

RESEARCH

Open Access



Feasibility properties of the EQ-5D-3L and 5L in the general population: evidence from the GP Patient Survey on the impact of age

Ole Marten* and Wolfgang Greiner

Abstract

Background: There is evidence to suggest that the proportion of missing values is slightly higher in the older population resulting in lower completion rates of the EQ-5D. However, existing studies rarely provide a within-sample comparison of feasibility properties across age groups to quantify this difference. Hence, this study examines feasibility properties of the EQ-5D-3L and 5L in the general population and explores the impact of age on the completion of EQ-5D instruments.

Methods: We pool five waves from the English GP Patient Survey, where respondents self-report their health in either EQ-5D-3L or 5L. Descriptive analysis was undertaken to analyse the distribution and proportion of missing values and completion rates stratified by age and EQ-5D version; logistic regression models were specified to quantify the impact of age, gender and potential long-term conditions on the completion of each of the EQ-5D instruments.

Results: The total sample comprises ~4.36 million observations, of which 2.88 million respondents report their health in 5L and 1.47 million in 3L, respectively. Respondents over 64 years have slightly more missing values in each dimension than younger respondents. The highest share was observed for the oldest age group in the dimension anxiety/depression (3L 9.1% vs. 5L 7.6%), but was otherwise below 5%. Consequently, completion rates (observed and predicted) decreased with older age and at a higher rate after the age of 64; this was more pronounced for the 3L.

Conclusion: Evidence from our study suggests that both the EQ-5D-3L and 5L have good feasibility properties. In comparison to younger populations there appears to be a higher proportion of respondents with incomplete responses beyond the age of 64 years. Overall, the 5L version compares more favourably in terms of missing values, completion rates as well as with regard to the expected probability of an incomplete descriptive system.

Keywords: EQ-5D-3L, EQ-5D-5L, Health-related quality of life, Feasibility, Older population, GPPS

* Correspondence: ole.marten@uni-bielefeld.de

School of Public Health, Department of Health Economics and Health Care Management, Bielefeld University, Universitaetsstrasse 25, Bielefeld, Germany



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Over the past decades the consideration of the patient perspective has gained enormous importance for health care and policy decision making [1–3]. With regard to this, the measurement of self-reported health-related quality of life (HRQoL) is a major component on this pathway. The EQ-5D is a well-established and widely used instrument developed by the EuroQol Group specifically intended for that purpose [4]. Over the years the EQ-5D has developed a significant role in economic decision making, since prominent health technology assessment bodies in the United Kingdom, the Netherlands, Spain or France recommend or clearly specify that HRQoL should be measured using the EQ-5D [5, 6]. Since the release of the EQ-5D-5L – with five response levels – [7], extensive research has been conducted to compare the measurement properties of the EQ-5D-3L and EQ-5D-5L (hereafter 3L and 5L, respectively). A recent literature review confirms that both the 3L and the 5L are applicable to a wide range of populations, while confirming improved informativity, less ceiling effects and better distributional properties for the 5L [8]. Further studies examined the measures' feasibility, which was commonly operationalised in terms of missing values and completion rates at the individual level. The former is either defined as unit nonresponse or as item nonresponse, where information is unavailable for the respondent as a whole or just on individual items [9]. Whereas the latter construct is defined as the share of computable EQ-5D index values, which requires complete information on all five items of the descriptive system [10]. Studies by Janssen et al. [11] and Agborsangaya et al. [12] reported very good feasibility of both 3L and 5L with missing values of less than 2%, whereas Buchholz et al. [8] conclude on a proportion of missing values of less than 5% from reviewing 15 studies.

However, there is evidence to suggest that the proportion of missing values is slightly higher in the older population resulting in lower completion rates of the instrument [10]. Even though studies conducted with older respondents [13–17] report missing values well within the margin reported by Buchholz et al. [8], there are other studies reporting proportions of up to 10% [18–21]. On the contrary, the samples in Janssen et al. [11], Agborsangaya et al. [12] and other studies [22–24] were younger than 65 years and reported considerable fewer missing values. Hence, we suspect that the share of missing responses and incomplete descriptive systems may be driven by an age-dependent effect, as this was also described for the SF-36 [25–27]. Terwee et al. [28] argue that missing values may be indicative of problems with item interpretability, which is confirmed by findings from Hulme et al. [19] and van Leeuwen et al. [29] who report this kind of response issues on the 3L for older people.

Hence, this study aims to assess the magnitude of missing values and incomplete responses for both the 3L and 5L using five waves of the large-scale General Practitioner Patient Survey (GPPS) based on age-stratified comparisons of these feasibility parameters in the English general population, which allows a within-sample assessment of differences between the older and the younger general public.

Methods

Data

For the analysis we utilise individual-level self-reported EQ-5D data from the GPPS [30]. This is a large-scale cross-sectional survey undertaken on behalf of the National Health Service (NHS) England. Since 2007 the survey is sent yearly to more than 2 million adults asking them about their experience with their general practitioner and other NHS services. The questionnaire is primarily posted to participants; however, the survey can also be completed online or by telephone. Moreover, it is available in a variety of languages. The overall GPPS samples are obtained by drawing proportionately stratified samples from each practice using registration data held by the NHS Digital database. Individuals are eligible for inclusion in the survey, if they are 18 years or above, hold a valid NHS registration number and were continuously registered with a general practitioner (GP) for at least 6 months [31]. Even though respondents are recruited via GP registries, we would like to argue that the underlying sample is in effect recruited from the general population, since registration with a GP does not necessarily imply that respondents are currently under treatment for a condition or an illness.

Variables

The primary aim of the GPPS survey is to assess patients' experiences with their GP and other local NHS services; questions include aspects such as access to services, appointments, waiting times and how people manage their health. In addition to that, respondents answer the EQ-5D, which has been used in the survey between 2011 and 2017 [31]. The EQ-5D is a standardised generic measure of HRQoL developed by the EuroQol Group. The EQ-5D descriptive system entails five dimensions: mobility, self-care, usual activities, pain or discomfort and anxiety or depression. The initially developed EQ-5D-3L has three response levels, allowing respondents to describe their health status based on three options: no problems (level 1); some or moderate problems (level 2); or extreme problems/unable to (level 3) [32]. The EQ-5D-5L is a re-developed version covering the same five dimensions, but expanding the available response options to five levels, again, ranging from no problems (level 1), over slight, moderate and severe

problems to extreme problems/unable to (level 5) [7]. The response from each dimension-level can be concatenated to form a health profile, which can be linked to a value set – a scoring algorithm with preference-based weights for each dimension-level – to generate a single summary index score [33]. The second component of the EQ-5D – the visual analogue scale (EQ VAS) – is not included in the GPPS survey [34].

In this study, we pool data from several years. We use data from 2012, as this is the last year the 3L was used, also capturing the effect of changing the instrument to the 5L between the first and the second wave of 2012. Further, we examine data from 2016 and 2017, since this was the most recent data when we applied for the data set. The survey mode was equivalent across all 3 years in the sense that the vast majority of respondents answered the questionnaire paper-based and only 4–6% of the respondents answered using the online survey, while telephone responses were negligible [30, 35–38]. Since the publicly available analysis tool does not allow in depth examination of all EQ-5D data, we submitted an application for individual-level data to NHS England. Further, we were granted access to reported background information, which is based on gender, age groups and existence of any of the following long-term conditions: Alzheimer's disease/ dementia, angina/heart problem, arthritis/joint problem, asthma/chest problem, blindness, cancer, deafness, diabetes, epilepsy, high blood pressure, kidney or liver disease, long-term back problem, long-term mental health problem or long-term neurological problem [31]. The information on the administration mode was not included in the individual-level data set and, hence, could not be controlled for.

Analysis

We examine feasibility of the EQ-5D in older persons in comparison to the general population by investigating distributional properties of EQ-5D data as well as the prevalence and distribution of missing values, which ultimately prevent the calculation of an EQ-5D index value. We do so by conducting descriptive analysis based on the proportion of respondents per level in both 3L and 5L for the whole sample as well as stratified by age groups. We expect to observe a lower proportion of level 1 responses (i.e. at the ceiling) on the 5L in general and more pronounced in respondents aged 65 and above.

As suggested by Janssen et al. [11], we examine feasibility for both 3L and 5L in terms of missing values separately for each dimension and stratified for age groups. We further report completion rates based on the same criteria. We analyse the proportion of missing values by age groups using chi-square tests to examine potential associations with age. Given the large-scale of this exercise, we report standardised effect sizes based on

Cramer's V to quantify the magnitude of observed differences [39]. We further explored the impact of age, gender and having a long-term health condition on the probability of returning an incomplete EQ-5D using logistic regression analysis. We used "incomplete response" as a binary dependent variable where 1 indicates that at least one EQ-5D item was not answered and, thus, we were unable to calculate an index value. We used 'female' and 'condition' as binary independent variables, where 1 represents being female or having a long-term condition, respectively. Further, age group was added as a categorical variable into the model with 18–24 years as the reference category. We used STATA's *margins* post-estimation command to calculate predicted probabilities of returning an incomplete EQ-5D for each age group holding the other variables at their sample means. We apply the conventional significance level of 5%. All analysis was conducted using STATA 16 [40].

Results

Sample description

After pooling five different waves of the GPPS the total sample comprised 4,358,700 observations. Of those, 1,476,395 contributed to the 3L sample, whereas 2,882,305 respondents were represented in the 5L sample. As Table 1 suggests, the sample characteristics were similar across the 3L and 5L sample including slightly more women. About one third of the sample was 65 years and above (3L: 33.8%; 5L: 36.7%) and about 60% reported at least one long-standing health condition. The most prevalent long-term condition was high blood pressure (23%) followed by arthritis or joint problems (16%). Mental health problems including Alzheimer's disease/dementia and neurological problems were reported by 15.9 and 15.5% for the 3L and 5L sample, respectively.

Comparison of response distribution

Tables 2 and 3 provide an overview of the response distribution for each dimension stratified by age groups for both the 3L and 5L, respectively. Unsurprisingly, problems were always least prevalent in the youngest age groups with a monotonically increasing trend with increasing age. Problems were more frequently reported when using the 5L and limitations were spread wider across the severity range. Generally, self-care appears to be the least affected dimension with a considerable ceiling effect. Even in the highest age group only 40% report any problems with self-care, whereas 82% report problems in mobility and pain or discomfort in that age group. Interestingly, problems with pain or discomfort and anxiety or depression were the most frequent in younger age groups (around 30% vs. self-care 5% vs. mobility 9%). While limitations in pain or discomfort

Table 1 Sample characteristics of five waves of GP Patient survey data

Characteristics	EQ-5D-3L sample		EQ-5D-5L sample		Total	
	N	%	N	%	N	%
Gender						
Male	636,076	43.1	1,255,846	43.6	1,891,922	43.4
Female	840,324	56.9	1,626,459	56.4	2,466,783	56.6
Age group						
18–24 years	65,729	4.5	116,338	4.0	182,067	4.2
25–34 years	148,966	10.1	267,430	9.3	416,396	9.6
35–44 years	203,616	13.8	363,407	12.6	567,023	13.0
45–54 years	259,698	17.6	501,875	17.4	761,573	17.5
55–64 years	300,329	20.3	577,672	20.0	878,001	20.1
65–74 years	273,922	18.6	595,753	20.7	869,675	20.0
75–84 years	169,686	11.5	347,876	12.1	517,562	11.9
85 or over	54,454	3.7	111,954	3.9	166,408	3.8
Long-standing health condition						
Yes	895,175	60.6	1,771,617	61.5	2,666,792	61.2
No	535,210	36.3	1,022,939	35.5	1,558,149	35.8
Don't know/ can't say	27,722	1.9	58,320	2.0	86,042	2.0
Missing	18,293	1.2	29,429	1.0	47,722	1.1
Long-term condition						
Alzheimer/ dementia	9929	0.7	21,101	0.7	31,030	0.7
Angina/ heart problems	96,707	6.6	179,424	6.2	276,131	6.3
Arthritis/ joint problems	240,480	16.3	466,970	16.2	707,450	16.2
Asthma/ chest problems	147,325	10.0	297,099	10.3	444,424	10.2
Blindness/ visual problems	18,500	1.3	32,530	1.1	51,030	1.2
Cancer in the last 5 yrs	54,362	3.7	118,398	4.1	172,760	4.0
Deaf/ hearing problems	70,311	4.8	139,745	4.9	210,056	4.8
Diabetes	123,405	8.4	266,795	9.3	390,200	9.0
Epilepsy	15,531	1.1	28,152	1.0	43,683	1.0
High blood pressure	337,554	22.9	668,507	23.2	1,006,061	23.1
Kidney or liver problems	25,649	1.7	56,508	2.0	82,157	1.9
Long-term back problems	157,361	10.7	306,127	10.6	463,488	10.6
Long-term mental health problems	51,121	3.5	119,783	4.2	170,904	3.9
Long-term neurological problems	26,744	1.8	60,278	2.1	87,022	2.0
Long-term other health problems	172,282	11.7	366,061	12.7	538,343	12.4

increase considerably with age, the proportion of any reported problems in anxiety or depression remains fairly stable; this pattern is constant across both EQ-5D versions. Overall, floor effects, where respondents respond with the worst answer category, are not observable in this general population sample. Severe and extreme problems are least prevalent in the dimensions self-care and anxiety or depression. However, while severe and extreme problems with self-care increase with age the opposite seems to be the case for anxiety or depression. Again, this pattern is consistent across both the 3L and

5L, with the exception being level 3 in mobility in the 3L ('confined to bed'), which was the least frequent overall.

Feasibility of the EQ-5D-3L and 5L

Table 4 summarises the share of missing values and completion rates by age groups based on the 3L and 5L. Overall, the proportion of missing values in any of the EQ-5D dimensions was very low but increasing with age. Chi-square tests suggest that the proportion of missing values in any dimension are not independent of the respondents' age ($p < 0.001$). Given the large sample size,

Table 2 Distribution of EQ-5D-5L responses by dimension and age group

Parameter	Level	Total N	Age group Proportion in %							
			18–24	25–34	35–44	45–54	55–64	65–74	75–84	85+
Mobility	No problems	1,958,750	91.2	90.2	86.2	78.5	68.6	58.8	39.6	18.0
	Slight problems	403,522	4.6	5.0	7.0	10.5	14.7	18.9	23.9	23.0
	Moderate problems	268,905	1.6	1.8	2.9	5.3	8.5	12.2	20.3	29.2
	Severe problems	170,235	0.6	0.8	1.6	3.4	5.7	7.3	12.1	22.0
	Unable to	22,630	0.4	0.4	0.4	0.5	0.6	0.7	1.3	4.1
Self-care	No problems	2,484,621	94.9	94.7	92.9	89.6	86.2	84.3	76.6	60.2
	Slight problems	149,380	1.7	1.8	2.5	3.7	5.0	6.1	9.5	15.0
	Moderate problems	118,583	1.0	1.0	1.6	3.0	4.4	5.0	7.2	12.5
	Severe problems	41,437	0.4	0.4	0.7	1.2	1.7	1.6	2.0	4.7
	Unable to	20,059	0.4	0.4	0.4	0.4	0.5	0.5	1.2	4.3
Usual activities	No problems	1,943,526	85.3	85.4	81.8	75.0	67.7	61.4	45.3	24.6
	Slight problems	446,180	7.8	7.9	9.5	12.6	15.9	19.3	24.3	24.3
	Moderate problems	259,886	3.3	3.0	4.0	6.1	8.6	10.9	17.2	25.0
	Severe problems	114,562	1.2	1.3	1.9	3.1	4.4	4.5	6.4	11.5
	Unable to	55,120	0.6	0.6	0.8	1.1	1.5	1.7	3.9	10.9
Pain/Discomfort	No	1,272,899	72.4	68.6	60.3	49.3	40.3	33.6	24.6	18.0
	Slight	874,407	18.2	20.7	24.9	29.7	33.2	33.6	34.6	31.2
	Moderate	460,993	5.6	6.4	8.7	12.4	16.1	20.0	26.7	33.8
	Severe	171,788	1.5	2.0	3.1	4.9	6.8	7.3	9.5	11.3
	Extreme	39,224	0.4	0.6	1.0	1.6	1.8	1.4	1.5	2.0
Anxiety/Depression	No	1,879,193	67.0	68.6	67.5	64.5	64.5	67.3	62.0	53.6
	Slight	556,159	17.3	17.7	18.4	19.4	19.4	19.0	20.7	24.1
	Moderate	252,616	9.0	8.0	8.0	9.2	9.4	7.8	8.9	12.4
	Severe	61,933	3.1	2.4	2.4	2.9	2.6	1.3	1.2	1.6
	Extreme	30,483	1.7	1.3	1.4	1.6	1.2	0.5	0.5	0.7

this test result is not surprising and mitigated by the negligible association (Cramér's V). Nonetheless, it appears as if there is a steeper increase in the last two age categories. The highest proportions were found in anxiety or depression for respondents 65 years and above, where the proportion of missing responses peaks at 7.6% (5L) and 9.1% (3L) for the oldest respondents. Apart from this, the proportion of missing responses is less than 5% across all dimensions, and generally lower for the 5L in comparison to the 3L. Missing value patterns stratified for age groups can be found in the Appendix (see Appendix Tables A1 and A2). Among those respondents with missing values, patterns with just one missing item account for ~60–70% depending on age and EQ-5D version. Moreover, patterns with two to four missing responses only accumulate between 22 up to 30% of respondents with missings. Interestingly, the proportion of complete non-response to both 3L and 5L is highest among young adults (18–24 years - 5L: 19.8%; 3L:

18.4%) and drastically decreases with higher age (85 years and over - 5L: 2.8%; 3L: 2.6%).

Further, we report completion rates based on age groups, i.e. proportion of respondents with all five items completed (see also Table 4). Across all age groups the 5L completion rate was found to be higher in comparison to the 3L and for both completion was negatively associated with age, however, this effect was very weak. While on average only 4% of the 5L utilities cannot be calculated in the youngest age group, this figure increases to more than 14% in those 85 years and above. Correspondingly, these figures range from 5% (18–24 years) to 16.8% (85+ years) for the 3L.

Figure 1 presents predicted probabilities for returning an incomplete 3L or 5L for each age group controlling for gender and the presence of any reported long-term condition. Firstly, the probability of an incomplete EQ-5D response was lowest in the age group 25–34 years for the 3L (5.3%) and for the 5L

Table 3 Distribution of EQ-5D-3L responses by dimension and age group

Parameter	Level	Total N	Age group Proportion in %							
			18–24	25–34	35–44	45–54	55–64	65–74	75–84	85+
Mobility	No problems	1,074,533	92.7	92.0	88.2	82.0	73.2	63.6	45.5	23.1
	Some problems	362,930	4.9	5.7	9.3	15.6	24.5	33.6	51.0	72.5
	Confined to bed	4545	0.3	0.3	0.2	0.2	0.2	0.2	0.5	1.7
Self-care	No problems	1,307,043	95.6	95.5	93.8	91.4	88.8	86.2	79.9	64.7
	Some problems	118,593	2.0	2.2	3.7	6.0	8.4	10.2	14.3	25.0
	Unable to	11,759	0.4	0.4	0.4	0.4	0.5	0.7	1.6	5.5
Usual activities	No problems	1,061,308	88.1	87.9	83.6	77.8	71.6	66.3	51.3	30.4
	Some problems	332,279	8.9	9.2	12.8	18.0	23.7	27.9	38.5	49.6
	Unable to	47,921	0.9	1.0	1.5	2.2	2.8	3.4	6.5	15.6
Pain/Discomfort	No	780,144	73.1	76.3	68.8	58.5	49.2	40.8	29.9	22.6
	Moderate	571,456	17.5	19.8	25.6	33.7	41.6	49.0	57.5	63.1
	Extreme	84,005	1.3	1.7	3.2	5.4	7.0	7.2	8.4	9.6
Anxiety/Depression	Not	1,057,588	78.6	77.7	74.4	70.7	70.8	71.9	66.8	59.3
	Moderately	314,019	16.1	17.2	19.4	22.3	22.5	20.9	23.4	29.5
	Extremely	38,506	2.7	2.6	3.2	3.7	3.0	1.5	1.4	2.1

(4.3%) in those 18–24 years, respectively. Up until the age of 64, the probability only marginally increased by 1.2 percentage points for both the 3L and 5L. However, beyond the age of 64 years the probability of an incomplete EQ-5D response accelerated quickly peaking at 13.9% for the 3L and 11.5% for the 5L in those being 85 years and above. Secondly, the probability of an incomplete response was found

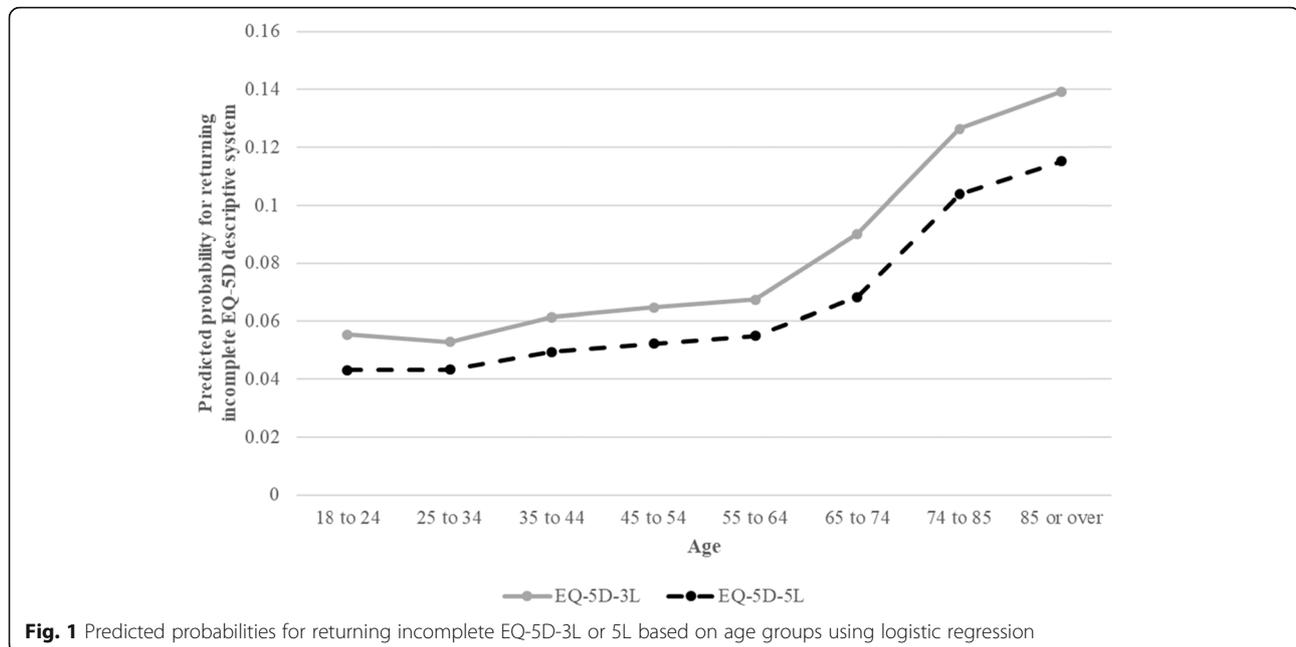
to be lower at any given age for the 5L in comparison to the 3L. The difference was between 1.0 (25–34 years) and 2.4 (85 or over) percentage points (see Fig. 1) with the spread being wider after the age of 64 years.

Discussion

The aim of this study was to assess feasibility of the 3L and 5L for older respondents in direct comparison to

Table 4 Proportion of missing values and overall EQ-5D completion rate stratified by age and EQ-5D version

Parameter	Missing N (%)	Age group Proportion in %								Cramér's V
		18–24	25–34	35–44	45–54	55–64	65–74	75–84	85+	
EQ-5D-5L										
Mobility	58,263 (2.02)	1.61	1.68	1.89	1.88	1.89	2.04	2.71	2.72	0.0229
Self-care	68,225 (2.37)	1.66	1.72	1.98	2.07	2.22	2.57	3.52	3.37	0.0363
Usual activities	63,034 (2.19)	1.85	1.88	2.06	2.00	1.94	2.11	2.98	3.76	0.0310
Pain/Discomfort	62,994 (2.19)	1.82	1.82	2.05	2.03	1.93	2.08	3.05	3.78	0.0328
Anxiety/Depression	101,921 (3.54)	1.98	2.01	2.37	2.46	2.89	4.12	6.78	7.56	0.0897
Completion rate	93.11%	96.01	95.89	95.18	94.68	94.06	92.24	87.78	85.92	0.1103
EQ-5D-3L										
Mobility	34,391 (2.33)	2.10	2.05	2.24	2.18	2.07	2.53	3.00	2.77	0.0205
Self-care	39,003 (2.64)	1.98	1.97	2.21	2.20	2.27	2.89	4.22	4.86	0.0496
Usual activities	34,890 (2.36)	2.06	2.03	2.15	2.01	1.87	2.43	3.68	4.42	0.0441
Pain/Discomfort	40,795 (2.76)	2.19	2.18	2.48	2.42	2.29	2.98	4.26	4.61	0.0439
Anxiety/Depression	66,287 (4.49)	2.66	2.54	2.99	3.28	3.79	5.69	8.36	9.14	0.0963
Completion rate	91.57%	94.88	94.91	93.98	93.39	92.70	89.83	85.20	83.14	0.1199



younger adults in the general population. The 5L showed better feasibility than the 3L across all age groups. The superiority of the 5L was more noticeable in older age groups, which was indicated by fewer missing values, higher completion rates and an overall lower likelihood of an incomplete response to the descriptive system.

With respect to the descriptive system we observe an improved response distribution when measured with the 5L as compared to the 3L version. As one might expect, the proportion of respondents at the ceiling is lower on the 5L and further decreasing with increasing age. This finding is consistent with other studies comparing the 3L and 5L version [8, 11]. Similarly, the response distribution in individual dimensions in the older population was also similar to that found in earlier studies, where self-care is the least informative dimension, whereas pain or discomfort is the most informative. Again, the response distribution in anxiety or depression does not seem to be moderated by age in both 3L and 5L [41–43].

The cross tabulation of incomplete responses per dimension with age categories clearly shows an increasing trend in higher age groups (Table 4). Generally, our findings are in line with proportions of missing values reported earlier [11, 12]. Especially in younger age groups we rarely observe more than 2% missing values per dimension, whereas the proportions are only marginally higher in older age and fall well within the overall margin reported in Buchholz et al. [8]. As an exception to this rule, we observed a higher proportion of missing values in the dimension of anxiety or depression, which go as high as 7.6% (5L) and 9.1% (3L) in those being 85 years and over. Holland and colleagues [44] report that

this dimension caused some embarrassment in older respondents. This may partially explain the higher prevalence of missing values in that particular dimension. Moreover, we found that in each dimension and for any given age group the proportion of missing responses was lower for the 5L – even though the difference was less than half a percent. This findings are also in line with earlier studies [8, 17, 45].

While the share of missing response was relatively low at the dimension-level, the completion rate, i.e. the ability to generate the utility value from the respondent's reported health state, was just above 90% on the total sample for both 3L and 5L. A potential explanation for this observation may be that missing values resulted from relatively many individuals with just one missing item rather than from respondents with multiple missing responses, which was suggested by the analysis of missing value patterns (see Appendix Tables A1 and A2). Moreover, completion of the 3L and 5L decreased by approximately 10 percentage points from the youngest to the oldest age group and more rapidly after the age of 64 years. For the total sample, we find that our observed completion rates for the 3L and 5L are about 5 percentage points lower than those reported in other studies in the general population [12, 46–49]. Considering the older population, 3L completion rates were 5–10 percentage points lower than those reported in a study from Switzerland, which also reports age group-specific completion [50]. However, the differences in the data collection process may largely account for this variance, since Luthy et al. [50] used computer-assisted personal interviews to collect data instead of self-reports as was the

case in this sample. Overall, literature on the feasibility properties of the 3L and 5L in the general public is scarce, where completion rates are predominantly reported for the overall sample. We are unaware of other studies providing age-specific completion rates, which limits further comparison with our findings.

We further provide predicted probabilities for an incomplete response based on a logistic regression model controlling for age groups, gender and presence of a long-term condition. The predicted probabilities largely follow the pattern from the uncontrolled cross-tabulations (Table 4) confirming the hypothesis of an age-dependent impact on the EQ-5D's completion with a more pronounced effect beyond the age of 64 years. Importantly, the 5L performs better than the 3L in the sense that the predicted probability of returning an incomplete descriptive system is consistently lower for the 5L. Evidence from the literature suggests that the length of the response scale has an effect on the data quality [51]. The question-answer process may be distorted, if the intended response does not match the available response options, which may cause the respondent to refuse to answer [29, 52]. Hence, it may be assumed that the lack of sensitivity in the 3L is in part responsible for the higher prevalence of missing values in the descriptive system, since respondents may lack the ability to report an appropriate level of problems on the three-level scale. Therefore, the improved sensitivity of the 5L [53, 54] may lead to improved feasibility as well, which is supported by our findings of better feasibility of the 5L in terms of reduced missing values, higher completion rates and lower probability of reporting an incomplete descriptive system. This notion is further underpinned by findings from Janssen and colleagues [55], where respondents argued that the 5L was easier to use and better reflects their response in comparison to the 3L.

Depending on the purpose of future studies, the slightly higher propensity for missing responses among aged respondents may have different implications. If the aim is to collect HRQoL data in a limited sample and to calculate quality-adjusted life years, researchers could consider assisted or interviewer-based approaches to mitigate the risk of bias due to incomplete response from older respondents [13, 56, 57]. At the same time, applying interviewer-based approaches may come at the cost of introducing other types of biases such as interviewer effects, socially desirable answers or a reduced willingness to disclose sensitive information, which may trade-off the gains of increased completeness [58]. However, on an aggregate level, such as in a population health survey like the GPPS, the extent of missing values can be rated as good or negligible [8, 11]. Nevertheless, our results suggest that missing values vary systematically by age groups and were more prevalent in older

adults. This ultimately implies a bias in estimated utility values against older respondents, which needs to be addressed adequately in statistical analyses.

A strength of our study is the huge sample size, which we gained by pooling data from several years of a consistent population health survey. In addition to that, we are able to compare responses to the descriptive system of both versions of the EQ-5D and across all age groups, however, it was not possible to compare the 3L and 5L on a like-for-like comparison, since respondents did not complete both measures. Due to the origin of the data, we had no information on how independently respondents answered the EQ-5D, i.e. whether respondents may have received help filling in the questionnaire and, hence, the level of feasibility problems for a self-report survey may be underestimated. A major limitation of our study is the missing EQ VAS component, which was not included in the survey and, hence, we were unable to investigate its feasibility properties. An in-depth analysis of the EQ VAS' feasibility properties in the general population seems desirable, since it is known to present problems to older adults [10]. Similarly, a qualitative study may facilitate a better understanding of the differences in feasibility properties between the 3L and 5L, which would also be welcomed for the EQ VAS. Future research should further explore the impact of different administration modes, i.e. paper-based vs. online completion, as we were not able to control for this factor even though the sample size would have been sufficient. Additionally, the GPPS data may allow an in-depth exploration of the impact of different long-term conditions on completion of both EQ-5D versions.

Conclusion

Evidence from our study suggests that both the 3L and 5L have good feasibility properties. The proportion of missing values is acceptable and low across all age groups. However, in comparison to younger populations there appears to be a higher proportion of respondents with incomplete responses, thus resulting in lower completion rates. Predicted probabilities for an incomplete response significantly increased beyond the age of 64 years for both versions of the EQ-5D, indicating a higher likelihood of missing values. Generally, we conclude that either version of the EQ-5D is applicable and feasible in the older population. However, the 5L version compares more favourably in terms of missing values, completion rates as well as with regard to the expected probability of an incomplete descriptive system.

Abbreviations

3L: EQ-5D-3L; 5L: EQ-5D-5L; EQ VAS: EQ-5D visual analogue scale; GP: General practitioner; GPPS: General Practitioner Patient Survey; HRQoL: Health-related quality of life; NHS: National Health Service.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13561-022-00374-y>.

Additional file 1.

Acknowledgements

The GP Patient survey is run by NHS England and NHS Improvement, who provided the data included in this paper. The views expressed by the authors do not necessarily reflect the view of the EuroQol Group.

Authors' contributions

OM and WG contributed to the conception of the work. OM applied for the data as well as analysed and interpreted the data; WG provided critical feedback to the analyses. OM drafted the manuscript. WG revised the manuscript. OM and WG read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This study was funded by EuroQol (grant number: 2015200).

Availability of data and materials

The data that support the findings of this study are available from NHS England but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of NHS England.

Declarations

Ethics approval and consent to participate

This study is part of a larger research study, which was approved by the Ethics Committee of Bielefeld University, Number 2017–207. All analyses are based on anonymised secondary data and did not require additional research ethics approval.

Consent for publication

Not applicable.

Competing interests

OM and WG are members of the EuroQol Group. There are no other conflicts of interest.

Received: 17 January 2022 Accepted: 11 May 2022

Published online: 20 May 2022

References

- Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993;118(8):622–9. <https://doi.org/10.7326/0003-4819-118-8-199304150-00009>.
- Devlin NJ, Appleby J. Getting the most out of PROMS: putting health outcomes at the heart of NHS decision-making: the king's fund; 2010.
- National Institute for Health and Care Excellence. Shared decision making: NICE guideline. 2021. <https://www.nice.org.uk/guidance/ng197/resources/shared-decision-making-pdf-66142087186885>. Accessed 14 Oct 2021.
- Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. *Appl Health Econ Health Policy*. 2017;15(2):127–37. <https://doi.org/10.1007/s40258-017-0310-5>.
- Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics*. 2017;35(5):11–9. <https://doi.org/10.1007/s40273-017-0544-y>.
- Haute Autorité de Santé. Choices in Methods for Economic Evaluation. 2012. https://www.has-sante.fr/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf. Accessed 31 Mar 2020.
- Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36. <https://doi.org/10.1007/s11113-011-9903-x>.
- Buchholz I, Janssen MF, Kohlmann T, Feng Y-S. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics*. 2018;36(6):645–61. <https://doi.org/10.1007/s40273-018-0642-5>.
- de Leeuw ED. Reducing missing data in surveys: an overview of methods. *Qual Quant*. 2001;35:147–60. <https://doi.org/10.1023/A:1010395805406>.
- Marten O, Brand L, Greiner W. Feasibility of the EQ-5D in the elderly population: a systematic review of the literature. *Qual Life Res*. 2021;31(6):1621–37. <https://doi.org/10.1007/s11136-021-03007-9>.
- Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. 2013;22(7):1717–27. <https://doi.org/10.1007/s11136-012-0322-4>.
- Agborsangaya CB, Lahtinen M, Cooke T, Johnson JA. Comparing the EQ-5D 3L and 5L: measurement properties and association with chronic conditions and multimorbidity in the general population. *Health Qual Life Outcomes*. 2014;12(1):74. <https://doi.org/10.1186/1477-7525-12-74>.
- Coast J, Peters TJ, Richards SH, Gunnell DJ. Use of the EuroQol among elderly acute care patients. *Qual Life Res*. 1998;7:1–10. <https://doi.org/10.1023/a:1008857203434>.
- Kunz S. Psychometric properties of the EQ-5D in a study of people with mild to moderate dementia. *Qual Life Res*. 2010;19(3):425–34. <https://doi.org/10.1007/s11136-010-9600-1>.
- Orgeta V, Edwards RT, Hounsborne B, Orrell M, Woods B. The use of the EQ-5D as a measure of health-related quality of life in people with dementia and their carers. *Qual Life Res*. 2015;24(2):315–24. <https://doi.org/10.1007/s11136-014-0770-0>.
- Davis JC, Liu-Ambrose T, Richardson CG, Bryan S. A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes? *Qual Life Res*. 2013;22(5):969–77. <https://doi.org/10.1007/s11136-012-0225-4>.
- Christiansen ASJ, Møller MLS, Kronborg C, Haugan KJ, Køber L, Højberg S, et al. Comparison of the three-level and the five-level versions of the EQ-5D. *Eur J Health Econ*. 2021;22(4):621–8. <https://doi.org/10.1007/s10198-021-01279-z>.
- Brazier JE, Walters SJ, Nicholl JP, Kohler B. Using the SF-36 and Euroqol on an elderly population. *Qual Life Res*. 1996;5(2):195–204. <https://doi.org/10.1007/bf00434741>.
- Hulme C, Long AF, Kneafsey R, Reid G. Using the EQ-5D to assess health-related quality of life in older people. *Age Ageing*. 2004;33(5):504–7. <https://doi.org/10.1093/ageing/afh178>.
- Grund S, Breiting E, Fricke S, Alpers G, Hundsdorfer W, Schäfer H. Health-related quality of life measurement inpatient geriatric rehabilitation: a comparison of the feasibility and suitability between the SF-36 and EQ-5D-5L questionnaires. *J Gerontol Geriatr Res*. 2017;06(06). <https://doi.org/10.4172/2167-7182.1000458>.
- Michalowsky B, Xie F, Kohlmann T, Gräse J, Wübbeler M, Thyriar JR, et al. Acceptability and validity of the EQ-5D in patients living with dementia. *Value Health*. 2020;23(6):760–7. <https://doi.org/10.1016/j.jval.2020.01.022>.
- Badia X, Schiaffino A, Alonso J, Herdman M. Using the EuroQol 5-D in the Catalan general population: feasibility and construct validity. *Qual Life Res*. 1998;7(4):311–22. <https://doi.org/10.1023/A:1024933913698>.
- Hughes DA. Feasibility, validity and reliability of the welsh version of the EQ-5D health status questionnaire. *Qual Life Res*. 2007;16(8):1419–23. <https://doi.org/10.1007/s11136-007-9238-9>.
- Yfantopoulos J, Chantzaras A, Kontodimas S. Assessment of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in psoriasis. *Arch Dermatol Res*. 2017;309(5):357–70. <https://doi.org/10.1007/s00403-017-1743-2>.
- Gerard K, Nicholson T, Mullet M, Mehta R, Roderick P. EQ-5D versus SF-6D in an older, chronically ill patient group. *Appl Health Econ Health Policy*. 2004;3(2):91–102. <https://doi.org/10.2165/00148365-200403020-00005>.
- Peyre H, Coste J, Leplège A. Identifying type and determinants of missing items in quality of life questionnaires: application to the SF-36 French version of the 2003 decennial health survey. *Health Qual Life Outcomes*. 2010;8(1):16. <https://doi.org/10.1186/1477-7525-8-16>.
- Coste J, Quinquin L, Audureau E, Pouchot J. Non response, incomplete and inconsistent responses to self-administered health-related quality of life measures in the general population: patterns, determinants and impact on the validity of estimates - a population-based study in France using the MOS SF-36. *Health Qual Life Outcomes*. 2013;11:44. <https://doi.org/10.1186/1477-7525-11-44>.

28. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
29. van Leeuwen KM, Jansen APD, Muntinga ME, Bosmans JE, Westerman MJ, van Tulder MW, et al. Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. *BMC Health Serv Res*. 2015;15:201. <https://doi.org/10.1186/s12913-015-0862-8>.
30. IPSOS MORI. GP patient survey: national report - July 2017 publication. 2017. <http://gp-patient.co.uk/surveysandreports2017>. Accessed 6 May 2022.
31. IPSOS MORI. GP Patient Survey: Technical Annex. 2017. <https://gp-patient.co.uk/downloads/archive/2017/GPPS%202017%20Technical%20Annex%20PUBLIC.pdf>. Accessed 31 Mar 2020.
32. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72. [https://doi.org/10.1016/0168-8510\(96\)00822-6](https://doi.org/10.1016/0168-8510(96)00822-6).
33. EuroQol Research Foundation. EQ-5D-5L User Guide. 2019. <https://euroqol.org/publications/user-guides/>. Accessed 31 Mar 2020.
34. IPSOS MORI. GP Patient Survey: Questionnaire redevelopment. 2018. <https://www.gp-patient.co.uk/Files/GPPS%20Y12%20Questionnaire%20Redevelopment%20report%20v1%20PUBLIC.PDF>. Accessed 31 Mar 2020.
35. IPSOS MORI. GP patient survey: national summary report - January 2016 publication. 2016. <https://gp-patient.co.uk/surveysandreports-10-16>. Accessed 6 May 2022.
36. IPSOS MORI. GP patient survey: national summary report - July 2016 publication. 2016. <https://gp-patient.co.uk/surveysandreports-10-16>. Accessed 6 May 2022.
37. IPSOS MORI. GP Patient Survey: Year 2011/2012 Summary Report. 2012. <https://gp-patient.co.uk/surveysandreports-10-16>. Accessed 6 May 2022.
38. IPSOS MORI. GP patient survey: January - September 2012 Summary Report 2012. <https://gp-patient.co.uk/surveysandreports-10-16>. Accessed 6 May 2022.
39. Rea LM, Parker RA. Designing and conducting survey research: A comprehensive guide. 4th ed. San Francisco: Jossey-Bass; 2014.
40. Corp S. Stata Statistical software: release 16. College Station: StataCorp LLC; 2019.
41. García-Gordillo MA, Adsuar JC, Olivares PR. Normative values of EQ-5D-5L: in a Spanish representative population sample from Spanish health survey, 2011. *Qual Life Res*. 2016;25(5):1313–21. <https://doi.org/10.1007/s11136-015-1164-7>.
42. Grochtdreis T, Dams J, König H-H, Konnopka A. Health-related quality of life measured with the EQ-5D-5L: estimation of normative index values based on a representative German population sample and value set. *Eur J Health Econ*. 2019;20(6):933–44. <https://doi.org/10.1007/s10198-019-01054-1>.
43. Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. EQ-5D-3L derived population norms for health related quality of life in Sri Lanka. *PLoS One*. 2014;9(11):e108434. <https://doi.org/10.1371/journal.pone.0108434>.
44. Holland R, Smith RD, Harvey I, Swift L, Lenaghan E. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQL. *Health Econ*. 2004;13(8):793–805. <https://doi.org/10.1002/hec.858>.
45. Corbacho B, Keding A, Chuang L-H, Ramos-Goni JM, Joshi K, Cockayne S, et al. Comparison of the EQ-5D-5L and the EQ-5D-3L using individual patient data from the REFORM trial. *F1000Res*. 2021;10:974. <https://doi.org/10.12688/f1000research.54554.1>.
46. Młyńczak K, Golicki D. Validity of the EQ-5D-5L questionnaire among the general population of Poland. *Qual Life Res*. 2021;30(3):817–29. <https://doi.org/10.1007/s11136-020-02667-3>.
47. Perneger TV, Combescurie C, Courvoisier DS. General population reference values for the French version of the EuroQol EQ-5D health utility instrument. *Value Health*. 2010;13(5):631–5. <https://doi.org/10.1111/j.1524-4733.2010.00727.x>.
48. Ferreira LN, Ferreira PL, Ribeiro FP, Pereira LN. Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. *Health Qual Life Outcomes*. 2016;14(1):89. <https://doi.org/10.1186/s12955-016-0491-x>.
49. Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whyhnes DK, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged or = 45 years. *Health Econ*. 2008;17(7):815–32. <https://doi.org/10.1002/hec.1298>.
50. Luthy C, Cedraschi C, Allaz A-F, Herrmann FR, Ludwig C. Health status and quality of life: results from a national survey in a community-dwelling sample of elderly people. *Qual Life Res*. 2015;24(7):1687–96. <https://doi.org/10.1007/s11136-014-0894-2>.
51. DeCastellarnau A. A classification of response scale characteristics that affect data quality: a literature review. *Qual Quant*. 2018;52(4):1523–59. <https://doi.org/10.1007/s11135-017-0533-4>.
52. de Leeuw E, Hox J, Huisman M. Prevention and treatment of item nonresponse. *J Off Stat*. 2003;19:153–76.
53. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics*. 2018;36(6):675–97. <https://doi.org/10.1007/s40273-018-0623-8>.
54. Mulhern B, Feng Y, Shah K, Janssen MF, Herdman M, van Hout B, et al. Comparing the UK EQ-5D-3L and English EQ-5D-5L value sets. *Pharmacoeconomics*. 2018;36(6):699–713. <https://doi.org/10.1007/s40273-018-0628-3>.
55. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value Health*. 2008;11(2):275–84. <https://doi.org/10.1111/j.1524-4733.2007.00230.x>.
56. Hickson M, Frost G. An investigation into the relationships between quality of life, nutritional status and physical function. *Clin Nutr*. 2004;23(2):213–21. [https://doi.org/10.1016/S0261-5614\(03\)00127-4](https://doi.org/10.1016/S0261-5614(03)00127-4).
57. Rohr M, Brandstetter S, Plomer A-S, Loss J, Kretschmer R, Apfelbacher C. A qualitative study exploring content validity and feasibility of frequently used generic health-related quality of life measures in older people with hip fracture: the patients' perspective. *Injury*. 2021;52(2):134–41. <https://doi.org/10.1016/j.injury.2020.09.061>.
58. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health (Oxf)*. 2005;27(3):281–91. <https://doi.org/10.1093/pubmed/fdi031>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

