inclusion of co-

variates to predict individual's latent class membership [21, 22].
We performed the LCA two times including the same variables: on the whole study population and on the subsample of testable individuals. Since we expected to obtain - in the first analysis - a group populated by only non-testable individuals, we excluded those people in the second analysis in order to capture more heterogeneity of the health status on the remaining oldest-old. The effect of the covariates has been estimated with the "one-step" technique in order to obtain less biased coefficients: they are estimated simultaneously as part of the latent class model [23].
Suppose a latent class model with Cclasses to be estimated according to m categorical variables and a covariate x. Let $Y_i = (Y_{i1}, ..., Y_{iM})$ be the vector of individual's response to the M variables where $Y_{im} = 1, 2, ..., r_m$. Let $c_i = 1, 2, ..., C$ be the latent class membership of the individual to the class, let $I(y = k)$ be the indicator function that is 1 if y is equal to k and 0 otherwise and let λ be the probability of membership in each latent class. Then the latent class model can be expressed as the following:

$$P(Y = y|x_i) = \sum_{c=1}^{C} \lambda_c(x_i) \prod_{m=1}^{M} \prod_{k=1}^{r_m} \rho_{mk|c}^{I(y_{im}=k)} \tag{1}$$

where $\lambda_c(x_i) = P(C_i = c|x_i)$ is a standard baseline-category for multinomial logistic model. In the case of one covariate, λ can be expressed as the following:

$$\lambda_c(x_i) = P(C_i = c|x_i) = \frac{exp\{\beta_{0c} + x_i\beta_{1c}\}}{1 + \sum_{j=1}^{C} exp\{\beta_{0j} + x_i\beta_{1j}\}} \tag{2}$$

for $c = 1, ..., C - 1$ where C is the reference class in the logistic regression. As a result, the log-odds of an individual to fall into latent class c relative to the reference class C, given x_i as value for the covariate, is the following:

$$log(\frac{\lambda_{c|c}(x_i)}{\lambda_{C|c}(x_i}) = \beta_{0c|c} + \beta_{1c|c}x_i \tag{3}$$

Multiple imputation was necessary to deal with missing values (Missing At Random - MAR) in order to avoid loss of precision in the analysis. K-nearest neighbor imputation method has been used for its high performance with survey data [**?**]. In order to obtain unbiased results, neighbors are found considering all the variables available in the dataset except the ones that are included in the models. Five neighbors have been considered to calculate the aggregated values to impute. Education, work did in the past, MMSE score, ADL performed, number of chronic disease, PCS and MCS have been imputed. Non of them had more than 7% missing values. More information about data imputation are included in Table A1.
Statistical analysis was performed using R version 3.5.0 [24] and VIM [25] and poLCA package [26].

Results

Participants were 504 with high prevalence of women (369): the sex ratio female/male of 2.73 confirms higher longevity of women. Mean age ± standard deviation was 93.1 ± 3.3 in the whole study population: men's mean age (92.5) resulted lower than women (93.3), t-test p = 0.01. Men were more educated: 64.5% of male vs 46.1% of female completed more than 3 years of school, but performed more physical jobs: 79.2% of male vs 52.6% of female were farmers or low-skilled workers. Overall, male showed a better condition in all the health measures considered in analysis. Large gender differences were found in the cognitive and functional status (60.7% of male vs 37.1% of female were not cognitively impaired; 61.5% of male vs 43.6% of female were autonomous). The gap in the remaining health measures is mainly due to the larger number of non-testable women (Table 1).
Three latent classes have been found both when considering the whole study population and the subsample of testable individuals. This number was chosen according to the "meaning" of the classes together with the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) which values are shown in Table 2. Every latent class has been labeled according to the posterior probabilities (λ) of finding a certain characteristic in the class showed in Table 3.
LCA performed on the whole study population resulted in three health profiles. The first class is characterized by high probability of being autonomous ($\lambda = 0.89$), not depressed ($\lambda = 0.80$), not cognitively impaired ($\lambda = 0.78$), perceiving good health ($\lambda = 0.92$) with values of PCS and MCS higher than or equal to the average (respectively $\lambda = 0.73$ and 0.65). This class, labeled "healthy group", includes 217 individuals (43.1% of the whole study population). The second class is characterized by high probability of being semi/not autonomous (respectively $\lambda = 0.47$ and 0.44), cognitively impaired ($\lambda = 0.97$) and not testable for depression ($\lambda = 0.97$) and self perceived health, including PCS and MCS indicators ($\lambda = 1$ for the three indicators). This class has been labeled: "unhealthy group". It includes 110 individuals (21.8% of the whole study population) which were almost all the non-testable nonagenarians according to the scales in analysis that included this category (self-rated health, depression, PCS and MCS). The third class includes nonagenarians with high probability of being semi-autonomous ($\lambda = 0.72$), mild/severely cognitively impaired (respectively $\lambda = 0.33$ and 0.40), depressed ($\lambda = 0.74$) and having PCS and MCS scores lower than the average (respectively $\lambda = 0.74$ and 0.66). Despite of how they performed in the objective health measures, they frequently declare a better health status: $\lambda = 0.43$ for declaring good health conditions is relatively high (bad self-perceived health: $\lambda = 0.57$). For this reason the last class - composed by 177 (35,1%) individuals - has been labeled: "unhealthy optimistic group".
LCA performed on the subsample of testable individual resulted in three

Table 1: Characteristics of the study population

Characteristics	Gender						p*
	Male		Female		Total		
	n	%	n	%	n	%	
Study population	135	26.8	369	73.2	504	100	
Age (m, sd)	92.5	2.8	93.3	3.4	93.1	3.3	<0.001
Education (years)							
0-2	16	11.9	49	13.3	65	12.9	<0.001
3	32	23.7	150	40.7	182	36.1	
4-5	63	46.7	142	38.5	205	40.7	
6+	24	17.8	28	7.6	52	10.3	
Work (level)							
farmer	87	64.4	163	44.2	250	49.6	<0.001
low	20	14.8	31	8.4	51	10.1	
middle	19	14.1	67	18.2	86	17.1	
high	9	6.7	108	29.3	117	23.2	
Self-rated health							
excellent/very good/good	84	62.2	191	51.8	275	54.6	<0.001
acceptable/poor	34	25.2	85	23.0	119	23.6	
non-testable	17	12.6	93	25.2	110	21.8	
Mini-Mental State Examination							
24-30	82	60.7	137	37.1	219	43.5	<0.001
18-23	24	17.8	75	20.3	99	19.6	
0-17	29	21.5	157	42.5	186	36.9	
Activities of Daily Living							
5	83	61.5	161	43.6	244	48.4	<0.001
4-1	44	32.6	158	42.8	202	40.1	
0	8	5.9	50	13.6	58	11.5	
Geriatric Depression Scale							
0-4	77	57.0	141	38.2	218	43.3	<0.001
5+	40	29.6	130	35.2	170	33.7	
non-testable	18	13.3	98	26.6	116	23.0	
Physical Component Summary							
$\geq$ average	75	55.6	130	35.2	205	40.7	<0.001
$<$ average	43	31.9	146	39.6	189	37.5	
non-testable	17	12.6	93	25.2	110	21.8	
Mental Component Summary							
$\geq$ average	66	48.9	136	36.9	202	40.1	0.005
$<$ average	52	38.5	140	37.9	192	38.1	
non-testable	17	12.6	93	25.2	110	21.8	
Chronic diseases (number)							
0	17	12.6	25	6.8	42	8.3	0.112
1	31	23.0	90	24.4	121	24.0	
2+	87	64.4	254	68.8	341	67.7	

*Male vs Female from Pearson χ^2 test or t-test as appropriate

Table 2: Model fit statistics for 2- to 6-class models

N. classes	2	3	4	5	6
Whole study population (n=504)					
AIC	5208.68	**4856.85**	4865.77	7179.39	7232.64
BIC	5369.14	**5118.65**	5228.91	7643.88	7798.47
Testable subsample (n=385)					
AIC	3794.22	**3649.22**	4103.42	4139.44	4063.46
BIC	3912.82	**3846.89**	4380.15	4495.23	4498.31

AIK: Akaike Information Criterion; BIC: Bayesian Information Criterion

health profiles too. The first class is characterized by high probability of being autonomous ($\lambda = 0.88$), not depressed ($\lambda = 0.81$), not cognitively impaired ($\lambda = 0.84$), self-reporting good health conditions ($\lambda = 0.91$), with PCS and MCS scores higher than or equal to the average (respectively $\lambda = 0.71$ and 0.66). This class has been labeled: "healthy group". It includes 209 individuals (54.3% of the testable subsample). The second class is characterized by high probability of being semi/not autonomous (respectively $\lambda = 0.47$ and 0.44), depressed ($\lambda = 0.81$), self-reporting bad health ($\lambda = 0.78$), with PCS and MCS scores lower than the average (respectively $\lambda = 0.93$ and 0.65). This group of 119 individuals (30.9% of the testable subsample) has been labeled: "unhealthy group". The third group is characterized by high probability of self-reporting good health conditions ($\lambda = 1$) and being semi-autonomous ($\lambda = 0.62$), mild/severe cognitive impairment (respectively $\lambda = 0.47$ and 0.48) with MCS score lower ($\lambda = 0.69$) but PCS score higher than or equal to the average ($\lambda = 0.84$). Posterior probabilities for depression are similar: $\lambda = 0.45$ not-depressed vs $\lambda = 0.55$ depressed. This class' label was: "semi-autonomous senile group". It included 57 nonagenarians (14.9% of the testable subsample). All the posterior probabilities are reported in Table 3.

The first class has been labeled "healthy group" in both analysis: posterior probabilities followed a similar pattern especially in terms of (good) health status items as shown by the circles black and white in Figure 1. The second class was named "unhealthy group" both when considering the whole study population and the subsample of testable individuals (see squares black and white in Figure 1). Despite the same label these two classes were different in terms of characteristics. When considering the whole study population, the second class was composed by almost all the non-testable nonagenarians: individuals in the worst health conditions. Excluding the non-testable nonagenarians, mainly individuals populating the third classes moved to the second resulting in an "unhealthy group" with less *extreme* health characteristics. The consequence of this exclusion was more evident for the last (third) class obtained in both analyses. When considering all the nonagenarians, we obtained the "unhealthy optimistic group": people mainly in bad health

conditions but not always declaring (SRH) their real health status. When excluding the non-testable nonagenarians, part of the individuals populating the third group obtained in the previous analysis moved to the second class found in the second analysis. As shown in Figure 1, the "unhealthy optimistic group" (first analysis) and the "unhealthy group" (second analysis) had similar posterior probabilities for the (good) health status indicators, especially in terms of functional and cognitive status. Within the second analysis, the 57 out of 385 nonagenarians composing the "semi-autonomous senile group" had higher probability of declaring good health condition and to obtain a high PCS score than the "healthy group" but they had poor cognitive health, sometimes depression and they were mainly semi-autonomous nonagenarians.

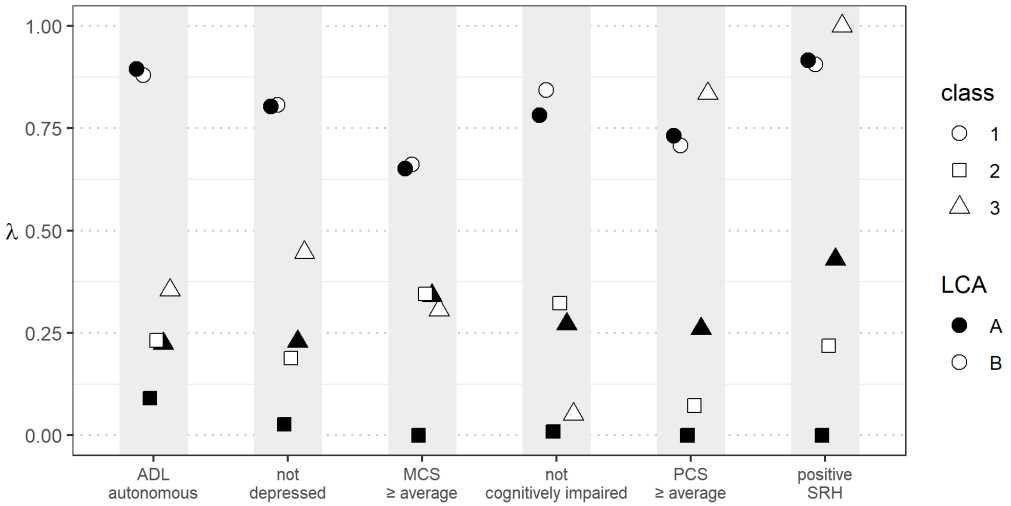
Table 3: Health status indicator probabilities (λ) per health status profile resulting from the two LCAs

Variable	Item	Whole study population (n=504)			Testable subsample (n=385)		
		Latent class			Latent class		
		1	2	3	1	2	3
n	(%)	217(43.1%)	110(21.8%)	177(35.1%)	209(54.3%)	119(30.9%)	57(14.8%)
Activities of Daily Living	autonomous	0.89	0.09	0.22	0.88	0.23	0.35
	semi-autonomous	0.11	0.47	0.72	0.12	0.70	0.62
	non-autonomous	0.00	0.44	0.06	0.00	0.07	0.03
Geriatric Depression Scale	non-depressed	0.80	0.03	0.23	0.81	0.19	0.45
	depressed	0.18	0.00	0.74	0.19	0.81	0.55
	non-testable	0.01	0.97	0.04			
Mental Component Summary	$\geq$average	0.65	0.00	0.34	0.66	0.35	0.31
	<average	0.35	0.00	0.66	0.34	0.65	0.69
	non-testable	0.00	1.00	0.00			
Mini-Mental State Exam.	24-30	0.78	0.01	0.27	0.84	0.32	0.05
	18-23	0.18	0.02	0.33	0.14	0.32	0.47
	0-17	0.04	0.97	0.40	0.02	0.35	0.48
Chronic Diseases (number)	0	0.11	0.13	0.02	0.10	0.02	0.03
	1	0.27	0.23	0.21	0.27	0.23	0.18
	2+	0.62	0.65	0.77	0.63	0.75	0.79
Physical Component Summary	$\geq$average	0.73	0.00	0.26	0.71	0.07	0.84
	<average	0.27	0.00	0.74	0.29	0.93	0.16
	non-testable	0.00	1.00	0.00			
Self-rated Health	excellent/very good/good	0.92	0.00	0.43	0.91	0.22	1.00
	acceptable/poor	0.08	0.00	0.57	0.09	0.78	0.00
	non-testable	0.00	1.00	0.00			

Empty items are due to the subsampling: non-testable individuals are not included in the second analysis

For both analysis 1: "healthy group"; 2; "Unhealthy group" and respectively 3: "Unhealthy optimistic group" and "Semi-autonomous senile group"

Figure 1: (Good) health status item probabilities (λ) per health status resulting from the two Latent Class Analysis (LCA)



Note1: Class 1: "healthy group"; 2: "Unhealthy group", for both first (A) and second (B) LCAs; Class 3 for LCA-A: "Unhealthy optimistic group"; for LCA-B: "Semi-autonomous senile group"; **Note2:** MCS: Mental Component Sumary; PCS: Physical Component Summary; Positive self-rated health: excellent/very good/good self-rated health

Results are controlled for age, gender, education and past-work (Table 4). In the analysis on the whole Mugello's nonagenarians, older individuals and entrepreneur or technical workers are more likely to be part of the "unhealthy group" instead of the "healthy group" (92-94 vs 90-91 Odds Ratio (OR) = 2.78; 95+ vs 90-91 OR = 7.60; high level worker vs farmer OR = 2.54) while being more educated reduces this odds (4-5 vs 3 years of education OR = 0.48; 5+ vs 3 OR = 0.06). Being older increases also the odds of being in "unhealthy optimistic group "instead of the "healthy group" (95+ vs 90-91 OR = 5.31) while both being male and middle level (qualified office) worker reduces it (male vs female OR = 0.40; middle level worker vs farmer OR = 0.37).
In the analysis on the subsample of testable individuals, being older increases the odds of being in the second "unhealthy group" instead of the "healthy group" (95+ vs 90-91 OR = 5.86) while both being male and middle level (qualified office) worker reduces it (male vs female OR = 0.45; middle level work vs farmer OR = 0.38). Finally, both being older and less educated reduces the odds of being in "semi-autonomous senile group "instead of the "healthy group" (95+ vs 90-91 OR = 8.13; 0-2 vs 3 years of education OR = 6.90).

Table 4: Odds ratios of demographic and socio-economic characteristics for the health profiles

LCA		Whole study population (n=504)					Testable subsample (n=385)				
Variable	Item	Coeff.	OR	Std. err.	t value	Pr(> |t|)	Coeff.	OR	Std. err.	t value	Pr(> |t|)
				2 vs 1					**2 vs 1**		
	(Intercept)	-0.93	0.39	0.34	-2.78	0.01	-0.34	0.71	0.34	-1.02	0.31
Age class	92-94	1.02	2.78	0.34	3.00	0.00	0.12	1.13	0.38	0.32	0.75
(ref. 90-91)	95+	2.03	7.60	0.37	5.45	0.00	1.77	5.86	0.38	4.65	0.00
Sex (ref. female)	male	-0.82	0.44	0.38	-2.19	0.03	-0.80	0.45	0.36	-2.21	0.03
Education	0-2	-0.65	0.52	0.46	-1.39	0.16	0.49	1.64	0.55	0.89	0.37
(ref. 3 years)	4-5	-0.74	0.48	0.32	-2.29	0.02	-0.31	0.73	0.35	-0.90	0.37
	> 5	-2.85	0.06	0.97	-2.93	0.00	-1.45	0.23	0.84	-1.72	0.09
Work	low level	0.08	1.08	0.50	0.16	0.88	0.12	1.12	0.46	0.26	0.80
(ref. farmer)	middle level	-0.28	0.75	0.42	-0.68	0.50	-0.96	0.38	0.46	-2.08	0.04
	high level	0.93	2.54	0.37	2.53	0.01	-0.24	0.79	0.45	-0.53	0.60
				3 vs 1					**3 vs 1**		
	(Intercept)	-0.15	0.86	0.31	-0.49	0.62	-2.02	0.13	0.64	-3.17	0.00
Age class	92-94	0.24	1.27	0.32	0.74	0.46	1.16	3.20	0.62	1.89	0.06
(ref. 90-91)	95+	1.67	5.31	0.34	4.92	0.00	2.10	8.13	0.65	3.24	0.00
Sex (ref. female)	male	-0.92	0.40	0.32	-2.87	0.00	-0.78	0.46	0.70	-1.12	0.27
Education	0-2	0.26	1.30	0.42	0.63	0.53	1.93	6.90	0.69	2.79	0.01
(ref. 3 years)	4-5	-0.25	0.78	0.32	-0.80	0.42	-0.58	0.56	0.60	-0.98	0.33
	> 5	-0.69	0.50	0.53	-1.30	0.20	0.21	1.24	0.82	0.26	0.80
Work	low level	0.06	1.06	0.42	0.15	0.88	-0.80	0.45	1.09	-0.73	0.47
(ref. farmer)	middle level	-0.99	0.37	0.41	-2.42	0.02	-1.38	0.25	0.97	-1.42	0.16
	high level	0.12	1.13	0.37	0.33	0.75	0.58	1.79	0.55	1.06	0.29

For both analysis 1: "healthy group"; 2; "Unhealthy group" and respectively 3: "Unhealthy optimistic group" and "Semi-autonomous senile group"

Discussion

For identifying health profiles among nonagenarians from Mugello (Tuscany - Italy), the latent class analysis (LCA) has been performed two times: first on the whole study population and then on the subsample of testable individuals - nonagenarians in the "extreme" (worst) conditions have been excluded from the analysis. Removing those individuals from the analysis allowed us to capture more heterogeneity of health among the remaining oldest-old.

Four different health profile have been found, within the two LCAs performed, labeled according to the posterior probabilities of finding certain health characteristics into them. A "healthy group" (a) that is consistent in both analysis when considering the whole study population and the subsample of testable individuals. An "unhealthy group" (b), which can be also called "non-testable group", resulted from the first analysis. This group confirms that non-testable individuals are a stand-alone group of people in extremely bad health conditions. The other two groups have been obtained by excluding the non-testable individuals from the analysis. The "unhealthy optimistic group", resulted from the analysis on the whole study population, splitted into a less extreme "unhealthy group" (c): nonagenarians in bad health conditions with respect to all the health indicators considered and a "semi-autonomous senile group" (d): individuals with good self-rated health and physical condition but bad cognitive status. Groups (a) and (c) are commonly found by other scholars interested in health profiles among elderly [6, 8, 27], while the last (d) one is similar to what some researchers identified on younger adults [6, 27].

Certain demographic and socio-economic characteristics were found to be associated with being part of some of the final groups. Older nonagenarians are more likely to be in worse health conditions. This result confirm literature data: at very old ages, the health status completely changes within one or two years [28, 29]. Male have a lower probability of being in worse general health conditions, confirming the so-called "gender paradox": men are healthier than women at older ages [30, 31]. More educated people are less likely to be in extremely bad health conditions and being low educated increases the probability of being cognitively impaired at older ages. Those results are similar to what have been found by Ng et al. [7] on younger elderly. Working experience is associated with health conditions, showing different results. In line with the existing literature, office workers have a lower probability of being in bad health conditions at older ages with respect to farmers [32, 33]. But people who had a high level job seem surprisingly more likely to be in the worst health conditions.

Considering health as a multidimensional concept by finding health profiles could help to understand the right care-needed according to the different health profile of each person [8, 34]. The "semi-autonomous senile group" of

nonagenarians, for instance, needs help mainly from a cognitive perspective while the "unhealthy group" needs a more complex care service with respect to the severity of the general health status.
This study has limitations which needs to be noted. It is based on a cross-sectional dataset: health characteristics has been collected only one time. For this reason, we were not allowed to study the causal relationship between socio-demographic characteristics and health status and profiles. Furthermore, many of the information about the health status are self reported and cutoff points - chosen according to the existing literature - did not equate to a clinical diagnosis, it would be useful to verify their veracity with objective measures. A longitudinal setting would help to solve some of those limitations and allow us to study mortality levels among different health classes and analyze the dynamic of the resulting health profiles among the elderly over time.

References

[1] Colin Mathers, Doris Ma Fat, J. T. Boerma, and World Health Organization, editors. The Global Burden of Disease: 2004 Update. World Health Organization, Geneva, Switzerland, 2008. OCLC: ocn264018380.

[2] Martin J Prince, Fan Wu, Yanfei Guo, Luis M Gutierrez Robledo, Martin O'Donnell, Richard Sullivan, and Salim Yusuf. The burden of disease in older people and implications for health policy and practice. The Lancet, 385(9967):549–562, February 2015.

[3] Other International Agreements of the United States of America 1776-1949 (Bevans) WHO. World Health Organization. page 20, 1948.

[4] World Population Prospects - Population Division - United Nations. https://esa.un.org/unpd/wpp/.

[5] M. Huber, J. A. Knottnerus, L. Green, H. v. d. Horst, A. R. Jadad, D. Kromhout, B. Leonard, K. Lorig, M. I. Loureiro, J. W. M. v. d. Meer, P. Schnabel, R. Smith, C. v. Weel, and H. Smid. How should we define health? BMJ, 343(jul26 2):d4163–d4163, July 2011.

[6] Louise Lafortune, François Béland, Howard Bergman, and Joël Ankri. Health status transitions in community-living elderly with complex care needs: A latent class approach. BMC Geriatrics, 9(1), December 2009.

[7] Charis W. L. Ng, Nan Luo, and Bee Hoon Heng. Health status profiles in community-dwelling elderly using self-reported health indicators: A latent class analysis. Quality of Life Research, 23(10):2889–2898, December 2014.

[8] Li-Fan Liu, Wei-Hua Tian, and Hui-Ping Yao. Utilization of health care services by elderly people with National Health Insurance in Taiwan: The heterogeneous health profile approach. Health Policy, 108(2-3):246–255, December 2012.

[9] Li-Fan Liu and Pei-Fang Su. What factors influence healthy aging? A person-centered approach among older adults in Taiwan. Geriatrics & Gerontology International, 17(5):697–707, May 2017.

[10] Paul Gellert, Petra von Berenberg, Thomas Zahn, Julia Neuwirth, Adelheid Kuhlmey, and Dagmar Dräger. Multimorbidity Profiles in German Centenarians: A Latent Class Analysis of Health Insurance Data. Journal of Aging and Health, page 089826431773789, October 2017.

[11] Mugello Study. http://www.mugellostudy.com/.

[12] Raffaello Molino-Lova, Francesco Sofi, Guido Pasquini, Annamaria Gori, Federica Vannetti, Rosanna Abbate, Gian Franco Gensini, and Claudio Macchi. The Mugello Study, a survey of nonagenarians living in Tuscany: Design, methods and participants' general characteristics. European Journal of Internal Medicine, 24(8):745–749, December 2013.

[13] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh. "Mini-mental state". Journal of Psychiatric Research, 12(3):189–198, November 1975.

[14] Marshal Folstein, James C. Anthony, Irma Parhad, Bonnie Duffy, and Ernest M. Gruenberg. The Meaning of Cognitive Impairment in the Elderly. Journal of the American Geriatrics Society, 33(4):228–235, April 1985.

[15] Sidney Katz. Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. JAMA, 185(12):914, September 1963.

[16] S. Katz. Assessing self-maintenance: Activities of daily living, mobility, and instrumental activities of daily living. J Am Geriatr Soc, 31(12):721–727, December 1983.

[17] Jerome A. Yesavage and Javaid I. Sheikh. 9/Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version. Clinical Gerontologist, 5(1-2):165–173, November 1986.

[18] Jeffrey M. Lyness, Tamson Kelly Noel, Christopher Cox, Deborah A. King, Yeates Conwell, and Eric D. Caine. Screening for Depression in Elderly Primary Care Patients: A Comparison of the Center for Epidemiologic Studies—Depression Scale and the Geriatric Depression Scale. Archives of Internal Medicine, 157(4):449–454, February 1997.

[19] Giovanni Apolone, Paola Mosconi, Luciana Quattrociocchi, Emilio A. L. Gianicolo, Nicola Groth, and Josh E. Jr Ware. Questionario sullo stato di salute SF-12. Versione italiana. Milano, Istituto di Ricerche Farmacologiche Mario Negri, 2005.

[20] ISTAT - NOMENCLATURA E CLASSIFICAZIONE DELLE UNITÀ PROFESSIONALI. http://professioni.istat.it/sistemainformativoprofessioni/cp2011/.

[21] C. Mitchell Dayton and George B. Macready. Concomitant-Variable Latent-Class Models. Journal of the American Statistical Association, 83(401):173–178, 1988.

[22] Jacques A. Hagenaars and Allan L. McCutcheon, editors. Applied Latent Class Analysis. Cambridge University Press, Cambridge ; New York, 2002.

[23] Annabel Bolck, Marcel Croon, and Jacques Hagenaars. Estimating Latent Structure Models with Categorical Variables: One-Step Versus Three-Step Estimators. Political Analysis, 12(01):3–27, 2004.

[24] R Core Team. R: The R Project for Statistical Computing. https://www.r-project.org/, 2017.

[25] Alexander Kowarik and Matthias Templ. Imputation with the R Package **VIM**. Journal of Statistical Software, 74(7), 2016.

[26] Drew A. Linzer and Jeffrey B. Lewis. **poLCA** : An R Package for Polytomous Variable Latent Class Analysis. Journal of Statistical Software, 42(10), 2011.

[27] Andrea R Zammit, John M Starr, Wendy Johnson, and Ian J Deary. Profiles of physical, emotional and psychosocial wellbeing in the Lothian birth cohort 1936. BMC Geriatr, 12:64, October 2012.

[28] S. E. Hardy, J. A. Dubin, T. R. Holford, and T. M. Gill. Transitions between States of Disability and Independence among Older Persons. American Journal of Epidemiology, 161(6):575–584, March 2005.

[29] Ardo van den Hout and Fiona E. Matthews. Multi-state analysis of cognitive ability data: A piecewise-constant model and a Weibull model. Statistics in Medicine, 27(26):5440–5455, November 2008.

[30] Anna Oksuzyan, Knud Juel, James W. Vaupel, and Kaare Christensen. Men: Good health and high mortality. Sex differences in health and aging. Aging Clinical and Experimental Research, 20(2):91–102, April 2008.

[31] Luca Padua, Patrizio Pasqualetti, Daniele Coraci, Isabella Imbimbo, Alessandro Giordani, Claudia Loreti, Camillo Marra, Raffaello Molino-Lova, Guido Pasquini, Ilaria Simonelli, Federica Vannetti, and Claudio Macchi. Gender effect on well-being of the oldest old: A survey of nonagenarians living in Tuscany: The Mugello study. Neurological Sciences, 39(3):509–517, March 2018.

[32] K. Wickrama, J. A. Mancini, K. Kwag, and J. Kwon. Heterogeneity in Multidimensional Health Trajectories of Late Old Years and Socioeconomic Stratification: A Latent Trajectory Class Analysis. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68(2):290–297, March 2013.

[33] Sanna Read, Emily Grundy, and Else Foverskov. Socio-economic position and subjective health and well-being among older people in Europe: A systematic narrative review. Aging & Mental Health, 20(5):529–542, May 2016.

[34] S Jay Olshansky and Bruce A Carnes. Ageing and health. The Lancet, 375(9708):25, 2010.

Table A1: Marginal distribution pre and post missing values imputation of characteristics of the study population. Absolute values, percentages and differences

Characteristics	Pre-		Post-		Difference (%)
	n	%	n	%	
Education (years)					
0-2	65	13,5	65	12,9	-0,6
3	166	34,5	182	36,1	1,6
4-5	198	41,2	205	40,7	-0,5
6+	52	10,8	52	10,3	-0,5
total	481	100	504	100	
Work (level)					
farmer	245	49,4	250	49,6	0,2
low	51	10,3	51	10,1	-0,2
middle	86	17,3	86	17,1	-0,3
high	114	23,0	117	23,2	0,2
total	496	100	504	100	
Mini-Mental State Examination					
24 30	213	43,8	219	43,5	-0,4
18 23	95	19,5	99	19,6	0,1
0 17	178	36,6	186	36,9	0,3
total	486	100	504	100	
Activities of Daily Living					
5	235	47,6	244	48,4	0,8
4-1	201	40,7	202	40,1	-0,6
0	58	11,7	58	11,5	-0,2
total	494	100	504	100	
Physical Component Summary					
$\geq$ average	187	39,1	205	40,7	1,6
<average	181	37,9	189	37,5	-0,4
non-testable	110	23,0	110	21,8	-1,2
total	478	100	504	100	
Mental Component Summary					
$\geq$ average	182	38,1	202	40,1	2,0
<average	186	38,9	192	38,1	-0,8
non-testable	110	23,0	110	21,8	-1,2
total	478	100	504	100	
Chronic diseases (number)					
0	42	9,0	42	8,3	-0,7
1	111	23,9	121	24,0	0,1
2+	312	67,1	341	67,7	0,6
total	465	100	504	100	