OHE Office of Health Economics

A MULTINATIONAL QUANTITATIVE STUDY

Understanding the preferences of people with acute leukemia for different health outcomes

David Mott
Hannah Hussain
Sulayman Chowdhury
Samantha Nier
Chris Skedgel

ALAN
Acute Leukaemia Advocates Network

SEPTEMBER 2025

A MULTINATIONAL QUANTITATIVE STUDY

Understanding the preferences of people with acute leukemia for different health outcomes

David Mott
Office of Health Economics, London, UK

Hannah Hussain
Office of Health Economics, London, UK

Sulayman Chowdhury
Office of Health Economics, London, UK

Samantha Nier
Acute Leukemia Advocates Network, Bern, Switzerland

Chris Skedgel
Office of Health Economics, London, UK

OFFICE OF HEALTH ECONOMICS &
ACUTE LEUKEMIA ADVOCATES NETWORK

Please cite this report as:

Mott D.J., Hussain H., Chowdhury S., Nier S., Skedgel C. 2025. Understanding the preferences of people with acute leukemia for different health outcomes: a multinational quantitative study. OHE Grant-Funded Research Report, London: Office of Health Economics. Available at: https://www.ohe.org/publications/understanding-the-preferences-of-people-with-acute-leukemia-for-different-health-outcomes/ and https://www.acuteleuk.org/publications

Corresponding Author:
David Mott
dmott@ohe.org

Funding

This research paper was commissioned and funded by the Acute Leukemia Advocates Network (ALAN).

Acknowledgements

The authors are grateful to Professor Richard Norman for his review of an earlier draft.

About OHE Grant-Funded Research Reports

Many of the studies OHE performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications. Publication is at the client's discretion.

Studies published by OHE as OHE Grant-Funded Research Reports are subject to internal quality assurance and undergo external review, usually by a member of OHE's Editorial Panel. Any views expressed are those of the authors and do not necessarily reflect the views of OHE as an organization.

Table of Contents

Executive Summary

When we want to understand or measure someone's "health-related quality of life", we typically ask them to describe the problems they have with different aspects of their health. In this study, we wanted to (a) look at the health-related quality of life of people with acute leukemia (AL) and (b) understand which aspects, or "dimensions", of health-related quality of life are most important people with AL. We also sought to compare the responses of people with AL with people from the general public.

What is the health-related quality of life of people with acute leukemia?

Members of the general public and people living with AL were asked to complete a questionnaire about their health-related quality of life. The usual questionnaire is called the EQ-5D and it is commonly used to make decisions in healthcare. It has five questions about different dimensions of health: mobility, ability to look after oneself (self-care), ability to participate in usual activities, pain/discomfort, and anxiety/depression.

In this study, we asked people to complete the usual EQ-5D, and then we added two more dimensions we thought would be relevant to people with AL: tiredness and cognition. In each version, people described their experience on each dimension, from "no problems" to "extreme problems". The infographic below indicates the percentage of people that reported having some problems on each dimension.

% of people with AL that reported problems with...						
Tiredness*	Pain/ discomfort	Anxiety/ depression	Usual activities	Mobility	Cognition*	Self-Care
97%	91%	90%	86%	63%	56%	42%
Compared to these %'s in the general population...						
73%	58%	50%	24%	24%	38%	13%

*Non-core EQ-5D dimensions, i.e., bolt-ons

With the usual EQ-5D questionnaire, people with AL reported having particular problems with pain/discomfort, anxiety/depression, and doing their usual activities. The percentages of people in the general population reporting problems on these dimensions were much lower in comparison, illustrating the impact of AL on health-related quality of life. When we added the two additional dimensions, though, we found that almost 100% of persons with AL had problems with tiredness, and more than half experienced cognition problems. This is an important finding, because it means that the usual EQ-5D may be overlooking some of the problems associated with AL.

While everyday tasks can be challenging, people with AL report far greater problems with non-visible symptoms such as pain/discomfort, tiredness, anxiety/depression and cognitive difficulties. These hidden burdens are common - with severe tiredness and anxiety/depression affecting around one in three patients - and they highlight why patient-reported outcomes must look beyond what can be seen from the outside.

We also wanted to understand how health-related quality of life might differ in people that have different characteristics e.g., older/younger, male/female, type of AL. We found that:

People with a longer disease duration...		*People with prior relapse experience...*
	*and*	
*...had **better** health-related quality of life*		*...had **worse** health-related quality of life*

The longer disease duration finding may be because the people that feel the worst might not participate in research studies like this (known as selection bias) or alternatively that, over time, people adjust and find better ways to cope with their illness (known as adaptation). The relapse history finding could be because of the impact of the relapse itself, and/or its treatment.

Which aspects of health-related quality of life matter most to people with acute leukemia?

In addition to asking people about their problems on the different dimensions of health, we asked them to rank which aspects of health matter most to them. This was done using a "best-worst scaling" survey, where respondents chose the most and least important health problems in different scenarios.

	#1 Most important	#2	#3	#4	#5	#6	#7 Least important
Patients	Pain/ discomfort	Cognition*	Mobility	Anxiety/ depression	Usual activities	Tiredness*	Self-care
General Public	Pain/ discomfort	Mobility	Cognition*	Usual activities	Self-care	Anxiety/ depression	Tiredness*

*Non-core EQ-5D dimensions, i.e., bolt-ons

Patients and the general public agreed that being free from pain/discomfort is most important. Beyond this, patients placed greater emphasis on cognition, while the public focused more on mobility and self-care. Patients also rated tiredness as slightly more important than the public did. These differences reflect the lived experience of acute leukemia - patients adapt to some daily limitations but continue to prioritize symptoms that affect how they think, feel, and cope day-to-day.

What this means for patients and advocates

Improving care for people with AL means listening closely to what matters most to patients. This goes beyond controlling the disease, it's also about recognizing and addressing the hidden symptoms that affect daily life. By capturing these experiences in routine care and research, we can make sure treatment strategies and health policies reflect the patient's voice.

In everyday care

Treat the whole person, not just the disease

- Pain/discomfort, fatigue, cognition and anxiety/depression should be regularly assessed.
- Early detection of these problems can guide better support and treatment.

Focus where patients focus

- Patients report that tiredness and cognition matter just as much as physical function.
- Making space to discuss these symptoms in clinic visits ensures patients feel heard.

In treatment and research decisions

Value what improves daily life

- Treatments that control disease, but also reduce fatigue, pain/discomfort or cognitive difficulties can make a big difference.
- These benefits may not always be captured if decisions rely only on "standard" measures.

Tools must evolve to reflect patient voice

- Adding "bolt-on" questions on cognition and tiredness helps capture what matters most.
- This ensures future studies and health policy decisions take patient priorities into account.

1. Introduction

Acute leukemia (AL) comprises a family of rapidly progressing hematological malignancies: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML) and acute promyelocytic leukemia (APL) - that share an urgent need for treatment but differ in terms of their underlying biology, typical age of onset and therapeutic pathways. AML and APL occur predominantly in adults and usually present abruptly; ALL is most common in childhood but is increasingly seen in older adults, who face a distinct clinical and psychosocial trajectory: their biological vulnerability limits treatment tolerance, prognosis is comparatively poor (with just 30-40% long-term remission rates), and they often endure heightened symptom burden and reduced quality of life due to both disease and therapy (Terwilliger and Abdul-Hay, 2017). Disease prognosis and quality of life (QoL) outcomes may vary across subtypes and patient demographics, making personalized and patient-centered care increasingly important. Survival has improved over the last two decades, yet treatment remains highly intensive and is frequently accompanied by prolonged hospitalization, severe toxicities and lasting functional limitations such as fatigue, pain and cognitive impairment (Andrés-Jensen et al., 2024).

Recent evidence highlights the heterogeneity of patients' experiences. In a 76-country cross-sectional survey of >2,500 individuals with leukemia (all types, including chronic), large and systematic variation was shown in self-reported health-related quality of life (HRQoL) by subtype, gender and age with the Hematological Malignancy Patient-Reported Outcome (HM-PRO) instrument. People with AL, particularly those with ALL, as well as women, consistently reported worse scores, while older age and longer disease duration were weakly associated with better HRQoL, suggesting possible adaptation and response shift to the condition over time (Salek et al., 2025).

Treatment decisions in AL are often complex, involving difficult trade-offs between survival, toxicity and quality of life. This is especially true in the context of relapsed or refractory disease, where treatment options may be more aggressive and carry significant burden. Patients can face intensive therapies, prolonged hospitalizations, and persistent side effects such as fatigue, pain, and cognitive impairment - symptoms that may continue well beyond active treatment. Yet clinical decision-making frameworks often underrepresent what

matters most to patients. A UK-based discrete-choice experiment in relapsed/refractory acute leukemia found that while chance of response was the most important driver of choice, quality of life outcomes also play an important role in decision-making - however there may be substantial preference heterogeneity across individuals (Mott et al., 2024). Capturing such heterogeneity systematically is a requirement for genuinely patient-centered decision making. Although previous research confirms that quality of life matters to people living with acute leukemia, less is known about which specific aspects they prioritize most. In a large best-worst scaling (BWS) study in AML, patients consistently ranked the possibility of dying and the risk of long-term side effects as their greatest concerns - rated more than twice as high as items related to care delivery and decision-making (Richardson et al., 2021). These findings highlight the importance of going beyond general quality of life assessments to understand which specific outcomes patients value most. This study builds upon existing work by identifying and quantifying quality of life priorities among people living with acute leukemia, to support more preference-informed care and person-centered decision-making.

The EQ-5D-5L holds a central role in health technology assessment (HTA), where it is the most commonly used generic measure for capturing health-related quality of life (HRQoL) in economic evaluations (Rowen et al., 2023), and is the preferred instrument recommended by NICE for use in submissions to inform reimbursement decisions (NICE, 2022). Its five 'core' dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) are valued using general population preferences and summarized into a single index score (a 'utility'), which can be used to generate quality-adjusted life years (QALYs) for use in economic evaluation. The current five-level version (EQ-5D-5L) offers five severity levels for each dimension, improving sensitivity and reducing ceiling effects compared to the earlier three-level version (EQ-5D-3L).

EQ-5D-5L is a generic tool that has been designed to measure HRQoL across a wide range of diseases and populations and it can be applied across interventions, allowing results to be compared in a consistent manner. However, its brevity means it may not fully capture the symptoms and functional challenges that matter most to patients in specific clinical contexts. This has led to criticism that the EQ-5D-5L may overlook key domains of importance in some conditions. In response, the EuroQol Group is developing an "EQ-5D

Bolt-on Toolbox", whereby additional items will be developed such that they can be appended to the core five dimensions when warranted (Devlin et al., 2025). Among the initial set of bolt-ons for inclusion in the toolbox are fatigue (tiredness), which is well established in the literature, and cognition, which is undergoing advanced testing and validation (Rencz and Janssen, 2024). Both symptoms are particularly relevant in acute leukemia (Meyers, Albitar and Estey, 2005; Tomaszewski et al., 2016), where patients frequently report sustained tiredness, cognitive difficulties and treatment-related exhaustion that may not be fully captured by the core dimensions of EQ-5D-5L (van Dongen-Leunis, Redekop and Uyl-de Groot, 2016).

A further consideration is that EQ-5D value sets reflect preferences of the general public, for a variety of normative reasons (Versteegh and Brouwer, 2016). However, a growing body of evidence shows that patients may evaluate health states differently from people that have never experienced them (Gandhi et al., 2017; Peeters and Stiggelbout, 2010). Lived experience can lead to adaptation (being less concerned about specific health issues after becoming accustomed to them), response shift (a recalibration of internal standards for what constitutes "good health"), or simply a more nuanced appreciation of treatment burden. For example, cancer survivors consistently place greater weight on fatigue and cognitive fog than general public respondents do - even when generic quality of life instruments treat these issues as secondary (Fayers and Machin, 2013). On the other hand, the general public may over-penalize impairments that patients regard as manageable.

Given these different effects, examining preferences from both stakeholder perspectives offers complementary insights: general public values remain essential for economic evaluation and equity considerations, while patient-derived preferences highlight how those living with the disease actually trade off benefits and burden in day-to-day life.

In this study, we aimed to better understand the relative importance of seven health outcomes: the core EQ-5D-5L dimensions plus two bolt-ons, tiredness and cognition, from the perspective of people living with acute leukemia, as well as the general population. We used profile-case (case II) BWS to quantify the value people place on these different aspects of HRQoL.

1.1 Study Objectives

The primary objective of this study was to:

- Explore the preferences of people with acute leukemia for the five health outcomes included in EQ-5D instruments as well as two additional items: tiredness and cognition.

Secondary objectives were to:

- Compare patients' preferences with general population preferences.
- Examine the HRQoL of patients and contrast this with the general population.
- Explore how patient HRQoL may vary based on different demographic and clinical variables.

In the following sections, we set out the methodology used to address the study objectives (Section 2), describe the results (Section 3), and discuss the implications of the findings (Section 4), followed by a conclusion (Section 5).

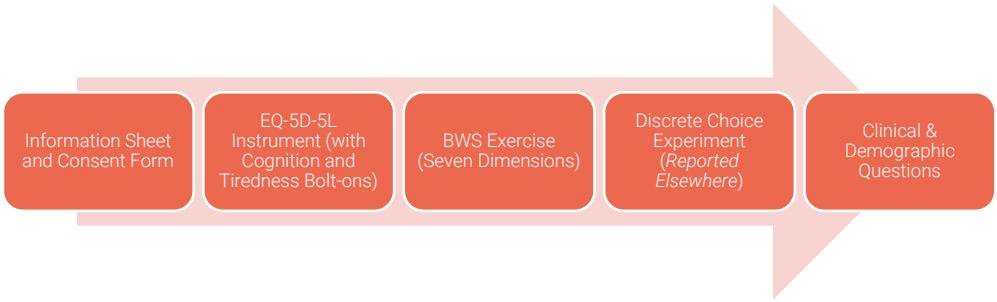
2. Methods

2.1 Overview

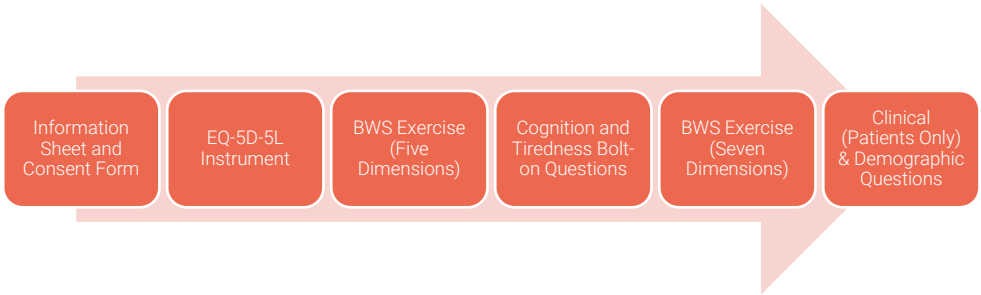
This study comprised online surveys delivered to two different populations (patients and general population) across six different study countries (UK, France, Germany, Italy, Spain, and USA). The content of the surveys varied depending on the sample, including the BWS exercises. Figure 1 shows the two different versions of the surveys.

FIGURE 1. SURVEY VERSIONS

Survey Version A



Survey Version B



Survey version A served two purposes. First, it enabled a geographical extension of an earlier project that sought to explore patients' treatment preferences using a discrete choice experiment (DCE), in a relapsed/refractory setting in the UK (Mott et al., 2024). The results of this geographical extension will be reported elsewhere (Mott et al., 2025). Second, it provided data for the project reported here, via the BWS exercise. Patients in all countries, except for the UK, saw survey version A.

Survey version B related solely to the project reported here. Given that value sets for EQ-5D instruments are typically produced based on general population preferences, we included general population samples in this study to enable comparisons with patient preferences. Additionally, a second BWS exercise with only the core five dimensions of EQ-5D was included in this survey version, to explore the impact of adding bolt-on dimensions (these comparisons will be reported elsewhere). Finally, as a UK patient sample had already completed the DCE, the UK patient sample for this study completed survey version B rather than version A. In summary, all general population samples, and the UK patient sample, saw survey version B.

The remainder of this section describes the design of the BWS exercises, study recruitment processes, and analytical approaches.

2.2 Design of the BWS exercise

BWS encompasses a range of techniques, to be discussed in more detail below, but is based on the ranking of different elements. However, rather than ordering all possible elements, as in a conventional ranking task, it simplifies the exercise by asking respondents to identify only the best (or most-preferred) and worst (or least-preferred) elements from the larger set of options. It is based on the idea that respondents can more easily identify the best and worst, or most and least important, elements from a set than provide a complete ranking of all the elements in that set (Flynn et al., 2007).

2.2.1 Choice of variant

A profile-case BWS (sometimes known as 'Case II') was chosen for this study. This version of BWS is specifically intended to capture the relative importance of different elements, and the trade-offs between those elements (Flynn, 2010), and has previously been used for assessing health states defined by a multi-attribute utility instrument (Ratcliffe et al., 2012). While this method cannot be used to generate a value set (at least not without additional 'anchoring' data being collected), this was not a problem for this study given our objectives.

2.2.2 Structure of the exercise

As described in section 2.1, there were two different BWS exercises. In both exercises, participants are shown a health state that is described using a set of

different dimensions of HRQoL. Each dimension has a different severity level e.g., "no problems", "moderate problems", or "extreme problems". In the task, respondents are asked to select the best (or "least bad") and the worst (or "most bad") levels. They are then shown a new health state, with different levels, and asked the same question again.

The difference between the two BWS exercises is the number of dimensions included. In one of the exercises, completed by the general population and UK patient samples, only the core five dimensions of EQ-5D are included (hereafter referred to as "BWS-5D"). In the other task, completed by all samples, the core five dimensions of EQ-5D-5L are included alongside two additional 'bolt-on' items: cognition and tiredness (hereafter referred to as "BWS-7D").

The EQ-5D-5L instrument describes each dimension using five severity levels (from no problems to extreme problems). The five-level version of EQ-5D has better measurement properties compared to the earlier three-level version (EQ-5D-3L) (Buchholz et al., 2018; Thompson and Turner, 2020). While we use the five-level version of EQ-5D when collecting respondents' self-reported HRQoL data in the surveys (see Figure 1), in the BWS exercises we opted to use only three of the five severity levels. This was to reduce the complexity of the task. Specifically, we used the best level (no problems or level 1), the middle level (moderate problems or level 3), and the worst level (extreme problems or level 5) from EQ-5D-5L in the BWS exercises. As such, we describe our BWS exercises as using an "adapted" EQ-5D-5L descriptive system, rather than the EQ-5D-3L descriptive system.

Table 1 sets out the dimensions and levels used in the BWS exercises, with the wording from the UK surveys.

TABLE 1. DIMENSIONS AND LEVELS USED IN THE BWS EXERCISES (UK WORDING)

	DIMENSION	LEVELS
Core dimensions[1]	**Mobility (MO)**	1. No problems in walking about 3. Moderate problems in walking about 5. Unable to walk about
	Self-care (SC)	1. No problems washing or dressing myself 3. Moderate problems washing or dressing myself 5. Unable to wash or dress myself
	Usual activities (UA)	1. No problems doing my usual activities 3. Moderate problems doing my usual activities 5. Unable to do my usual activities
	Pain/discomfort (PD)	1. No pain or discomfort 3. Moderate pain or discomfort 5. Extreme pain or discomfort
	Anxiety/depression (AD)	1. Not anxious or depressed 3. Moderately anxious or depressed 5. Extremely anxious or depressed
Bolt-on dimensions[2]	**Tiredness (TI)**	1. Not tired 3. Moderately tired 5. Extremely tired
	Cognition (CO)	1. No problems with cognition 3. Moderate problems with cognition 5. Extreme problems with cognition

[1]Included in both BWS exercises. [2]Only included in the BWS-7D exercise.

On occasion hereafter, acronyms will be used when referring to specific dimension levels in line with Table 1, for example, MO1 (no problems in walking about), PD5 (extreme pain or discomfort) and TI3 (moderately tired).

Tiredness and cognition were chosen due to their relevance in acute leukemia and their potential for consideration as formal bolt-ons for EQ-5D as part of a future EQ-5D Bolt-on Toolbox (Devlin et al., 2025). The wording used for tiredness is from the bolt-on developed by Yang et al. (2015). The wording used for cognition was based on a bolt-on developed in an ongoing research study led by OHE and the University of Technology, Sydney.

A screenshot of a health state from the BWS-7D exercise is included in Figure 2.

FIGURE 2. EXAMPLE HEALTH STATE FROM THE BWS-7D EXERCISE (UK WORDING)

What is the best (least bad) aspect?	Choose one best and one worst aspect	What is the worst (most bad) aspect?
○	**Unable** to walk about	○
○	**Unable** to wash or dress myself	○
○	**Unable** to do my usual activities	○
○	**Extreme** pain or discomfort	○
○	**Extremely** anxious or depressed	○
○	**Extremely** tired	○
○	**Extreme problems** with cognition	○

2.2.3 Experimental design

The experimental design of the BWS-7D exercise was based on a seven-dimension orthogonal array containing 27 rows, which was identified using R. Severity levels were assigned to the array to minimize the number of 'dominant' options, as these options are not informative. For example, a dominant option would be a health state where only one dimension was set at "no problems", making it a logical choice to select as the "best" level. For the BWS-5D task, the same overall design as the BWS-7D was used, with two dimensions removed.

Given that 27 health states are too many to feasibly ask respondents to consider in a survey, the designs were subsequently split up (blocked) in R with level balance prioritized. For participants completing survey version A (see Figure 1), the BWS-7D design was split into three blocks of nine health states. For participants completing survey version B, both the BWS-5D and BWS-7D designs were split into two blocks, with the worst health state (highest severity

level in all dimensions) included in both, resulting in 14 health states being assessed by each participant, in each task. The experimental design can be found in Appendix A1.

2.3 Sample and recruitment

For the patient samples, recruitment targeted approximately n=80 patients in the US, n=40 in the UK and n=100 patients across Germany, France, Italy and Spain (EU4). UK participants were invited through the charity Leukemia Care, while patients in the US and EU4 were identified and screened by a specialist patient recruitment agency. Adult patients were eligible with a diagnosis of acute leukemia of any type, at any treatment stage, and no specific age or gender quotas were employed.

The target general population sample in each country was 100 respondents, conducted via an online panel. Respondents were eligible if they were aged >18 years, and no specific age or gender quotas were employed.

2.4 Ethics approval

Ethical approval for this study provided by the Economics Research Ethics Committee at City St George's, University of London (formerly City, University of London), application ID ETH2425-0034.

2.5 Statistical analysis

2.5.1 Self-reported HRQoL

Responses to each dimension of the EQ-5D-5L, along with the two bolt-on dimensions, were summarized using descriptive statistics (proportions responding with each severity level) and split by sample type (patients and general population). It is important to note that, in contrast to the BWS exercises (which had three severity levels), for the self-reported HRQoL, respondents had a choice between five severity levels for each dimension (where level one represents "no problems" and level five represents "extreme

problems"). Dimension scores were also examined by leukemia type (ALL, AML and APL).

Three different summary scores were also examined:

- **EQ VAS:** This is the response to the visual analogue scale component of the EQ-5D-5L instrument, where respondents rate their health on a scale of 0-100. A score of 0 represents the "worst health imaginable" and 100 represents the "best health imaginable".

- **Level Sum Score (LSS):** Sometimes called a 'misery score', this is a sum of the scores on each level. For example, imagine health state 21111, which represents slight problems walking about and no problems on any of the other four core dimensions. The LSS for this health state is six (2+1+1+1+1). As such, the best possible LSS is five (1+1+1+1+1) and the worst possible LSS is 25 (5+5+5+5+5). LSS can be compared across countries and between different studies. Note that we do not incorporate the bolt-on items into the LSS calculation in order to preserve comparability with external studies.

- **Utility:** This is the utility for the respondent's health state, based on the US EQ-5D-5L value set by Pickard et al. (2019). This value set was chosen *a priori* because the US was the country with the largest target sample size for patients, and the use of one value set in our analysis enabled comparisons of utilities across countries within our study (comparisons would be limited if we applied different value sets for respondents in different countries).

Descriptive statistics (means, standard deviations) were prepared for the above three summary scores for each sample type (patient and general population), as well as by leukemia type (ALL, AML, and APL). The distributions of responses for each summary score, for each sample type, were also examined. Finally, descriptive statistics (means, standard deviations) by country were also prepared for the three summary scores. This analysis was performed in Stata version 15.1.

2.5.2 Determinant of patient HRQoL

To examine the factors that influence patient HRQoL, we conducted ordinary least squares (OLS) regression analyses. Based on the findings by Salek et al. (2025), we hypothesized that:

- **Age:** Older patients would have better HRQoL than younger patients.
- **Gender:** Female patients would have worse HRQoL than non-female patients.
- **Leukemia type:** People with ALL would have worse HRQoL than people with AML/APL.
- **Disease duration:** People living with the disease for longer would have better HRQoL.

In addition to testing the above hypotheses, we also sought to explore the impact of experiencing a relapse on HRQoL, with the hypothesis that those that have experienced a relapse in the past would have worse HRQoL than those that have not.

We used two different summary scores as the dependent variables: LSS and utility. We opted not to use EQ VAS scores in this analysis, as these scores are more subjective and may be impacted by response scale bias.

The model specification used is set out in Equation 1.

$$HRQoL_i = \alpha + \beta_1 Age_i + \beta_2 Female_i + \beta_3 AML_i + \beta_4 APL_i + \beta_5 Diagyrs_i + \beta_6 Relapsed_i \quad (1)$$

Where $HRQoL_i$ is the summary score for individual i (either LSS or utility, depending on the model), α is a constant, Age_i is the individual's age, $Female_i$ is a dummy variable that equals one if the individual is female, AML_i and APL_i are dummy variables that equal one if the individual has AML or APL respectively, $Diagyrs_i$ is the number of years since the individual's diagnosis, and $Relapsed_i$ is a dummy variable that equals one if the individual has experienced a relapse in the past. This analysis was performed in Stata version 15.1.

2.5.3 BWS data

Two different approaches were taken for analyzing the BWS data: a counting approach, and a modelling approach.

In the counting approach, the number of times a dimension level is chosen as 'best' or 'worst' are examined (Louviere, Flynn and Marley, 2015). Best-worst scores are then produced, which subtract the number of times a dimension level is chosen as 'worst' from the number of times a dimension level is chosen as 'best'. Best-worst scores are positive if the dimension level was chosen as 'best' more frequently than it was chosen as 'worst', and negative if the reverse is true. The best-worst scores are then standardized by taking the number of appearances into account (i.e., divided by the number of times the dimension level appeared in the BWS tasks) to enable comparisons. This analysis was conducted in Microsoft Excel.

While the counting approach is intuitive, it does not take into account all of the available information. For example, it does not consider the alternative options available to respondents when they select a particular dimension level as 'best' or 'worst'. In contrast, modelling takes this information into account, whilst also enabling an examination of uncertainty.

In the modelling approach, we adopted a marginal-sequential model which assumes that respondents select the worst level from all five/seven levels in the health state and subsequently select the best level from the remaining four/six levels. Multinomial logit (MNL) models were estimated using the R package *supportBWS2*. The model specification for the BWS-7D exercise is presented in Equation 2.

$$U_j = \begin{pmatrix} \beta_1 MO_2 + \beta_2 MO_3 + \beta_3 SC_2 + \beta_4 SC_3 + \beta_5 UA_2 \\ +\beta_6 UA_3 + \beta_7 PD_2 + \beta_8 PD_3 + \beta_9 AD_2 + \beta_{10} AD_3 \\ +\beta_{11} TI_2 + \beta_{12} TI_3 + \beta_{13} CO_2 + \beta_{14} CO_3 \end{pmatrix} + \varepsilon_j \qquad (2)$$

Where U_j represents the utility of an individual for alternative j, which is a linear function of the dimension levels (using the notation introduced in Table 1), and ε_j is an unknown random component. The dimension level variables were effects coded. Relative importance (RI) scores were estimated using the resulting coefficients, by taking the full utility range for each dimension and dividing this by the total utility range (for all dimensions combined).

3. Results

3.1 Respondent characteristics

We did not meet our recruitment target in Spain, and therefore it was dropped from the analysis. As such, the European data were combined to create an 'EU3' group comprising France, Germany, and Italy. Overall, there were 723 respondents to the surveys, of which 212 were people with acute leukemia (29%) and 511 were from the general population (71%). Tables 2 and 3 provide an overview of the characteristics of the patient and general population respondents, respectively.

TABLE 2. DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF PATIENT SAMPLES

N (%)	FULL SAMPLE (N=212)	UK (N=40)	US (N=88)	EU 3 (N=84)
Age (Mean (SD))	53.7 (13.7)	41.2 (16.0)	59.7 (8.8)	53.4 (12.7)
Gender				
Male	77 (36%)	8 (20%)	36 (41%)	51 (61%)
Female	135 (64%)	32 (80%)	52 (59%)	33 (39%)
Ethnicity				
White	161 (76%)	36 (90%)	43 (49%)	82 (98%)
Black	17 (8%)	1 (3%)	16 (18%)	0 (0%)
Asian	9 (4%)	2 (5%)	7 (8%)	0 (0%)
Mixed	2 (1%)	1 (3%)	0 (0%)	1 (1%)
Other	12 (6%)	0 (0%)	11 (13%)	1 (1%)
Hispanic/Latino*	11 (5%)	-	11 (13%)	-
Education				
Completed high school	192 (91%)	33 (83%)	88 (100%)	71 (85%)
Has degree or equivalent	106 (50%)	7 (18%)	40 (45%)	47 (56%)
Marital status				
Married	121 (57%)	24 (60%)	57 (65%)	40 (48%)
Not married	91 (43%)	16 (40%)	31 (35%)	44 (52%)
Responsible for children	67 (32%)	18 (45%)	26 (30%)	23 (28%)
Yes	144 (68%)	22 (55%)	62 (70%)	60 (72%)
No	1 (<1%)	0 (0%)	0 (0%)	1 (1%)
Not reported				
Leukemia type				
ALL	62 (29%)	11 (28%)	40 (46%)	11 (13%)
AML	133 (63%)	28 (70%)	34 (39%)	71 (85%)
APL	17 (8%)	1 (3%)	14 (16%)	2 (2%)
Age at diagnosis				
(Mean (SD))	50.3 (14.4)	37.1 (15.4)	56.7 (8.8)	49.8 (14.4)
Years since diagnosis				

N (%)	FULL SAMPLE (N=212)	UK (N=40)	US (N=88)	EU 3 (N=84)
(Mean (SD))	3.4 (3.5)	4.2 (4.1)	3.0 (2.5)	3.6 (4.0)
Current treatment				
None	89 (42%)	21 (53%)	32 (36%)	36 (43%)
Induction therapy	7 (3%)	0 (0%)	6 (7%)	1 (1%)
Consolidation therapy	31 (15%)	0 (0%)	10 (11%)	21 (25%)
Maintenance therapy	38 (18%)	4 (10%)	21 (24%)	13 (16%)
Awaiting transplant	10 (5%)	1 (3%)	9 (10%)	0 (0%)
Recent transplant	22 (10%)	9 (23%)	6 (7%)	7 (8%)
Don't know/not sure	6 (3%)	5 (13%)	2 (2%)	4 (5%)
Other	9 (4%)	0 (0%)	2 (2%)	2 (2%)
Relapse history				
Not achieved remission	28 (13%)	1 (3%)	15 (17%)	12 (14%)
Never relapsed	98 (46%)	30 (75%)	21 (24%)	47 (56%)
One relapse	62 (29%)	9 (23%)	33 (38%)	20 (24%)
Two relapses	21 (10%)	0 (0%)	16 (18%)	5 (6%)
More than two relapses	3 (1%)	0 (0%)	3 (3%)	0 (0%)
Transplant history				
None	94 (44%)	16 (40%)	52 (59%)	26 (31%)
One	107 (51%)	22 (55%)	30 (34%)	55 (66%)
Two	9 (4%)	2 (5%)	4 (5%)	3 (4%)
More than two	2 (1%)	0 (0%)	2 (2%)	0 (0%)

A higher proportion of participants were female across both samples and countries (n=405 pooled, 56%), barring the pooled EU3 patient sample, whereby 61% (n=51) were male. The mean age of the general population sample was 48.2 (SD: 16.2) years and within the patient population sample was 53.7 (13.7). The majority of respondents were white in both respondent populations (general population, 91%; patient population, 76%) and educated to degree (or equivalent) level (general population, 66%; patient population, 50%).

In the patient sample, the most common diagnosis was AML (n=133, 63%), followed by ALL (n=62, 29%) and APL (n=17, 8%). The average age of patient sample was 54 (SD: 14), and the mean age at diagnosis was 50 (SD: 14), both of which are relatively young for acute leukemia. Patients from the UK were younger on average than those from the EU3 and US. The most common response to the treatment status question was "not receiving active treatment" (n=89, 42%). The most common response to the experience with relapse question was "I have achieved remission, but I have never relapsed" (n=98, 46%). A slim majority of patient respondents had received at least one transplant (n=118, 56%).

TABLE 3. DEMOGRAPHIC CHARACTERISTICS OF THE GENERAL POPULATION SAMPLES

N (%)	FULL SAMPLE (N=511)	UK (N=101)	US (N=102)	EU 3 (N=308)
Age (Mean (SD))	48.2 (16.2)	47.6 (17.2)	49.2 (18.3)	48.1 (15.1)
Gender				
Male	241 (47%)	48 (48%)	46 (45%)	147 (48%)
Female	270 (53%)	53 (53%)	56 (55%)	161 (53%)
Ethnicity				
White	463 (91%)	88 (87%)	80 (78%)	295 (96%)
Black	21 (4%)	5 (5%)	13 (13%)	3 (1%)
Asian	8 (2%)	3 (3%)	2 (2%)	3 (1%)
Mixed	5 (<1%)	3 (3%)	0 (0%)	2 (<1%)
Other	10 (3%)	2 (2%)	3 (3%)	5 (2%)
Hispanic/Latino*	4 (<1%)	-	4 (4%)	-
Education				
Completed high school	433 (85%)	81 (80%)	99 (97%)	253 (82%)
Has degree or equivalent	336 (66%)	60 (59%)	66 (65%)	210 (68%)
Marital status				
Married	248 (49%)	55 (55%)	52 (51%)	141 (46%)
Not married	263 (52%)	46 (46%)	50 (49%)	167 (54%)
Children				
Yes	190 (37%)	34 (34%)	31 (30%)	125 (41%)
No	315 (62%)	66 (65%)	70 (69%)	179 (5%9)
Not reported	6 (1%)	1 (1%)	1 (1%)	0 (0%)

*Note: this was only a specific category in the US survey.

3.2 Self-reported HRQoL

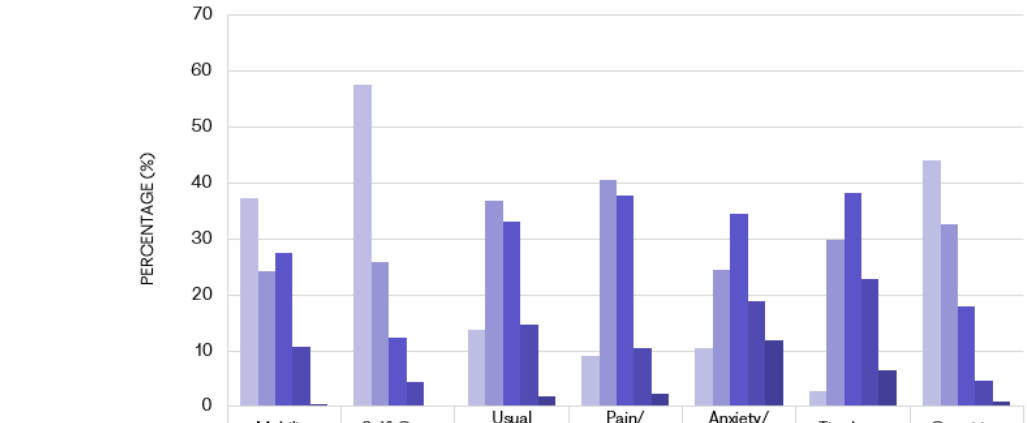
3.2.1 Issues by dimension

Figures 3 and 4 report severity-level distributions for the EQ-5D-5L descriptive system (plus tiredness and cognition bolt-ons) for both patients with acute leukemia and the general population sample. Across every dimension, the general population reported much better HRQoL.

The largest differences between patients and the general population related to physical functioning. Three-quarters (76%) of the general population sample reported that they had no problems with mobility and 89% reported that they have no problems washing or dressing themselves (self-care). In contrast, only 37% and 58% of patients, respectively, selected these levels. This effect was most extreme for usual activities, where 76% of the general population reported having no problems doing their usual activities, compared to just 14% of patients.

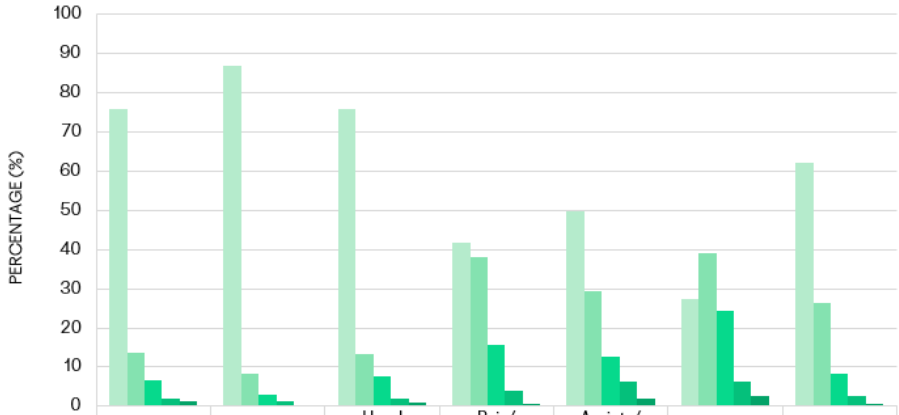
Non-observable health domains also revealed sharp contrasts between the two samples. While 42% of the general population reported having no pain or discomfort, this was only the case for 9% of the patient sample. Furthermore, 50% of the general population reported that they are not anxious or depressed, whereas this was only true for 10% of the patient sample.

FIGURE 3. PATIENT RESPONSES TO THE EQ-5D-5L AND TWO BOLT-ON ITEMS (%)



	Mobility	Self-Care	Usual Activities	Pain/ Discomfort	Anxiety/ Depression	Tiredness	Cognition
Level 1 (No problems)	37	58	14	9	10	3	44
Level 2	24	26	37	41	25	30	33
Level 3	27	12	33	38	34	38	18
Level 4	11	4	15	10	19	23	5
Level 5 (Extreme Problems)	1	0	2	2	12	7	1

FIGURE 4. GENERAL POPULATION RESPONSES TO THE EQ-5D-5L AND TWO BOLT-ON ITEMS (%)



	Mobility	Self-Care	Usual Activities	Pain/ Discomfort	Anxiety/ Depression	Tiredness	Cognition
Level 1 (No problems)	76	87	76	42	50	27	62
Level 2	14	8	14	38	29	39	26
Level 3	7	3	8	16	13	25	8
Level 4	2	1	2	4	6	6	3
Level 5 (Extreme Problems)	1	0	1	1	2	3	1

Differences were also observed with the bolt-on items. Over two thirds (68%) of patients reported experiencing moderate, or worse, tiredness. However, this was only the case for 34% of the general population sample. The responses to the cognition bolt-on were the most similar between the two samples but still differed. For example, 62% of the general population reported having no problems with cognition, compared to 44% of the patient sample.

Severe or extreme problems (levels 4 and 5) were relatively uncommon in both groups, but their prevalence was consistently higher among patients, particularly for anxiety/depression and tiredness. Overall, the dimension data confirm that the patient cohort experience a substantially greater symptomatic and psychosocial burden than the general population, with both tiredness and cognition resembling distinctive gaps not captured by the core five EQ-5D-5L dimensions.

The proportions of patients reporting problems on each dimension, split by leukemia type, can be found in Appendix A.2.

1.2.2 Combined scores

Table 4 provides the summary scores across the two samples, as well as patient scores by leukemia type. The general population sample has better mean scores on all metrics compared to the patient sample (higher EQ VAS, lower LSS, and higher utility). Amongst the patient sample, mean scores are very similar between patients with ALL and AML on all metrics. However, patients with APL have better scores on average compared to patients with ALL and AML, and a higher mean EQ VAS score than the general population average (73.5 vs. 72.8, respectively). Summary scores for patients, split by country, can be found in Appendix A3.

TABLE 4. EQ-5D-5L SUMMARY SCORES ACROSS SAMPLES AND BY LEUKEMIA TYPE

	EQ VAS	LEVEL SUM SCORE (LSS)	UTILITY (US VALUE SET)
General Population (n=511)	72.8 (20.3)	7.6 (3.0)	0.81 (0.23)
All Patients (n=212)	60.3 (22.4)	11.8 (3.9)	0.52 (0.33)
- Patients with ALL (n=62)	61.8 (22.9)	11.9 (4.1)	0.49 (0.37)
- Patients with AML (n=133)	57.8 (22.7)	12.0 (3.8)	0.51 (0.31)
- Patients with APL (n=17)	73.5 (11.6)	10.4 (4.1)	0.64 (0.31)

Mean scores are reported, with standard deviations in parentheses.

Figures 5-7 illustrate the distributions of each summary score (EQ VAS, LSS, and utility, respectively) for each sample (patients and general population). While the general population EQ VAS scores peak at around 80 to 90, Figure 5 illustrates that there is a wider spread of scores for patients, with a significant proportion below 50. In terms of LSS, Figure 6 shows that while the general population distribution is heavily right skewed, there is a more symmetric distribution for patients and a lot more variability. LSS above 15 (indicating an average of level three, or "moderate problems", on all dimensions) are very rare in the general population sample, but not uncommon in the patient sample.

FIGURE 5. EQ VAS SCORES FOR THE PATIENT AND GENERAL POPULATION SAMPLES

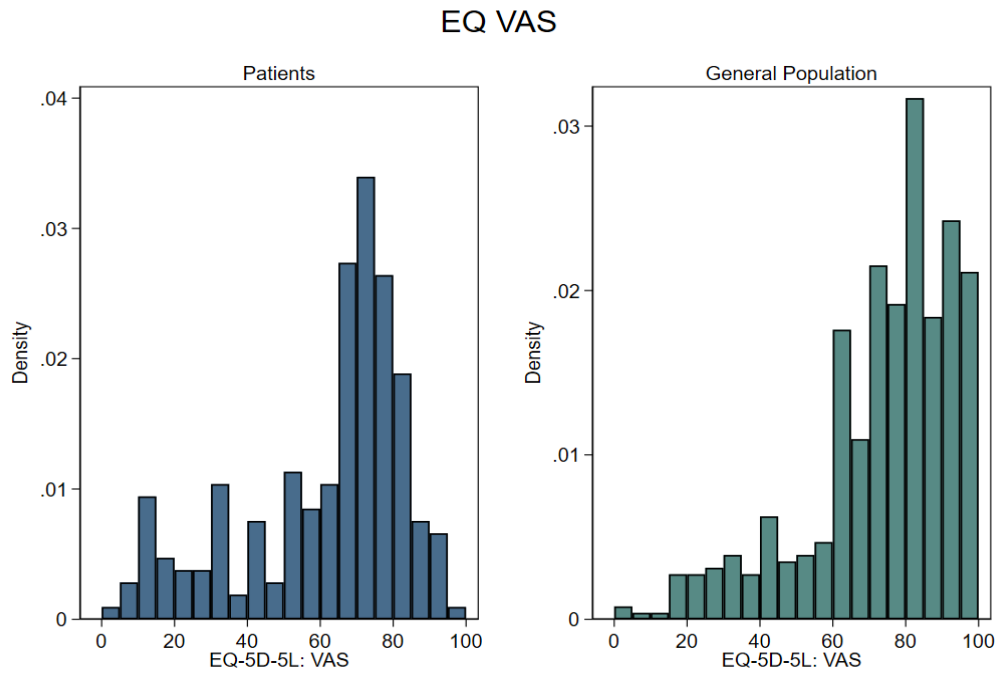


FIGURE 6. EQ-5D-5L LSS FOR THE PATIENT AND GENERAL POPULATION SAMPLES

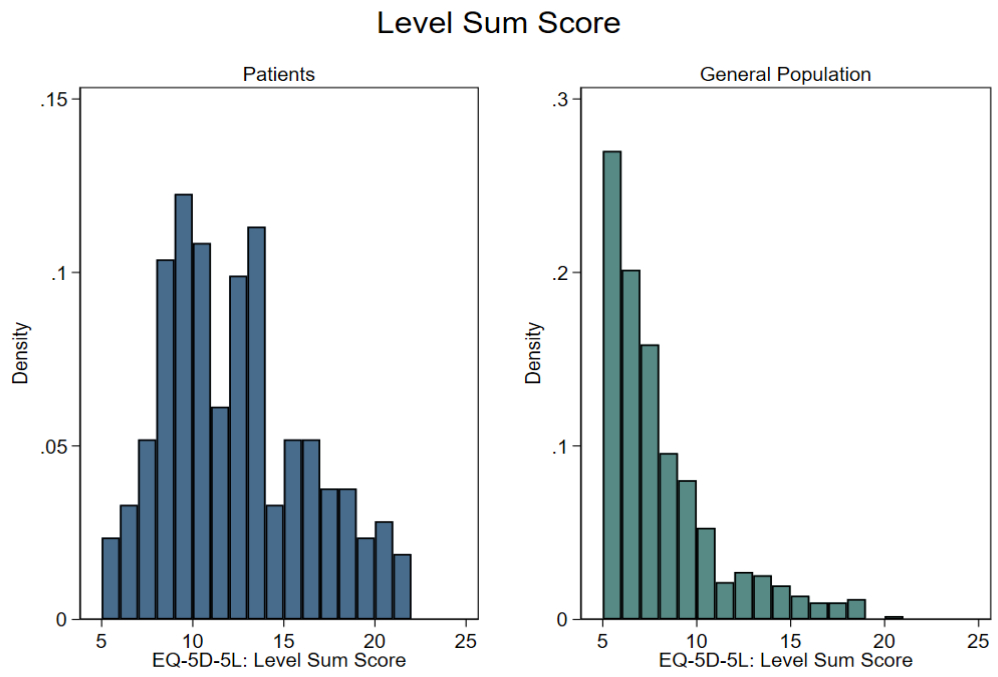
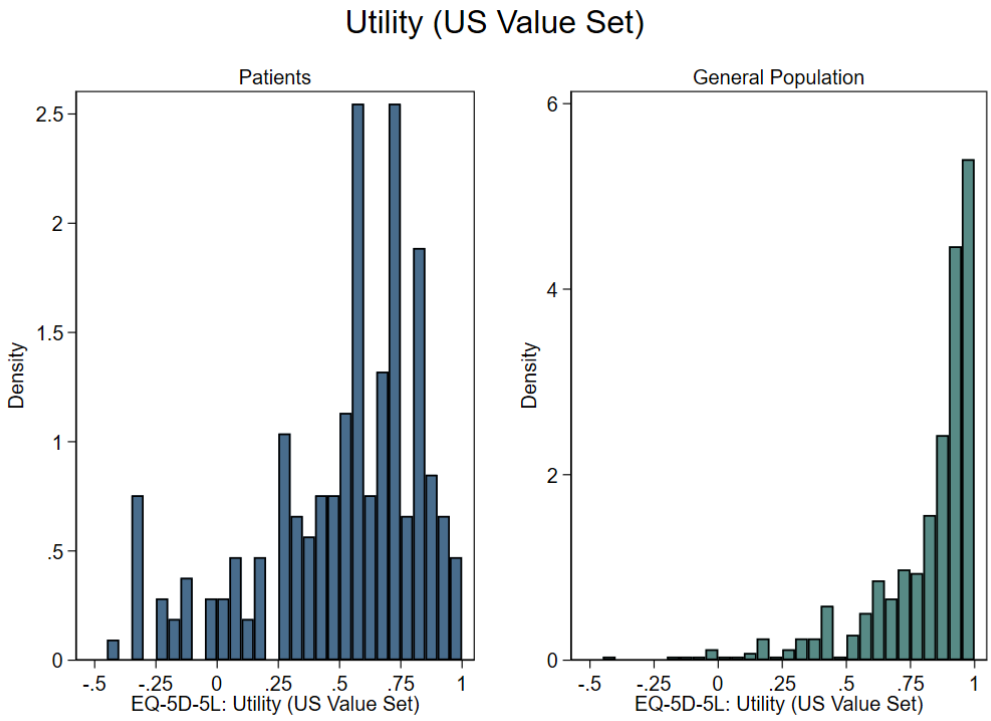


Figure 7 shows that utilities are clustered around 1 (full health) for the general population sample and strongly left skewed. In contrast there is a flatter distribution across the patient sample, with a significant number of respondents reporting utilities below 0.25, including several below 0 (indicating a self-reported health state that has been valued as being worse than dead).

FIGURE 7. EQ-5D-5L UTILITIES FOR THE PATIENT AND GENERAL POPULATION SAMPLES



3.3 Determinants of patient HRQoL

TABLE 5. REGRESSION ANALYSIS TO EXPLORE THE DETERMINANT OF PATIENT HRQOL

	LEVEL SUM SCORE			UTILITY (US VALUE SET)		
Independent variables	Coefficient	Std. Err.	P-value	Coefficient	Std. Err.	P-value
Age	0.032	0.019	0.100	-0.002	0.002	0.263
Female (reference = male)	-0.255	0.517	0.622	0.031	0.044	0.489
AML (reference = ALL)	0.291	0.560	0.604	0.012	0.048	0.811
APL (reference = ALL)	-1.626	0.981	0.099	0.161	0.084	0.057
Years since diagnosis	-0.353	0.071	0.000	0.028	0.006	0.000
Experienced a relapse	1.705	0.529	0.001	-0.136	0.045	0.003
Constant	10.752	1.175	0.000	0.536	0.100	0.000
R^2		0.196			0.168	

Higher level sum scores = worse HRQoL. Lower utilities = worse HRQoL.

The results from the regression analyses to explore the factors that influence patient HRQoL can be found in Table 5. The LSS model suggests that older patients have higher LSS (worse HRQoL) than younger patients, though the effect is small and only significant at the 10% level. Being female (vs. being male) and having AML (vs. ALL) did not have a statistically significant impact on LSS. In contrast, having APL (vs. ALL) was associated with a 1.6 lower LSS on average (better HRQoL). A longer disease duration (as indicated by the number of years since diagnosis) was also associated with lower LSS (better HRQoL) on average. This effect was quite substantial; for example, a disease duration of 12 years is associated with a 3.5 lower LSS on average compared to a disease duration of 2 years. Finally, experiencing a relapse in the past was associated with a 1.7 higher LSS (worse HRQoL) on average compared to those that had not experienced a relapse.

The utility model has similar patterns, though the coefficients are in the opposite direction due to the nature of the dependent variables. For example, experiencing a relapse in the past was associated with a 0.136 lower utility (worse HRQoL) on average compared to those that had not experienced a relapse.

3.4 BWS-7D results

3.4.1 Counting approach

The results of the count analysis for the patient samples can be found in Table 6. Overall, the dimension level most selected as 'worst' was PD5 (n=412; extreme pain or discomfort), followed by CO5 (n=262; extreme problems with cognition), MO5 (n=249; unable to walk about), AD5 (n=206; extreme anxiety or depression), and TI5 (n=183; extremely tired). The dimension level most selected as 'best' was PD1 (n=355; no pain or discomfort), followed by UA1 (n=279; no problems doing my usual activities), MO1 (n=252; no problems in walking about), CO1 (n=248; no problems with cognition) and TI1 (n=248; not tired).

An equivalent table for the general population samples can be found in Appendix A4. Table 7 presents a comparison of standardized best-worst scores between the patient samples and the general population samples. Due to slight differences in the mix of health states in the BWS exercise between the two samples (due to blocking), direct comparisons should be taken with some caution. However, the results suggest some notable differences between the two samples. Whilst the samples agreed that the worst aspects of the health states were PD5 ("extreme pain or discomfort") and CO5 ("extreme problems with cognition"), the ranking differed between the two samples: PD5 was worst for patients; CO5 was worst for the general population. However, whilst the difference between CO5 and PD5 was relatively substantial in the patient sample, it was very small in the general population sample.

TABLE 6. COUNT ANALYSIS FOR THE PATIENT SAMPLES

DIMENSION LEVEL	CHOSEN AS BEST (N) [*A*]	CHOSEN AS WORST (N) [*B*]	BEST-WORST SCORE [*A-B*]	APPEARANCES (N) [*C*]	STANDARDISED BEST-WORST SCORE [*(A-B)/C*]
Mobility					
Level 1 (MO1)	252	24	228	696	0.328
Level 3 (MO3)	12	76	-64	696	-0.092
Level 5 (MO5)	18	249	-231	716	-0.323
Self-Care					
Level 1 (SC1)	163	11	152	696	0.218
Level 3 (SC3)	26	23	3	696	0.004
Level 5 (SC5)	37	106	-69	716	-0.096
Usual Activities					
Level 1 (UA1)	279	18	261	696	0.375
Level 3 (UA3)	25	51	-26	696	-0.037
Level 5 (UA5)	39	166	-127	716	-0.177
Pain/Discomfort					
Level 1 (PD1)	355	16	339	696	0.487
Level 3 (PD3)	19	98	-79	696	-0.114
Level 5 (PD5)	4	412	-408	716	-0.570
Anxiety/Depression					
Level 1 (AD1)	236	13	223	696	0.320
Level 3 (AD3)	21	53	-32	696	-0.046
Level 5 (AD5)	16	206	-190	716	-0.265
Tiredness					
Level 1 (TI1)	248	16	232	696	0.333
Level 3 (TI3)	53	32	21	696	0.030
Level 5 (TI5)	21	183	-162	716	-0.226
Cognition					
Level 1 (CO1)	248	8	240	696	0.345
Level 3 (CO3)	26	85	-59	696	-0.085
Level 5 (CO5)	10	262	-252	716	-0.352

TABLE 7. STANDARDISED BEST-WORST SCORES COMPARING THE PATIENT AND GENERAL POPULATION SAMPLES

DIMENSION LEVEL	STANDARDISED BEST-WORST SCORES		
	PATIENTS [*A*]	GENERAL POPULATION [*B*]	DIFFERENCE [*A-B*]
Mobility			
Level 1 (MO1)	0.328	0.395	-0.067
Level 3 (MO3)	-0.092	0.000	-0.092
Level 5 (MO5)	-0.323	-0.198	-0.125
Self-Care			
Level 1 (SC1)	0.218	0.266	-0.047
Level 3 (SC3)	0.004	-0.013	0.017
Level 5 (SC5)	-0.096	-0.150	0.054
Usual Activities			
Level 1 (UA1)	0.375	0.369	0.006
Level 3 (UA3)	-0.037	-0.003	-0.034
Level 5 (UA5)	-0.177	-0.138	-0.039
Pain/Discomfort			
Level 1 (PD1)	0.487	0.308	0.179
Level 3 (PD3)	-0.114	-0.051	-0.063
Level 5 (PD5)	-0.570	-0.328	-0.241
Anxiety/Depression			
Level 1 (AD1)	0.320	0.128	0.193
Level 3 (AD3)	-0.046	-0.045	-0.001
Level 5 (AD5)	-0.265	-0.211	-0.055
Tiredness			
Level 1 (TI1)	0.333	0.102	0.231
Level 3 (TI3)	0.030	0.000	0.030
Level 5 (TI5)	-0.226	-0.088	-0.138
Cognition			
Level 1 (CO1)	0.345	0.235	0.110
Level 3 (CO3)	-0.085	-0.082	-0.003
Level 5 (CO5)	-0.352	-0.334	-0.018

There were also differences in relation to the best dimension levels. While the best dimension level was PD1 for patients, the best dimension level for the general population was MO1, and PD1 was ranked third best (after UA1). Similar to the worst dimension levels, the differences in scores for the best levels are greater in the patient sample, indicating stronger preferences in the patient sample towards the best dimension level (PD1).

3.4.2 Modelling BWS choices

The multinomial logit model results for the BWS-7D task for both samples can be found in Table 8. Across both samples there is a logical ordering of preferences within each dimension i.e., extreme problems are worse than moderate problems in each dimension. The coefficients on the level three variables are consistently large and negative, and in both samples PD5 (extreme pain or discomfort) has the largest coefficient in absolute terms, indicating the pain/discomfort is the most important dimension.

TABLE 8. MULTINOMIAL LOGIT MODEL RESULTS FOR THE BWS DATA

	PATIENTS				GENERAL POPULATION			
	Coeff	Std Err	Z Score	p-value	Coeff	Std Err	Z Score	p-value
MO3	-0.555	0.067	-8.264	<0.001	-0.272	0.038	-7.247	<0.001
MO5	-0.962	0.066	-14.581	<0.001	-1.266	0.035	-35.980	<0.001
SC3	-0.194	0.071	-2.747	0.006	-0.191	0.038	-5.018	<0.001
SC5	-0.638	0.069	-9.247	<0.001	-1.001	0.036	-27.705	<0.001
UA3	-0.349	0.070	-4.956	<0.002	-0.348	0.038	-9.256	<0.001
UA5	-0.867	0.069	-12.615	<0.001	-1.081	0.036	-30.304	<0.001
PD3	-0.636	0.066	-9.661	<0.001	-0.088	0.037	-2.383	0.017
PD5	-1.506	0.063	-23.810	<0.001	-1.474	0.035	-42.428	<0.001
AD3	-0.108	0.070	-1.549	0.121	0.031	0.038	0.809	0.419
AD5	-1.148	0.067	-17.230	<0.001	-0.968	0.036	-26.681	<0.001
TI3	-0.121	0.071	-1.705	0.088	0.003	0.039	0.083	0.934
TI5	-0.902	0.069	-13.008	<0.001	-0.623	0.037	-16.723	<0.001
CO3	-0.254	0.068	-3.716	<0.001	-0.055	0.037	-1.482	0.139
CO5	-1.412	0.065	-21.833	<0.001	-1.316	0.035	-38.046	<0.001

Coeff = coefficient; Std Err = standard error. All variables are effects coded.

Relative importance scores for the different dimensions for the patient sample, which were derived from the coefficients of the MNL model in Table 8, can be found in Figure 9. Overall, for patients, pain/discomfort was the most important dimension (RI=21.4%). While the count analysis results in the previous section suggested that the cognition (the second most important dimension) was considerably less important than pain/discomfort, the modelling results suggest that cognition is of similar importance (RI=18.0%). The third most important dimension was mobility (RI=14.5%), followed by

anxiety/depression (RI=14.1%), usual activities (RI=12.2%) and tiredness (RI=11.3%). The least important dimension was self-care (RI=8.6%).

FIGURE 9. RI SCORES FOR THE PATIENT AND GENERAL POPULATION SAMPLES

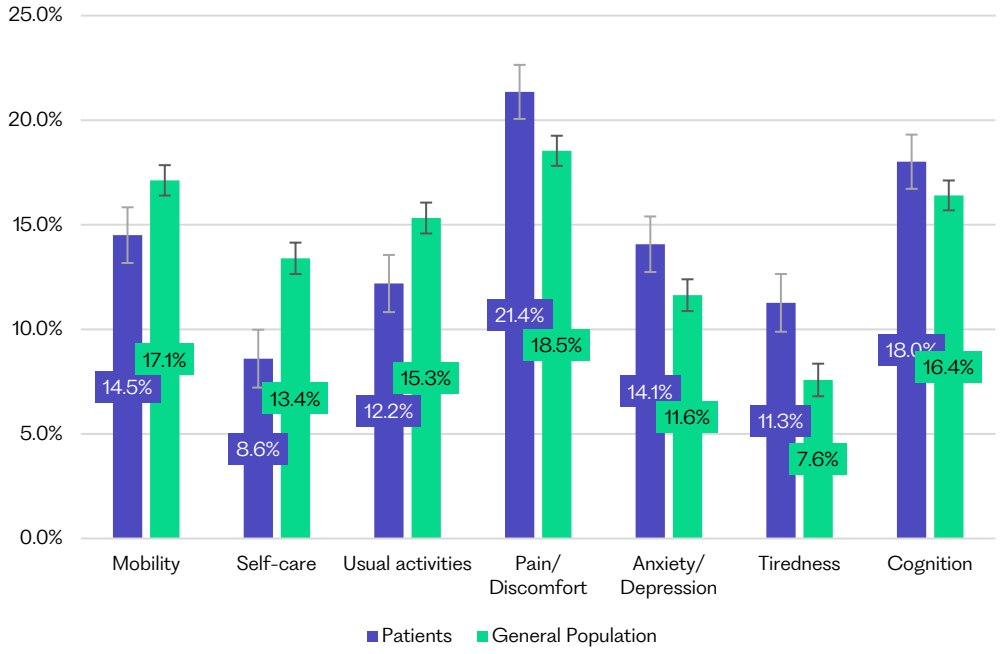


Figure 9 also provides the results for the general population samples alongside the results for the patient samples. There are some notable differences in scores between the two samples, the largest of which relate to self-care (RI=13.4% for the general population compared to 8.6% for patients), tiredness (RI=7.6% for the general population compared to 11.3% for patients), and usual activities (RI=15.3% for the general population compared to 12.2% for patients). The confidence intervals only overlap for cognition, suggesting that there are statistically significant differences in RI scores for all other dimensions when comparing the results from the two samples.

TABLE 9. DIMENSION RANK ORDERINGS IMPLIED BY THE RI SCORES

	PATIENTS	GENERAL POPULATION
1. Most important	Pain/discomfort	Pain/discomfort
2.	Cognition	Mobility
3.	Mobility	Cognition
4.	Anxiety/depression	Usual activities
5.	Usual activities	Self-care
6.	Tiredness	Anxiety/depression
7. Least Important	Self-care	Tiredness

Table 9 provides a summary of the implied rank order of the dimensions from the RI scores. In both samples, pain/discomfort is the most important dimension. The next two are cognition and mobility in both samples, though the rank order differs by sample (cognition second for patients, third for general population). For the final four dimensions there is considerably more variability. Patients rank self-care as the least important dimension, whereas the general population rank self-care as the fifth most important, and tiredness as the least important dimension.

4. Discussion

4.1 Summary of findings

4.1.1 Self-reported HRQoL

The EQ-5D-5L results, including the tiredness and cognition bolt-ons reveal a consistent and substantial HRQoL gap between people living with acute leukemia and the general population. Across every core EQ-5D-5L dimension, patients reported substantially greater levels of impairment than the general population, with the greatest differences observed in physical functioning. For example, 76% of general population reported having no problems with mobility compared with just 37% of patients, and 87% reported full independence in relation to self-care versus 58% of patients. Usual activities followed the same pattern (76% vs. 14% reporting having no problems).

Non-observable health domains also showed marked disparities. Being pain-free was reported by 42% of the general population compared with only 9% of patients, and moderate-to-severe anxiety/depression (level 3+) was more than three times as common in patients (65% vs. 21%). The bolt-on items further highlighted unmet needs — particularly tiredness, with only 3% of patients reporting no tiredness, compared with over a quarter (27%) of the general population sample. Moderate tiredness was common (around 40% of patients vs. 25% of general population) and prevalent across treatment stages, confirming tiredness as a major and burdensome symptom in this population. Self-reported problems with cognition were also more common in patients (57% reported some impairment vs. 38% in the general population).

Severe or extreme problems (levels 4-5) were relatively uncommon in both groups, but were consistently more frequent among patients, particularly for anxiety/depression and tiredness. This pattern suggests that while severe impairments are rare, moderate problems are more widespread and concentrated in symptom-driven domains.

Summary metrics reinforced these findings. Mean EQ VAS scores were lower for patients than for the general population (60.3 vs. 72.8). LSS scores were higher (11.8 vs. 7.6) and utilities (U.S. value set applied) were markedly lower

(0.52 vs. 0.81). Within the patient sample, ALL and AML groups had similar mean scores, whereas APL patients reported better outcomes — mean EQ VAS scores were higher than even general population average (73.5 vs. 72.8) and mean utility was higher than for other subtypes (0.64 vs. 0.49-0.51).

Considered together, these findings show that the HRQoL burden in acute leukemia is substantial and persistent, with tiredness and cognition emerging as key dimensions not fully captured by the EQ-5D-5L core set.

4.1.2 Preferences for different aspects of health

The BWS results show that all seven of the health outcomes included in the exercise are important to patients. However, patients' greatest concern overall relates to pain/discomfort. The dimension level PD5 (extreme pain or discomfort) was the most selected 'worst' option in the BWS exercise, and PD1 (no pain or discomfort) was the most selected 'best' option. The next most selected 'best' and 'worst' options were selected far less frequently in the exercises in comparison. Furthermore, based on the modelling results, pain/discomfort was the most important dimension overall (RI=21.4%), followed by cognition (RI=18.0%). In contrast, the least important dimension was self-care (RI=8.6%).

There were some notable differences between the preferences of the patient and general population samples. Although pain/discomfort was the worst dimension for both samples, there was relatively less emphasis on this dimension in the BWS exercise for the general population compared to patients based on the count analysis. Based on the relative importance scores from the modelling analysis, there were some notable differences in the relative importance of different dimensions, which resulted in different rank orderings. Self-care and usual activities were more important to the general population sample compared to the patient sample (differences of 4.8pp and 3.1pp respectively). In contrast, tiredness was more important to the patient sample compared to the general population sample (difference of 3.7pp).

4.2 Interpretation and comparison with other literature

4.2.1 Self-reported HRQoL

In our study, people with acute leukemia frequently reported problems on the core dimensions of EQ-5D-5L, as well as on the tiredness and cognition bolt-on items, reflecting the wide-ranging impacts of the disease. The mean utility of the patient sample is low (0.52) and the average in our general population sample was far higher (0.81). That said, the general population sample were older on average than the patient sample. Nonetheless, population norms (average values in the general population) in Western countries rarely dip below 0.75, and where this occurs, it only occurs in the oldest age category (75+) (Janssen and Szende, 2014). As such, our data would suggest that people with acute leukemia have considerably worse HRQoL, on average, compared to people in the general population.

There is relatively little EQ-5D-5L data available for people with acute leukemia to enable comparisons with our study. A 2012 conference presentation provided some data from n=86 people with acute leukemia, reporting a mean utility of 0.82 using EQ-5D-5L (Leunis et al., 2012). More recently, a longitudinal pilot study collected EQ-5D-5L data in older adults with AML and reported mean utilities between 0.69 and 0.75 (LoCastro et al., 2023). However, the sample size was only n=11 at baseline and decreased over time. The mean utility in our study is considerably lower than all these estimates, and our sample size far exceeds those from these earlier studies.

While the study by Salek et al. (2025) used a non-preference-based condition-specific measure, HM-PRO (Hematological Malignancy Patient Reported Outcomes), rather than EQ-5D-5L, we tested several hypotheses based on their research with our data. Like Salek et al., we found that a longer disease duration was associated with better health outcomes. However, we did not find significant differences in HRQoL between ALL and AML patients, nor did we find a significant association with age or gender. However, their sample sizes were significantly larger than ours (n=403 vs n=212). Salek et al. (2025) did not look at APL patients specifically, whereas in our study we found that people with APL had better HRQoL compared to people with ALL or AML. However, given that this subgroup consisted of only 17 patients, this finding should be taken with caution. Furthermore, unlike Salek et al. (2025), we were

able to explore the potential impact of individuals having experienced a relapse in the past on HRQoL, and found that this has a reasonably large and statistically significant negative impact on HRQoL, as would be expected.

4.2.2 Preferences for different aspects of health

While there have been several studies that sought to elicit preferences from people with acute leukemia, most have focused on a treatment decision-making context, and none have focused exclusively on preferences for different health outcomes (LoCastro et al., 2023; Mott et al., 2024; Richardson et al., 2021, 2020; Saini et al., 2023). Our study has shown that a variety of health dimensions are important to people with acute leukemia, with pain/discomfort, mobility, and cognition the three most important of those included in our BWS exercise.

In Western countries, pain/discomfort is often the most important dimension in EQ-5D-5L value sets (Roudijk, Janssen and Olsen, 2022), which are based on general population preferences. In this study, pain/discomfort was not only the most important dimension in the general population sample, but also the patient sample (and indeed was even more important in the latter). In contrast, anxiety/depression, which is often in the top two or three dimensions, was ranked fifth (out of the core five dimensions) in the general population sample in this study. While the latter is unexpected, these comparisons should be taken with caution. Any differences between our results and existing value sets may be due to differences in sampling, changes in preferences over time, and/or differences in methodology.

In our study, the biggest differences between patients and the general population were in relation to self-care and tiredness. Patients considered self-care to be far less important compared to the general population (8.6% and rank 7/7 vs. 13.4% and rank 5/7, respectively). This may be due to experience; people with acute leukemia have already been through an experience whereby their ability to look after themselves will have been significantly impacted. In contrast, for the general population, this is more likely to be hypothetical. The self-reported HRQoL scores provide some evidence for this: 42% of patients reported some issues with self-care, compared to only 13% of the general population.

In contrast, the general population considered tiredness to be far less important compared to patients (7.6% and rank 7/7 vs. 11.3% and rank 6/7, respectively). Although the general population sample was not unfamiliar with tiredness — 72.6% reported experiencing some level of tiredness — this dimension did not appear as important as the others in the BWS exercise. People with acute leukemia experience fatigue as both a symptom of the disease and as a side effect of treatment, and as such it is not surprising that this dimension would have had more weight in this group.

Relatively few studies have elicited *patient* preferences for EQ-5D health states, and none have done so using profile-case BWS or in the context of acute leukemia specifically. Ludwig et al. (2021) elicited preferences from German and Spanish patients with rheumatism and diabetes mellitus using a DCE. Whilst they also found differences in preferences between patients and the general population (based on existing EQ-5D-5L value sets), the patterns that they observed are different to those observed in this study. Ogorevc et al. (2019) elicited preferences from Spanish patients with metastatic breast cancer and rheumatoid arthritis using TTO and DCE. They found that, compared to the general population, patients valued mobility and self-care as less important, and pain/discomfort and anxiety/depression as more important. All of these effects were observed in our study, based on the relative importance scores from the BWS exercise.

4.3 Implications of the results

The results of this study have a range of potential implications. We utilized two bolt-on items for EQ-5D, both of which are potential candidates for inclusion in the forthcoming "EQ-5D Bolt-on Toolbox". Our results show that a meaningful proportion of people with acute leukemia, as well as people in the general population, report experiencing problems in relation to tiredness and cognition. As such, provided these additional items do not overlap with the core five dimensions or each other (which should be determined with robust psychometric studies), the inclusion of these dimensions may increase the sensitivity of EQ-5D to the experience of persons with acute leukemia as well as a wide range of other conditions that may be associated with tiredness and cognition problems.

Furthermore, our BWS results show that both tiredness and cognition are considered important to both patients and the general population when people consider a range of health problems. However, the relative importance of these two dimensions varied significantly. Cognition was nearly as important as pain/discomfort in the two samples and was ranked as the second and third most important dimension in the patient and general population samples, respectively. The precise wording of a future cognition bolt-on item may vary from the one used in this study. However, if future valuation studies continue to find that cognition (however it is described) is valued higher than most other core dimensions, the implications of its inclusion in future economic evaluations may be substantial. In contrast, tiredness was considerably less important than cognition. Nonetheless, tiredness was still considered more important to patients than one of the core dimensions (self-care), and although it was the least important dimension for the general population, it was still important overall. Furthermore, it may be the case that the prosaic label of "tiredness" in the BWS exercises led general population respondents to underestimate its potential impact, relative to more visceral labels such as "pain" and "anxiety". Had the wording of this bolt-on used more clinical terminology (e.g., fatigue), it may have come through as more important in the BWS exercise.

The results of this study also contribute to the relatively sparse literature comparing the relative importance of the dimensions of EQ-5D for patient and general population samples. While the patterns of differences vary, it is becoming increasingly clear that preferences differ between patient and general population samples. The variety of patterns observed in the literature may be explained by a range of factors, including valuation methodology. However, if patients' preferences are informed by their experiences, it is plausible that different patient groups will express different preferences in stated preference exercises. To that end, it is notable that our results better reflect those from a study that included metastatic cancer patients (Ogorevc et al., 2019), compared to a study that included only patients with chronic diseases (Ludwig et al., 2021).

4.4 Limitations and areas for future research

This is the first multinational study that sought to explore the preferences of people with acute leukemia for different health outcomes using a stated preference methodology. Alongside the BWS exercise, the study also collected EQ-5D-5L data (including two additional items), and incorporated general population samples, to generate policy-relevant data and to enable comparisons to be made.

However, the study is not without its limitations. Firstly, the sample sizes are relatively small. Acute leukemia is a rare and severe disease, and it is challenging and expensive population to recruit in research studies as a result. Profile-case BWS is relatively statistically efficient for examining the relative importance of different dimensions because each task produces two data points (a best and worst option), in contrast to other stated preference methodologies such as discrete choice experiments. Nonetheless, the relatively small sample sizes mean that our ability to explore preference heterogeneity (e.g., preferences by subgroups) is limited.

A second limitation, which exacerbates the first, is that we did not employ any quotas during recruitment. The aim of this was to avoid situations where people with acute leukemia were prevented from completing the survey. However, it means that in some countries the sample composition is very different to others. For example, in the EU3, 85% of patients had AML, whereas in the US only 39% of patients had AML.

Future research into the preferences of people with acute leukemia would be valuable to validate these findings, as well as findings from the broader literature around treatment preferences. However, given the recruitment challenges, it may be advisable to focus on methods that require relatively minimal sample sizes, and to recruit directly through hospitals and other health facilities, if possible, as per LoCastro et al. (2023).

Furthermore, based on the results of this study, further research to develop the cognition and tiredness bolt-ons for EQ-5D would certainly appear to be warranted. In relation to valuation of these bolt-ons specifically, it would be useful to explore whether these dimensions are equally important when alternative methods that incorporate trade-offs with death, such as TTO, are used. Additionally, it would be informative to see how different wordings for the

tiredness bolt-on, especially any that use the term 'fatigue', have an impact on (a) the extent of self-reported issues from general population samples and (b) the relative importance of this dimension in valuation tasks.

5. Conclusion

This study provides further insights into the HRQoL impacts faced by people with acute leukemia, and how well current tools capture the outcomes that matter most. We found that patients' self-reported HRQoL is significantly worse than that of the general population. The addition of tiredness and cognition bolt-on items helped highlight unmet needs, particularly around tiredness, which stood out as both a common and burdensome issue.

Our preference data illustrates the importance of different dimensions of HRQoL to people with acute leukemia, with pain/discomfort and cognition ranked highest in importance. Patients' preferences for different health outcomes differed to those of the general population, suggesting that health care decisions might differ if value sets reflected patient preferences.

Our findings have important implications for future research and practice. There is a case for augmenting generic tools like the EQ-5D-5L with additional items that reflect key symptoms in specific populations. Furthermore, while the use of value sets based on general population preferences is commonplace, these value sets may not accurately reflect the priorities of patients. This highlights the importance of examining patients' preferences in order to make better and more informed healthcare decisions.

References

Andrés-Jensen, L., Nielsen, C.G., van den Heuvel-Eibrink, M.M. and Schmiegelow, K., 2024. Acute Toxicity and Late Effects Related to Acute Lymphoblastic Leukemia Treatment. In: G. Escherich and V. Conter, eds, *Acute Lymphoblastic Leukemia in Children and Adolescents*. [online] Cham: Springer Nature Switzerland, pp.279—303. DOI: 10.1007/978-3-031-71180-0_18.

Buchholz, I., Janssen, M.F., Kohlmann, T. and Feng, Y.-S., 2018. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. *PharmacoEconomics*, 36(6), pp.645—661. DOI: 10.1007/s40273-018-0642-5.

Devlin, N.J., Xie, F., Slaap, B. and Stolk, E., 2025. Measuring and Valuing Health Using EuroQol Instruments: New Developments 2025 and Beyond. *Applied Health Economics and Health Policy*. [online] DOI: 10.1007/s40258-025-00989-2.

van Dongen-Leunis, A., Redekop, W.K. and Uyl-de Groot, C.A., 2016. Which Questionnaire Should Be Used to Measure Quality-of-Life Utilities in Patients with Acute Leukemia? An Evaluation of the Validity and Interpretability of the EQ-5D-5L and Preference-Based Questionnaires Derived from the EORTC QLQ-C30. *Value in Health*, 19(6), pp.834—843. DOI: 10.1016/j.jval.2016.05.008.

Fayers, P.M. and Machin, D., 2013. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes*. John Wiley & Sons.

Flynn, T.N., 2010. Valuing citizen and patient preferences in health: recent developments in three types of best—worst scaling. *Expert Review of Pharmacoeconomics & Outcomes Research*, 10(3), pp.259—267. DOI: 10.1586/erp.10.29.

Flynn, T.N., Louviere, J.J., Peters, T.J. and Coast, J., 2007. Best--worst scaling: what it can do for health care research and how to do it?'. *Journal of Health Economics*, 26, pp.171—89. DOI: 10.1016/j.jhealeco.2006.04.002.

Gandhi, M., Tan, R.S., Ng, R., Choo, S.P., Chia, W.K., Toh, C.K., Lam, C., Lee, P.T., Latt, N.K.Z., Rand-Hendriksen, K., Cheung, Y.B. and Luo, N., 2017. Comparison of health state values derived from patients and individuals from the general population. *Quality of Life Research*, 26(12), pp.3353—3363. DOI: 10.1007/s11136-017-1683-5.

Janssen, B. and Szende, A., 2014. Population Norms for the EQ-5D. In: A. Szende, B. Janssen and J. Cabases, eds, *Self-Reported Population Health: An International Perspective based on EQ-5D*. [online] Dordrecht: Springer Netherlands, pp.19—30. DOI: 10.1007/978-94-007-7596-1_3.

Leunis, A., Redekop, W.K., Lowenberg, B. and Uyl-de Groot, C., 2012. UT1 The Calculation of Quality of Life Utilities for Acute Leukemia: A Comparison Between EQ5D-5L and QlQ-C30. *Value in Health*, 15(7), p.A284. DOI: 10.1016/j.jval.2012.08.513.

LoCastro, M., Wang, Y., Sanapala, C., Jensen-Battaglia, M., Wittink, M., Norton, S., Klepin, H.D., Richardson, D.R., Mendler, J.H., Liesveld, J., Huselton, E. and Loh, K.P., 2023. Patient preferences, regret, and health-related quality of life among older adults with acute myeloid leukemia: A pilot longitudinal study. *Journal of Geriatric Oncology*, [online] 14(5). DOI: 10.1016/j.jgo.2023.101529.

Louviere, J.J., Flynn, T.N. and Marley, A.A.J., 2015. *Best-Worst Scaling: Theory, Methods and Applications*. 1st edn. [online] Cambridge University Press. DOI: 10.1017/CBO9781107337855.

Ludwig, K., Ramos-Goñi, J.M., Oppe, M., Kreimeier, S. and Greiner, W., 2021. To What Extent Do Patient Preferences Differ From General Population Preferences? *Value in Health*, 24(9), pp.1343—1349. DOI: 10.1016/j.jval.2021.02.012.

Meyers, C.A., Albitar, M. and Estey, E., 2005. Cognitive impairment, fatigue, and cytokine levels in patients with acute myelogenous leukemia or myelodysplastic syndrome. *Cancer*, 104(4), pp.788—793. DOI: 10.1002/cncr.21234.

Mott, D., Hitch, J., Nier, S., Pemberton-Whiteley, Z. and Skedgel, C., 2024. Patient Preferences for Treatment in Relapsed/Refractory Acute Leukemia in the United Kingdom: A Discrete Choice Experiment. *Patient Preference and Adherence*, Volume 18, pp.1243—1255. DOI: 10.2147/ppa.s442530.

Mott, D., Hussain, H., Chowdhury, S., Nier, S. and Skedgel, C., 2025. *Patient Preferences for Treatment in Relapsed/Refractory Acute Leukemia: A Multinational Study*. [Forthcoming OHE Grant-Funded Research Report] Available at: https://www.ohe.org/. [Accessed 12 June 2025].

NICE, 2022. *NICE health technology evaluations: the manual*. [online] London: National Institute for Health and Care Excellence. Available at: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation. [Accessed 19 Apr. 2023].

Ogorevc, M., Murovec, N., Fernandez, N.B. and Rupel, V.P., 2019. Questioning the differences between general public vs. patient based preferences towards EQ-5D-5L defined hypothetical health states. *Health Policy*, 123(2), pp.166—172. DOI: 10.1016/j.healthpol.2017.03.011.

Peeters, Y. and Stiggelbout, A.M., 2010. Health State Valuations of Patients and the General Public Analytically Compared: A Meta-Analytical Comparison of Patient and Population Health State Utilities. *Value in Health*, 13(2), pp.306—309. DOI: 10.1111/j.1524-4733.2009.00610.x.

Pickard, A.S., Law, E.H., Jiang, R., Pullenayegum, E., Shaw, J.W., Xie, F., Oppe, M., Boye, K.S., Chapman, R.H., Gong, C.L., Balch, A. and Busschbach, J.J.V., 2019. United States Valuation of EQ-5D-5L Health States Using an International Protocol. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 22(8), pp.931—941. DOI: 10.1016/j.jval.2019.02.009.

Ratcliffe, J., Flynn, T., Terlich, F., Stevens, K., Brazier, J. and Sawyer, M., 2012. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *PharmacoEconomics*, 30(8), pp.713—727. DOI: 10.2165/11597900-000000000-00000.

Rencz, F. and Janssen, M.F., 2024. Testing the Psychometric Properties of 9 Bolt-Ons for the EQ-5D-5L in a General Population Sample. *Value in Health*, 27(7), pp.943—954. DOI: 10.1016/j.jval.2024.03.2195.

Richardson, D.R., Crossnohere, N.L., Seo, J., Estey, E., O'Donoghue, B., Smith, B.D. and Bridges, J.F.P., 2020. Age at Diagnosis and Patient Preferences for Treatment Outcomes in AML: A Discrete Choice Experiment to Explore Meaningful Benefits. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the*

American Society of Preventive Oncology, 29(5), pp.942—948. DOI: 10.1158/1055-9965.EPI-19-1277.

Richardson, D.R., Oakes, A.H., Crossnohere, N.L., Rathsmill, G., Reinhart, C., O'Donoghue, B. and Bridges, J.F.P., 2021. Prioritizing the worries of AML patients: Quantifying patient experience using best—worst scaling. *Psycho-Oncology*, 30(7), pp.1104—1111. DOI: 10.1002/pon.5652.

Roudijk, B., Janssen, B. and Olsen, J.A., 2022. How Do EQ-5D-5L Value Sets Differ? In: N. Devlin, B. Roudijk and K. Ludwig, eds, *Value Sets for EQ-5D-5L*. [online] Cham: Springer International Publishing, pp.235—258. DOI: 10.1007/978-3-030-89289-0_6.

Rowen, D., Mukuria, C., Bray, N., Carlton, J., Cooper, S., Longworth, L., Meads, D., O'Neill, C. and Yang, Y., 2023. UK Valuation of EQ-5D-5L, a Generic Measure of Health-Related Quality of Life: A Study Protocol. *Value in Health*, 26(11), pp.1625—1635. DOI: 10.1016/j.jval.2023.08.005.

Saini, L., Griffin, J.D., Pandya, B.J., Shah, M.V., Zhou, M., Yang, H., Song, Y. and Marshall, D.A., 2023. Patient and Physician Preferences for Acute Myeloid Leukemia Maintenance Treatments Following Hematopoietic Stem Cell Transplantation. *Patient Preference and Adherence*, Volume 17, pp.2805—2819. DOI: 10.2147/PPA.S421871.

Salek, S., Nier, S., Pemberton-Whiteley, Z., Ionova, T., Ianni, G., Tripepi, G., York, N., Schroeter, N., Huntley, K., Rynne, M., Costello, D., Pecova, L. and Oliva, E.N., 2025. Impact of leukemia subtype and demographics on patient quality of life in 76 countries: a cross-sectional study. *Frontiers in Hematology*, [online] 3. DOI: 10.3389/frhem.2024.1502166.

Terwilliger, T. and Abdul-Hay, M., 2017. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer Journal*, 7(6), pp.e577—e577. DOI: 10.1038/bcj.2017.53.

Thompson, A.J. and Turner, A.J., 2020. A Comparison of the EQ-5D-3L and EQ-5D-5L. *PharmacoEconomics*, 38(6), pp.575—591. DOI: 10.1007/s40273-020-00893-8.

Tomaszewski, E.L., Fickley, C.E., Maddux, L., Krupnick, R., Bahceci, E., Paty, J. and van Nooten, F., 2016. The Patient Perspective on Living with Acute Myeloid Leukemia. *Oncology and Therapy*, 4(2), pp.225—238. DOI: 10.1007/s40487-016-0029-8.

Versteegh, M.M. and Brouwer, W.B.F., 2016. Patient and general public preferences for health states: A call to reconsider current guidelines. *Social Science & Medicine*, 165, pp.66—74. DOI: 10.1016/j.socscimed.2016.07.043.

Yang, Y., Rowen, D., Brazier, J., Tsuchiya, A., Young, T. and Longworth, L., 2015. An Exploratory Study to Test the Impact on Three "Bolt-On" Items to the EQ-5D. *Value in Health*, 18(1), pp.52—60. DOI: 10.1016/j.jval.2014.09.004.

Appendix

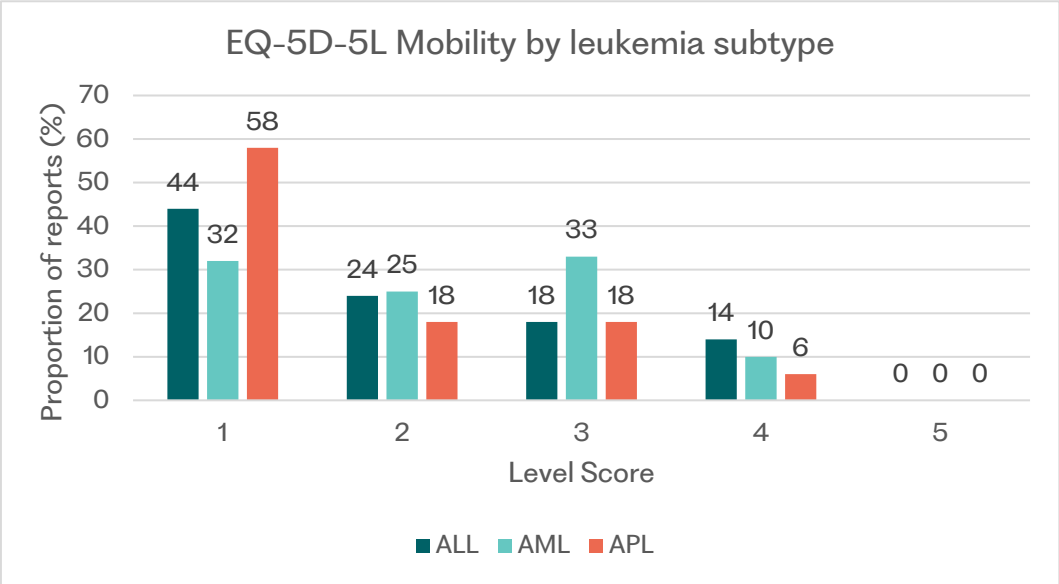
A1. BWS experimental design

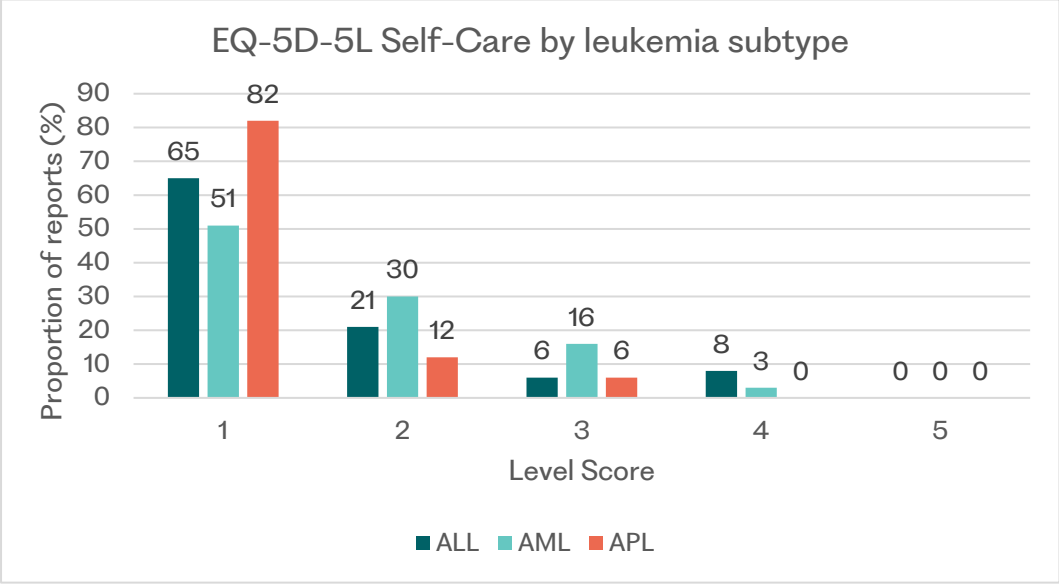
TABLE A1. BWS Experimental Design: OA(27, 7, 3, 2)

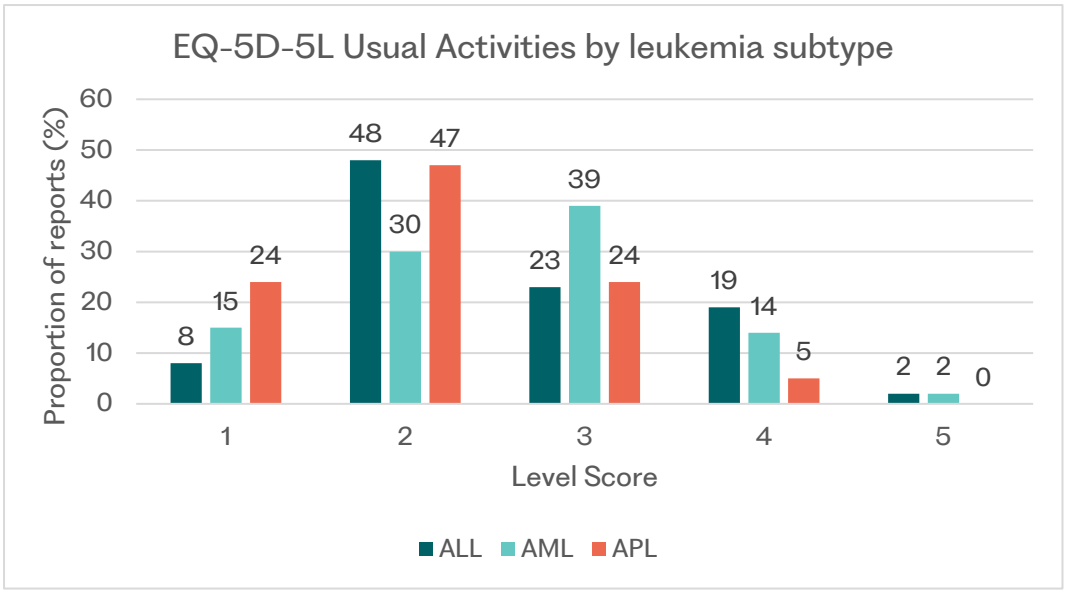
ROW	MO	SC	UA	PD	AD	TI	CO	2-BLOCK	3-BLOCK
1	3	3	3	3	3	3	3	Both	1
2	1	1	1	3	3	3	2	1	1
3	3	3	3	2	2	2	1	1	1
4	2	2	2	1	1	1	3	1	1
5	1	2	3	3	1	2	2	1	1
6	2	3	1	1	2	3	3	1	1
7	2	2	2	2	2	2	2	1	1
8	1	1	1	1	1	1	1	1	1
9	3	1	2	2	3	1	1	1	1
10	3	3	3	1	1	1	2	1	2
11	1	1	1	2	2	2	3	1	2
12	2	2	2	3	3	3	1	1	2
13	3	2	1	1	3	2	2	1	2
14	1	2	3	1	2	3	1	1	2
15	2	3	1	2	3	1	2	2	2
16	1	3	2	2	1	3	3	2	2
17	3	1	2	3	1	2	3	2	2
18	2	1	3	3	2	1	1	2	2
19	2	1	3	1	3	2	3	2	3
20	1	3	2	3	2	1	2	2	3
21	3	1	2	1	2	3	2	2	3
22	2	3	1	3	1	2	1	2	3
23	1	3	2	1	3	2	1	2	3
24	1	2	3	2	3	1	3	2	3
25	3	2	1	3	2	1	3	2	3
26	2	1	3	2	1	3	2	2	3
27	3	2	1	2	1	3	1	2	3

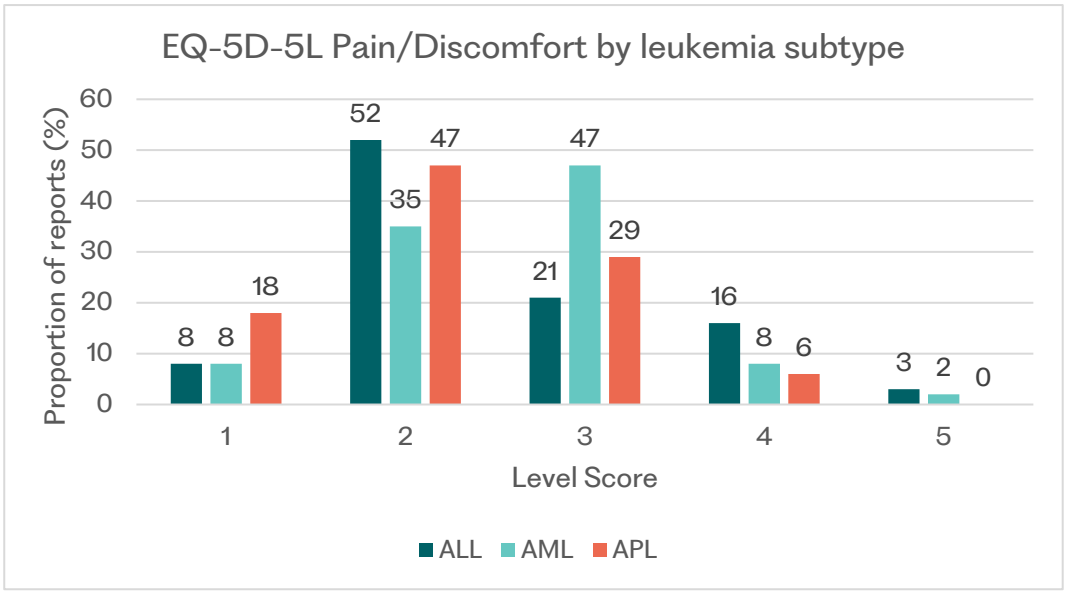
2-block and 3-block refer to the blocks when the design was split into two blocks and three blocks, respectively (see section 2.2.3 for an explanation of how the blocking was applied in practice). TI and CO were removed for the BWS-5D exercise, reducing the design to a different orthogonal array: OA(27, 5, 3, 2).

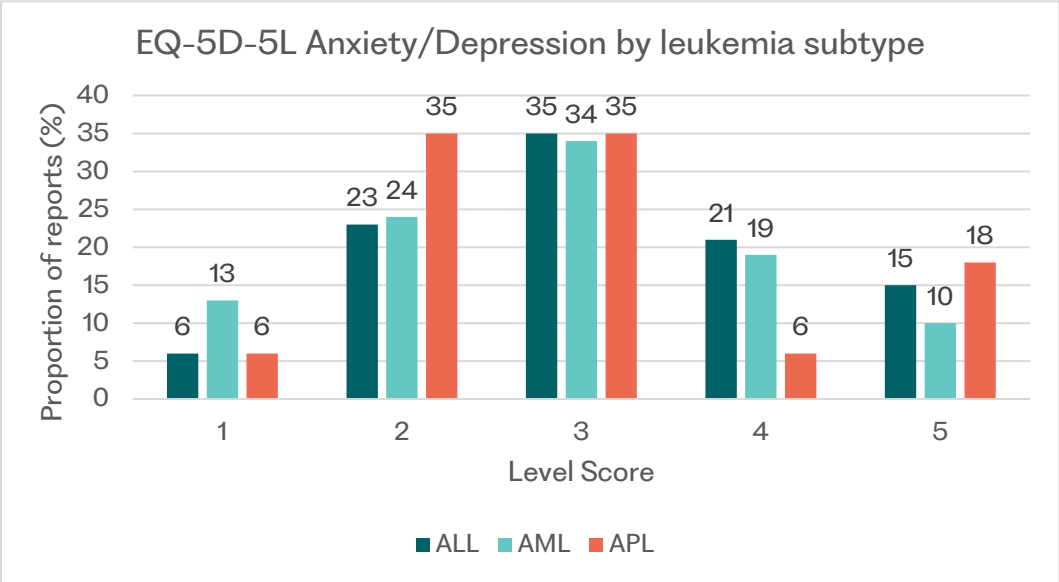
A2. Self-reported HRQoL: dimension scores by leukemia type

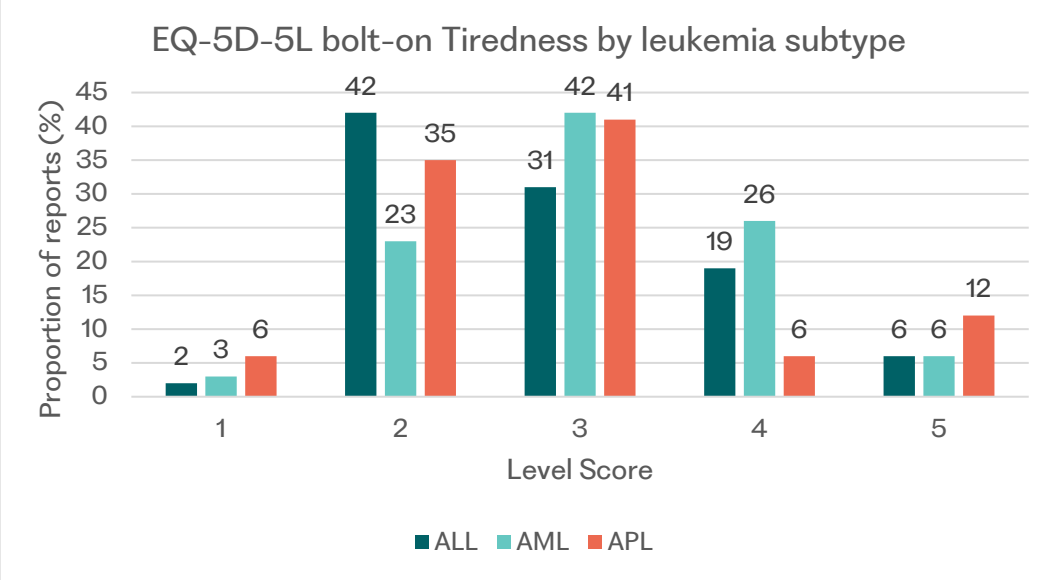


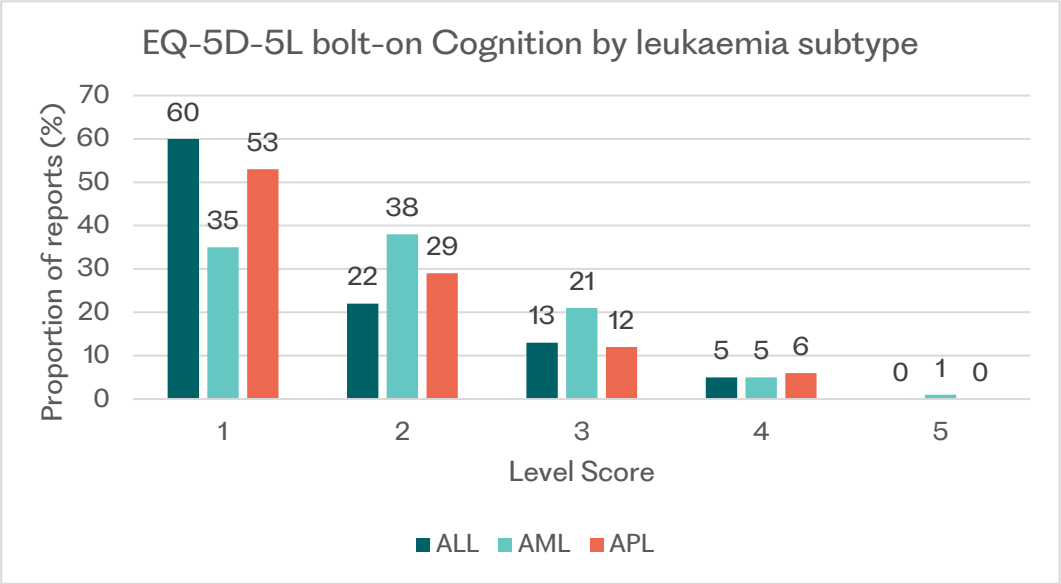



EQ-5D-5L Usual Activities by leukemia subtype
Proportion of reports (%)
60
50
40
30
20
10
0
8
15
24
48
30
47
23
39
24
19
14
5
2
2
0
1
2
3
4
5
Level Score
ALL
AML
APL


EQ-5D-5L Pain/Discomfort by leukemia subtype
Proportion of reports (%)
60
50
40
30
20
10
0
8
8
18
52
35
47
21
47
29
16
8
6
3
2
0
1
2
3
4
5
Level Score
ALL
AML
APL


EQ-5D-5L Anxiety/Depression by leukemia subtype
Proportion of reports (%)
40
35
30
25
20
15
10
5
0
6
13
6
23
24
35
35
34
35
21
19
6
15
10
18
1
2
3
4
5
Level Score
ALL
AML
APL


EQ-5D-5L bolt-on Tiredness by leukemia subtype
Proportion of reports (%)
45
40
35
30
25
20
15
10
5
0
2
3
6
42
23
35
31
42
41
19
26
6
6
6
12
1
2
3
4
5
Level Score
ALL
AML
APL



A3. Self-reported HRQoL: summary scores by country

TABLE A2. SUMMARY HRQOL SCORES FOR PATIENTS, BY COUNTRY

COUNTRY	N	EQ-VAS MEAN (SD)	EQ-VAS MEDIAN [IQR]	LSS MEAN (SD)	LSS MEDIAN [IQR]
UK	40	64.38 (23.38)	70 [31]	10.63 (3.84)	10 [6]
France	15	73.13 (17.69)	78 [20]	8.73 (3.61)	8 [3]
Germany	37	63.51 (15.65)	68 [24]	11.51 (2.67)	11 [3]
Italy	32	19.84 (8.47)	19 [18]	16.88 (2.87)	17 [4]
USA	88	69.51 (8.33)	71 [10]	11.24 (3.29)	10 [4]
Total	212	60.25 (22.43)	68 [25]	11.84 (3.93)	11 [5.5]

A4. BWS results: general population count analysis

TABLE A3. COUNT ANALYSIS FOR THE GENERAL POPULATION SAMPLES

DIMENSION LEVEL	CHOSEN AS BEST (N) [*A*]	CHOSEN AS WORST (N) [*B*]	BEST-WORST SCORE [*A-B*]	APPEARANCES (N) [*C*]	STANDARDISED BEST-WORST SCORE [*(A-B)/C*]
Mobility					
Level 1 (MO1)	997	88	909	2301	0.395
Level 2 (MO2)	168	167	1	2298	0.000
Level 3 (MO3)	198	704	-506	2555	-0.198
Self-Care					
Level 1 (SC1)	758	147	611	2298	0.266
Level 2 (SC2)	162	192	-30	2304	-0.013
Level 3 (SC3)	192	576	-384	2552	-0.150
Usual Activities					
Level 1 (UA1)	956	107	849	2301	0.369
Level 2 (UA2)	157	165	-8	2298	-0.003
Level 3 (UA3)	200	553	-353	2555	-0.138
Pain/Discomfort					
Level 1 (PD1)	808	101	707	2298	0.308
Level 2 (PD2)	144	261	-117	2301	-0.051
Level 3 (PD3)	134	973	-839	2555	-0.328
Anxiety/Depression					
Level 1 (AD1)	487	194	293	2298	0.128
Level 2 (AD2)	141	245	-104	2301	-0.045
Level 3 (AD3)	137	675	-538	2555	-0.211
Tiredness					
Level 1 (TI1)	400	165	235	2304	0.102
Level 2 (TI2)	175	174	1	2298	0.000
Level 3 (TI3)	208	432	-224	2552	-0.088
Cognition					
Level 1 (CO1)	624	84	540	2298	0.235
Level 2 (CO2)	61	250	-189	2298	-0.082
Level 3 (CO3)	47	901	-854	2558	-0.334

About OHE

With over 60 years of expertise, the Office of Health Economics (OHE) is the world's oldest independent health economics research organisation. Every day we work to improve health care through pioneering and innovative research, analysis, and education.

As a global thought leader and publisher in the economics of health, health care, and life sciences, we partner with Universities, Government, health systems and the pharmaceutical industry to research and respond to global health challenges.

As a government-recognised Independent Research Organisation and not-for-profit, our international reputation for the quality and independence of our research is at the forefront of all we do. OHE provides independent and pioneering resources, research and analyses in health economics, health policy and health statistics. Our work informs decision-making about health care and pharmaceutical issues at a global level.

About ALAN

The Acute Leukemia Advocates Network (ALAN) is an independent global network of patient organisations, dedicated to changing outcomes of patients with acute leukemias by strengthening patient advocacy in that area. We aim to maximise the capacity of members within the network to allow us, together, to deliver tailored services to acute leukemia patients and carers on the national level. All whilst joining forces between organisations on the policy and research level across countries.

ALAN is hosted under the umbrella of the Leukemia Patient Advocates Foundation (LePAF), a patient-led non-profit foundation based in Switzerland. As a foundation we connect leukemia patient organizations on all continents to strengthen advocacy work. The mission is to improve the lives and survival of patients affected by leukemia as well as their relatives by supporting leaders in providing help and support.

SEPTEMBER 2025