Identifying health and frailty in multiple myeloma: a clinical and biochemical approach to improve outcome prediction

A thesis submitted to the University of Manchester for the degree of Doctor of Medicine in the Faculty of Biology, Medicine and Health

2023

Luke J M Carter-Brzezinski

School of Medical Sciences, Division of Cancer Sciences, Department of Haematological oncology

Contents

Word count 48,485

1.1: Index of Figures

1:2: Index of Tables

1.3: Abbreviations

ADL- Activities of Daily Living
ANOVA- Analysis of variance
CCI- Charlson Co-morbidity index
CGA- Comprehensive Geriatric Assessment
DVTD- Daratumumab, velcade, thalidomide, dexamethasone
ECOG- Eastern Co-operative Oncology Group
EORTC- European Organisation for Research and Treatment of cancer
FCI- Frieburg Comorbidity Index
FDR- False Discovery Rate
FISH- Fluorescence in situ hybridisation
HCT-CI- Haematopoietic Stem Cell Transplantation Comorbidity Index
HR- Hazard Ratio
IADL- Instrumental Activities of Daily Living
ISS- International Staging system
IMWG- International Myeloma Working Group
IRAS: Integrated Access Research System
KPS- Karnofsky Performance Status
MGUS- Monoclonal Gammopathy of Uncertain Significance
MRP- Myeloma Risk Profile
NICE- National Institute of Clinical Excellence
PC- Principal Components
PCA- Principal component analysis
PS- Performance status
RF- Random Forest
R-ISS- Revised International Staging System
SBDC- Stollar Biomarker Discovery Centre
SWATH-MS- Sequential Window Acquisition of All Theoretical Mass Spectra

1.4: Abstract

Fitness and frailty can have a significant impact on the capability of a patient to tolerate cancer treatment. However, identifying these characteristics can be challenging in the setting of cancer where a range of physical and psychological factors may affect perceived fitness, and where cancer-related health deficits may be reversible. An objective tool to distinguish effects of frailty from those of disease could significantly aid patient management and treatment selection. This research investigated a real-world population of patients with myeloma seeking to define how relative impacts of different health impacts might be detected and quantified.

Consecutive patients undergoing therapy for newly diagnosed or relapsed multiple myeloma at a tertiary centre were recruited over a three-year period (n=91). Data comprised standard of care tests of myeloma or general health, together with serum samples for quantitative protein analysis using SWATH mass spectrometry. Additional quality of life (QoL) assessments was performed using standard validated tools. Data sets were tested as appropriate using standard testing, survival analysis, exploratory statistical analysis, and supervised machine learning to look for features primarily reflecting fitness frailty or myeloma disease.

The patient group had characteristics that were comparable with published cohorts. Initial use of Principal Component Analysis (PCA) suggested that three clinically distinct groups could be identified: those with high myeloma disease burden, those with myeloma and a high frailty burden and those with myeloma but with minimal adverse features; the groups had distinctive patterns of survival with a significantly inferior outcome in our novel high frailty burden group. Although these groups have potential clinical utility, large clinical datasets and PCA are not well suited to prospective patient classification in a routine clinical setting. Therefore, the serum mass spectrometry was performed on patient serum, and the datasets obtained were explored to identify candidate serum protein biomarkers that could prospectively identify the groups. Supervised analysis of serum proteins using Random Forest was performed on a representative group of patients to identify potential biomarkers. This was refined to identify a candidate set using standard significance approaches ANOVA, T-test and ROC analysis as well as looking for potential biological differences between proteins in the different subsets. The QoL results were used to

examine the relationship between frailty, myeloma and QoL. The EORTC MY-20 questionnaire was found to be a suitable and acceptable tool. Analysis of the results of QoL data demonstrated a deterioration of QoL as the number of lines of therapy increased and was related to the overall survival in our cohort. It was also shown that our novel frail cohort of patients had an inferior self-reported quality of life.

In summary, this study has identified clinically important subsets of patients that can be identified by standard of case testing and serum protein analysis. The findings of this research should require validation in prospective analyses but have the potential for changing treatment selection and management of patients with multiple myeloma.

1.5: Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

1.6: Copyright Statement

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and they have given the University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made. Presentation of Theses Policy You are required to submit your thesis electronically Page 10 of 24

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=2442 0), in any relevant Thesis restriction declarations deposited in the University Library, the University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in the University’s policy on Presentation of Theses.

1.7: Acknowledgement

I would like to acknowledge all the patients who allowed their data and samples to be used in this project.

1.8: Research Experience

Prior to this project, I undertook research in the area of chronic lymphocytic leukaemia as part of my undergraduate BSc. This work resulted in publication. I have no other formal laboratory research experience but do participate in large national trials as a clinician.

1.9: Experimental Approach

The overall approach to this piece of research was to distinguish frailty based on clinical parameters analysed using standard and exploratory statistical approaches applied to a patient cohort with a diagnosis of multiple myeloma. Then to use those groupings to explore whether biomarkers could be identified through proteomic and bioinformatic analysis of serum from the same population of patients to enable separation into distinct and useful clinical cohorts. These cohorts included: overall survival (prognosis), frailty and quality of life. Finally, to see how the quality of life of myeloma patients was affected by their disease using data analysis of a validated questionnaire. The overall aim was to improve and simplify the identification of important clinical cohorts within newly diagnosed and relapsed patients with myeloma, to help physicians make difficult clinical decisions in this challenging cohort.

2: Introduction Part 1: frailty and myeloma as separate entities

2.1: What is frailty?

Fitness and frailty represent two ends of a continuum of health and wellbeing. This is an important concept to consider when treating cancer patients.[1] At one end of the spectrum, if the fit patient is undertreated with a subtherapeutic (reduced) dose of therapy, there is a risk of failure to achieve remission or achieve cure.[2] Equally, if a frail patient is treated with an intensive chemotherapy regimen, the off-target side effects may impact significantly the individual.[3] However, fitness and frailty are not binary constructs and can therefore be difficult to confidently identify. In the haematological malignancy of multiple myeloma, this distinction is made more difficult by the superimposition of disease burden onto the individual at the start of treatment.[4]

Before discussing the concept of frailty in myeloma, it is important to define the term frail. There are a substantial number of definitions of frailty in the literature, with a combination of disease specific and non-disease specific models available. Frailty is often however described as "a clinical syndrome of older patients, which results in reduced physiological reserve, increased vulnerability to stressors, and is associated with poor disease outcome"[5, 6]. Frailty may also be a precursor to disability.[7] While this is an insightful description of frailty, its clinical utility is questionable as it does not help rapid identification of a frail individual in the consultation room. Indeed, The British Geriatric Society suggest that a comprehensive geriatric assessment is likely to take 1.5 hours to complete.[8]

The literature also highlights that frailty is a dynamic process, with frequent transitions between frailty states. A study of a population in the USA of over the age of 70 demonstrated that 57.6% of patients had at least one transition, (from non-frail to pre-frail, or pre-frail to frail) during a 54 months follow up.[9] Again, this is not a population with a diagnosis of cancer, but it does emphasise the dynamic nature of health and frailty, and suggests that frailty should be assessed as continuous variable, rather than at a single time point.

Finally, it is important to mention that frailty and age are not analogous. Whilst the incidence of frailty does increase with age, there is not a linear correlation between them. This is referred to as 'biological age' in literature.[10]

2.2: How is frailty measured?

Freid *et al* [11]sought to simplify and standardise the definition of frailty by reducing it to the following variables: unintentional weight loss, self-reported exhaustion, weak grip strength, low physical activity, and slow walking speed. An original cohort of 4537 patients were followed up for 7 years. The presence of three or more of the above variables at the start of the study was defined as 'the frailty phenotype'. This classification was a statistically significant predictor for activity of daily living disability, hospitalisation and death at three years.[11] However, there are challenges in measuring each of these variables. For example, a wheelchair bound individual would be unable to complete the gait assessment metric. This emphasises the need for appropriate frailty assessments suitable to the specific population being tested.[12] Furthermore, Freid definition has also been criticised for not including a measure of cognitive impairment, which can impact on an individual's ability to withstand physical stressors and disease states.[13]

While any attempt at simplification and standardisation of the concept of frailty is to be applauded, it may not be appropriate and generalisable to all patient groups. Specifically, using the example of patients with a new diagnosis of multiple myeloma, they may be suffering from disease specific and potentially reversible health issues.[4] As an example, at diagnosis, a patient may be confused secondary to hypercalcaemia and be bed bound due to spinal cord compression. Both issues are potentially reversible with therapy but would have a pejorative effect on the Freid 'frailty phenotype' at diagnosis. The risk would therefore be that patients were inappropriately labelled as 'frail', and not given appropriate 'intensive therapy regimens' [14] This example highlights the clinical need for a disease specific and dynamic frailty assessment in multiple myeloma.

An alternative approach is to calculate the 'Frailty Index' for patients.[15] This method, engineered by a Canadian group uses a statistical formula that incorporates 92 variables linked with frailty. They showed that their outcome measure (time to death) was more

closely correlated to the 'frailty index' than to chronological age. While this again is a potentially useful statistical model, its complexity would preclude use in the clinical setting, where the resource of time is a scarce commodity. It does however demonstrate that large data sets can yield significant prognostic information. Interestingly, a later study demonstrated a moderate correlation between 'Frailty Index' and 'Fried's Frailty Phenotype'.[16]

A further attempt at simplification of frailty assessment is the Comprehensive Geriatric Assessment (CGA) and is often used in routine clinical practice.[17] It has some similarities to the Frailty Index but focuses on just 7 domains of frailty including: functioning, physical health, cognitive health, mood, co-morbidities, medications and socioeconomic circumstance. Despite focusing on less domains than the Frailty Index, it is still time consuming, requiring 1.5 hours to complete.[8]

2.3: Is frailty associated with inferior outcome for cancer patients?

A premise of this research project is that frailty causes inferior outcomes in patients with cancer, specifically multiple myeloma. This premise is corroborated by a systematic review of cancer patients.[1] That review used data from 20 studies, with a total of 2916 patients included and found frailty to be independently associated with all-cause mortality, hazard ratio (HR) 1.87.[1] It therefore offers robust evidence that 'frail' cancer patients have worse outcomes. It was not however specific to a multiple myeloma cohort.

There is also evidence that frailty in cancer patients leads poor tolerance the treatment regimens.[18] This has resulted in the American Society of Clinical Oncology recommending that all patients over the age of 65 should have a geriatric assessment prior to commencing chemotherapy as the standard of care.[19] This is with the aim of individualising treatment regimens and improving outcomes and side effect tolerability.

2.4: Is frailty in cancer a significant epidemiological issue?

Having established the relationship between frailty and poor outcome in cancer, it is also important to appreciate the scale of the problem. Data published from the USA Cancer

Register demonstrating that 80% of the annual deaths from cancer occurred in the elderly population.[20] It is important to note that advanced age does not equate to frailty, but frailty does increase with age. This would therefore suggest that frailty in cancer is a significant epidemiological issue. Furthermore, Mohile *et al* reported that a diagnosis on a non-skin cancer was associated with an increased number of patients 'meeting the criteria for frailty', adjusted odds ratio (OR 1.46).[21] Despite the increased incidence of myeloma in the elderly, it can occur in younger patients. Cancer Research UK report that one in 100,000 of the UK population in the 35-39 age range was diagnosed with multiple myeloma in the period between 2016-2018.[22] It is therefore important that any assessment of frailty should be applicable to all age groups, not just the elderly.[23]

An important issue highlighted in a recent review article by Rodriguez-Manas is that there must be clinical utility in identifying frailty.[7] If there were to be no beneficial interventions based on identifying the syndrome, its identification would be futile. This is manifestly not the case in multiple myeloma, where existing treatment options are frequently based on a patients' perceived health or frailty. Indeed, Ethun *et al* suggest that frailty scores can form part of the decision-making process regarding regimen and dose of chemotherapy patients revieve.[18] This is of particular importance as the elderly population is underrepresented in clinical trials.

2.5: What is a biomarker?

The term biomarker is shorthand for biological marker. It was defined in 2001 by the Biomarkers Definitions Working Group as "A characteristic that is objectively measured and evaluated as an indicator of normal biological process, pathogenic process, or pharmacological responses to a therapeutic evaluation".24 In the context of this thesis a biomarker should be thought of as a surrogate marker of disease activity or level of frailty in our cohort of patients.

2.6: Do biomarkers of frailty exist?

As previously mentioned, there are multiple scoring systems designed to assess different aspects of frailty. While they have been shown to provide useful and statistically significant

prognostic information, they are often time consuming and difficult for the clinician to undertake as part of routine practice.[8] A biomarker (or set of biomarkers) that correlates with frailty is an appealing proposition. Indeed, they not only have the potential to improve the accuracy of the diagnosis of frailty, but also to guide and personalise treatment decisions.[25]

The rational for a biomarker in frailty is outlined by Ferrucci *et al*.[26] The hypothesis is that at least part of the cause of functional decline in older frail patients is the loss of a functional reserve and subsequent failure of compensatory mechanisms. Functional reserve is an organ specific concept involving the residual capacity of an organ to carry out its primary physiological function. For example, one might discuss the functional reserve of the myocardium after a myocardial infarction and its reduced ability to facilitate tissue perfusion due to reduced cardiac output. The concept of functional reserve is often used in discussions of frailty, where the physiological function of multiple organs decline.
The net result is thought to be a loss of homeostasis due to alterations in metabolic pathways such as increased insulin resistance, immunometabolic dysfunction and resultant oxidative stress, as well as the stress response through the hypothalamic, pituitary adrenal axis.[27] The rational is therefore that biomarkers can be identified as biproducts of dysfunction of these pathways.

Cardoso *et al* identified a set of biomarkers that have the potential to identify frail patients.[28] This meta-analysis identified potential biomarkers in literature and categorised them into the following categories: inflammation, mitochondria and apoptosis, calcium homeostasis, fibrosis, neuromuscular junction and neurones, cytoskeleton and hormones or 'other'. They then proceeded to identify 19 biomarkers of high importance, based on the strength of evidence in the literature. These potential biomarkers (shown in Table 1) offer a useful insight into how frailty might be assessed biochemically, rather than clinically.

Table 1: List of biomarkers of high importance in frailty

Inflammation Related	**Mitochondria and Apoptosis Related**	**Other markers**
CXCL10 (C-X-C motif chemokine ligand 10)	GDF15 (growth differentiation factor 15)	AHCY (adenosylhomocysteinase)
IL-6 (interleukin 6)	FNDC5 (fibronectin type III domain containing 5)	miRNA (micro Ribonucleic acid) disease and pathway specific panels
CX3CL1 (C-X3-C motif chemokine ligand 1)	Vimentin (VIM)	KRT18 (keratin 18)
Calcium Homeostasis Related	**Fibrosis Related**	**Neurone Related**
Regucalcin (RGN/SMP30)	PLAU (plasminogen activator, urokinase)	BDNF (brain derived neurotrophic factor)
Calreticulin	AGT (angiotensinogen)	Progranulin (PGRN)
Cytoskeleton and Hormone Related		
α-klotho (KL)		
FGF23 (fibroblast growth factor 23)		
FGF21 (fibroblast growth factor 21)		
Leptin (LEP)		

This list of frailty biomarkers is reproduced from the paper by Cardoso et al[28] They are subclassified according by physiological categories.

2.7: Frailty section summary

In summary, frailty is a difficult entity to define and identify, especially in the context of an acute disease process such as myeloma. It is best understood as a phenotype rather than by its underlying physiological causes. However, the identification of frailty is important as it impacts the intensity of the chemotherapy regimen selected and therefore on prognosis. A simple, robust and disease specific frailty identifier would therefore be of significant clinical use in the context of multiple myeloma.

Frailty is often used ambiguously in literature. Some literature uses it in the context of the precursor to disability, while other sources would include factors such as comorbidities. For this research, the term vulnerability will be used. The working definition of vulnerability is the combination of frailty, resultant disability and concurrent comorbidity. This will be expanded on in subsequent sections.

2.8: What is multiple myeloma?

Multiple myeloma is a haematological malignancy of plasma cells, a form of terminally differentiated B-cell involved in humoral immunity.[29] In health, B cells within the bone marrow undergo heavy and light chain rearrangement, before moving into the germinal centres of lymphoid tissue to mature. It is at the germinal centres that potentially oncogenic changes occur. Somatic hypermutation is a process by which antigenic affinity of the immunoglobulins is increased.[30] Class switch recombination involves removal of a small part of the heavy chain constituent of the immunoglobulin and permits the production of several isotypes, directed at the same antigen.[31] Both somatic hypermutation and class switch recombination can frequently give rise to oncogenic mutations, often involving chromosomal translocation. These often involve the IgH locus on chromosome 14. If an oncogene is translocated downstream to the IgH enhancer region, this can give rise to a significant pathogenic mutation,[32] with the result of increased proliferation and other oncogenic properties.

Multiple myeloma arises from its precursor states: smouldering myeloma and monoclonal gammopathy of uncertain significance (MGUS). These precursor conditions share the same genetic alterations as myeloma but by definition have a lower burden of disease and the absence of end organ dysfunction.[33] Multiple myeloma can present with a constellation of clinical features including: infections, anaemia, hyper-viscosity, bone pain with associated pathological fractures and acute renal injury.[34]

The prognosis of a patient diagnosed with myeloma is incredibly variable, with some patients dying within weeks, to others living over 20 years. However, Cancer Research UK's published data in 2020 shows an 82.7% one year survival, and a 52.3% five year survival.[35]

2.9: What are the cytogenetic changes that occur in myeloma?

Chromosomal changes in multiple myeloma are often split into primary and secondary abnormalities.[36] The premise is that the primary mutations are present when the condition is MGUS, and secondary aberrations occurring later in the process. The primary cytogenetic changes often involve the IgH locus on chromosome 14 (t(11;14), t(4;14), t(14;16), t(14;20) and rarely isolated monsomy 14) and others involve trisomic changes of the odd numbered chromosomes (excluding 1, 13 and 21). It is then thought that the secondary cytogenetic

abnormalities occur at progression and include 17p and 1q(amp), giving escape from cell cycle regulation with loss of function of tumour suppressor genes.[36] Abdallah et al have also shown that different cytogenetic changes give rise to specific disease features, with t(4;14), t(14;16) and t(14;20) associated with more severe anaemia and a higher beta2microglobulin (B2M), while those with t(4;14) have higher serum monoclonal protein values.[37]

2.10: Is myeloma a significant epidemiological issue?

The age standardised incident rate (ASIR) of multiple myeloma was found to be 2.1 per 100,000 people in 2016, with an increased incidence of 126% globally in the 16 years from 1990.[38] In the UK, it is reported that approximately 5000 patients are diagnosed with myeloma each year, making multiple myeloma the second most common haematological malignancy in the UK.[39] Given the fact that over two thirds of new diagnoses are in the over 70s category, there is irrefutable evidence of the problem of frailty in myeloma.

2.11: How is disease burden and prognosis assessed in multiple myeloma?

Scoring systems help clinicians classify burden of disease and link this with prognosis. The Myeloma International Scoring System (ISS) produced by Greipp *et al* was published in 2005.[40] ISS uses serum beta-2-microglobulin (B2M) and albumin to classify patients into three risk groups (stages I-III), each with statistically different overall survivals. Of note, this scoring system is simple to calculate and easy to assimilate into clinical practice as it requires the assimilation of only two standard of care tests.

B2M is a protein constituent of the HLA Class I antibody, found on the surface of most cells, though predominantly B-lymphocytes. It is found in increased concentration in serum in myeloma due to increased proliferation and turn over of B-cells, along with reduced glomerular excretion in renal dysfunction.[41] Albumin is a liver derived protein, important in maintaining osmotic pressures and transporting molecules in plasma. Reduced albumin levels are again not specific to myeloma but occur due to either reduced production as part of an IL-6 driven acute phase response, or due to increased excretion due glomerular damage and nephrotic syndrome.

The ISS score has been further refined Palumbo *et al*, who expanded the scoring system to include two other parameters: LDH and interphase fluorescent in-situ hybridisation (FISH).[42] LDH is the enzyme used to convert pyruvate to lactate and is raised in multiple malignancies due to changes in glucose metabolism in malignant cells.[43] FISH was used to identify the high risk changes: del 17p and/or t(14;16) and t(4,14). The presence of just one of these chromosomal abnormalities classifies a patient as high risk'. Both investigations are standard of care tests in the UK at diagnosis, making the widespread adoption of this tool a possibility. The paper reports that the revised-ISS (R-ISS) is superior to the ISS as it reclassifies 26% of patients.[42] The R-ISS reclassifies this cohort of patients from the good prognosis group into one other inferior prognostic groups (Stage II-III) and in doing so improves the separation of the survival curves. This results in improved prospective prognostication.

The R-ISS offers two interesting insights into scoring systems in myeloma. Firstly, it shows that aggregating different variables or scoring systems can improve prognostication (and assessment of disease burden). It does however also highlight the need for scoring systems to be sufficiently easy for the clinician to incorporate into their day-to-day practice. While bone marrow FISH of CD138 selected cells is a standard of care test in the UK, this is not the case in less economically developed countries with limited resources.[44]

2.12: Is the biology of myeloma different in different age groups?

There is debate as to whether multiple myeloma diagnosed in older patients has a more aggressive phenotype. If this was the case, then it could be argued that frailty by itself is a less important contributor to inferior survival in the elderly population. Ludwig *et al* published evidence that suggested patients under the age of 50 had more favourable features including lower International Staging System (ISS) score at diagnosis, as well as more preserved haemoglobin and renal function.[45] However, other sources have failed to reproduce this finding. One study reviewed the cytogenetic abnormalities found in 783 multiple myeloma cases and found there to be no difference between the number of high-risk cytogenetic abnormalities between young and elderly patients.[46] Loss of -y and 5+ were more common in the elderly, but these are not considered to be high-risk chromosomal

abnormalities. Indeed the incidence of t(4;14) was found to be lower in the over 65 cohort than in the under 65s, with the incidence 17p deletion being similar in all age groups.[47] There is therefore no definitive evidence to suggest that multiple myeloma in elderly cohorts is intrinsically more aggressive, suggesting other factors are important the inferior outcomes in this group.

2.13: How is multiple myeloma treated?

The treatment algorithm in multiple myeloma in the UK is complicated and evolving. At present patients are split into two groups at diagnosis: fit patients who are considered autologous transplant eligible, and unfit patients who are considered transplant ineligible. Patients deemed fit for an autologous stem cell transplant are usually treated with a drug combination of a proteasome inhibitor (bortezomib), an immunomodulatory drug (thalidomide), a monoclonal antibody (daratumumab) and a steroid (dexamethasone).[97] This regimen is known as DVT-D. The treatment algorithm for non- transplant eligible patients is less prescriptive but involves a combination of the classes of drugs previously mentioned, but at an attenuated level, often without a monoclonal antibody. As patients relapse they will be challenged by different drugs from the same classes of agents in the second and third line setting until a decision is made to stop myeloma directed thrapy.[48] There are also multiple trial therapies in multiple myeloma including chimeric antigen receptor T-cell therapy CAR-T), targeting the B-cell maturation antigen (BCMA) expressed on the surface of plasma cells.[49] Another therapeutic option is the group of drugs (such as teclistamab) know as the T-cell redirecting bispecific antibodies that also target specific plasma cell surface antigens.[50]

2.14: Is frailty an unmet need in current myeloma therapy?

The treatment landscape of multiple myeloma has changed dramatically over recent decades, from conventional chemotherapy, to the introduction of high dose melphlan autologous stem cell transplants, to novel antibody directed and pathway specific therapies.[51] Pozzi *et al* sought to establish whether the changing treatment paradigm had resulted in improved outcomes in different age subsets (<65 years, 65-74 years, >75

years).[52] Their real word data set demonstrated a significant improvement in overall survival in the younger two age groups, but no change in overall survival in the >75 age group in 1997-2005 cohort to patients treated after 2006. This is a remarkable finding, given the number and varieties of novel therapies licensed by the National Institute of Clinical Excellence (NICE).[14] This therefore raises the question as to whether there is an intrinsic property of chronological age that leads to worse outcomes despite treatment advances. Alternatively, it may be that age in this setting is simply grouping a more frail phenotype in the same cohort, as epidemiological data shows an increase in incidence of frailty with chronological age.[21]

2.15: Myeloma Section Summary

Kint *et al* argue that there is a strong base of evidence that suggests that myeloma therapies are improving outcomes in the younger age groups, and they have the potential to do so in selected elderly patients.[53] The challenge is to identify those older patients who would benefit from these novel and intensive anti-myeloma treatments. This is made more difficult since elderly and frail patients are underrepresented in clinical trials due to co-morbidities.[54]

Importantly however, the current UK national myeloma trial, the FITNESS study, is seeking to answer part of this question.[55] One of its primary objectives is to assess if upfront dose modification of therapy in non-transplant eligible patients has a significant effect on overall survival and treatment related morbidity and mortality. This is an important and welcomed first step in individualisation of treatment based on consideration of frailty.

3: Introduction Part 2: frailty and myeloma as a combined construct

3.1: How can disease, frailty, comorbidity and disability in myeloma be disambiguated?

The previous sections of the introduction have discussed the concept of frailty as a discrete entity, as well as provided information about multiple myeloma itself. The present section will explore the interaction between the two, and how they are currently assessed.

One of the principal objectives of this study is to disambiguate health and frailty in the setting of myeloma. The terms frailty, comorbidity and disability are often used interchangeably. However, Kint *et al* ague these terms are in fact distinct entities and can be encompassed by the term 'vulnerability'.[53] They go on to argue that each of these components should be assessed individually in order to decide on the optimal therapeutic approach for the individual.[53] Comorbidities are defined in that paper as pathological conditions or pre-existing conditions a patient has at the time of diagnosis. Examples include ischaemic heart disease, COPD or diabetes. Disability is described as the effect that illness has on the individuals' ability to carry out activities of daily living. Finally, as previously mentioned, frailty is the reduced physiological reserve of an individual, leading to an increased vulnerability to specific stressors. Each of these three variables can be assessed using different scores or tools.

Kint *et al* suggest that clinician's clinical judgement of a patient's overall vulnerability is often the primary method used.[53] This is inherently subjective and flawed assessment, subject to considerable bias and leading to poor inter-clinician reproducibility. The argument is therefore that vulnerability is multi-layered and is best assessed using objective methodology. The term vulnerability will therefore be used in this thesis as the preferred descriptor for the physiological state that encompasses frailty, disability and co-morbidities as a joint construct.

3.2: How are co-morbidities presently assessed in myeloma patients?

Objective methods already exist to measure the burden of co-morbidities in a multiple myeloma patient. Two such examples are the Charlson Comorbidity Index (CCI) and the

Freiburg Comorbidity Index (FCI). The FCI has been validated as an independent risk factor for reduced overall survival.[56] When compared directly in a retrospective analysis of myeloma patients, the FCI was a more sensitive tool for predicting overall survival than the CCI ($p > 0.001$ Vs p 0.059).[57]

While these scoring systems were not designed to be specific myeloma assessments, they are a commonly used tool in outcome analysis of myeloma patients. These data suggest that, while the overall number of comorbidities may be important, specific comorbidities are the more important factor. Bringhen *et al* found that renal failure at presentation (along with biological age) had a statistically significant negative impact on overall survival.[58] Engelhardt *et al* corroborated this finding, noting that renal impairment and respiratory comorbidities were independent variables that affect overall survival. [59] It is therefore logical to conclude that if some co-morbidities are more important than others, a disease specific co-morbidity score may have benefit.

3.3: How is performance status assessed in myeloma patients?

Another aspect of vulnerability that can be assessed in a structured (though not disease specific way) is 'performance status' which approximately equates to disability in the Kint classification.[53] Scores measuring 'activities of daily living' (ADL)[60] and 'independent activities of daily living' (IADL)[61] have been used since 1989.[62] Other commonly used scoring systems include the Karnofsky Performance status (KPS)[63] and the Eastern Cooperative Oncology Group (ECOG) Performance Score.[64] The latter has been shown in a Japanese cohort of patients (undergoing first line autologous stem cell transplant) to be independently associated with overall survival.[65]

While the assessment of performance status at baseline is a recommended assessment, the literature does suggest caution. Clinician performed performance scores, such as the KPS are subject to subjective bias and may not correlate with patient reported scores. In one small study of 40 patients, despite being given a score for greater than or equal to 80% (normal activity with some difficulty, some symptoms or signs), 70% had difficulty with moderate activity, and 25% had difficulties with bathing or dressing as recorded in patient

reported scores.[66] This finding emphasises the potential difference between clinician assessed and patient assessed outcomes.

3.4: How is frailty assessed in myeloma patients?

There have been multiple attempts to produce cancer specific frailty screens, many of which were assessed in a systematic review.[67] The logic for these screening tools was to facilitate a more rapid and less time-consuming method to identify frail patients than with the Comprehensive Geriatric Assessment (CGA). That review assessed a variety of tools in comparison to the GCA, with disappointing results.[67] While the sensitivity of some methods was promising in oncology patients (92%), the specificity and negative predictive value of these tools was disappointing, leading them to conclude that 'frailty screening methods have insufficient discriminative power' to assign a frailty status to an individual.

The use of these tools specifically in myeloma is difficult, as myeloma disease related comorbidities are reported to be a confounding factor.[68] It is suggested that the pathophysiology of myeloma acts an 'accelerant' on the aging process. Specific examples include the effect of myeloma induced anaemia on energy levels, and the effects of therapy (e.g. steroids) on muscle mass and by proxy the perception of frailty as measured in the assessments. Some of these effects will be reversible, while others will result in long term sequelae. It is this differentiation that may confound scores such as the CGA.

3.5: Are there biomarkers of frailty that are specific to multiple myeloma?

There is already some evidence to suggest that myeloma-specific biomarkers of frailty may exist and have potential clinical benefit. Cook *et al* outline three possible categories of biomarkers of frailty in myeloma including: cellular senescence, inflammaging and sarcopenia.[68] The specific biomarkers identified are summarised in Table 2. It is acknowledged that this data has not been prospectively validated, however it does identify a role for specific biomarkers of frailty, specific to a myeloma cohort.

Table 2: Potential biomarkers of frailty in multiple myeloma

Subgroup	Description	Examples of biomarkers
Cellular Senescence	Defined by: arrest in cell proliferation, resistance to apoptosis, production of senescence associated secretory phenotype (SASP)	DNA damage markers (γH2AX, ATM, MDC1) Telomere length/ telomere dysfunction foci from DNA extracted from peripheral blood Cell cycle arrest (p16INK4A, p53/p21 axis) Senescence associated B-galatosidase
Inflammaging/ immunesenescence	Accumulation of age dependent inflammatory mediators secondary to chronic low-level inflammation	IL-6, IL-8, IL-1, TNF-alpha, b-CHE, eHsp72 Selenium MicroRNA panels (disease or pathway specific) Alteration in immune cell subsets: Th12/ Treg ratio, CD8+, CD28-, KLRG-1+ quantification
Sarcopenia	Reduction in muscle mass	Apendicular skeletal muscle mass reduction (DXA or CT derived) Muscle strength (grip strength) Performance (up and go walking test) Potential biomarkers: myostatin, IGF-1, CRP

This table has been adapted from Cook et al[68] and details a list of potentially useful biomarkers of frailty, but this time specifically in multiple myeloma patients.

3.6: Can aggregate scores assess disease burden and frailty in myeloma patients?

In another retrospective data registry analysis, Offidani *et al* sought to investigate the utility of a 'vulnerability score' which combined the Charlson Co-morbidity index (CCI) with ECOG performance status.[69] Patients with a PS of 2-4 or a CCI score of 1-3 were assigned a point each and then grouped into low (score 0), intermediate (score 1), high (score 2) risk groups. They demonstrated that the high-risk group had a statistically significant reduction in overall survival, compared to the low-risk group.

Another important aggregate frailty score currently used in myeloma is the International Myeloma Working (IMWG) Group Frailty Score.[70] Although described as a 'frailty score' it is actually a composite vulnerability assessment that includes age, two activities of daily living assessments: Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) (which could be considered as assessments of disability) and the CCI (co-morbidity

assessment). The above variables split patients into three groups (fit, intermediate-fit and frail). This classification achieved statistical significance in separating the groups in terms of overall survival and progression free survival at three years. The IMWG Classification is reproduced in Appendix 1.

While the IMWG Frailty Score has achieved statistical significance for prognosis, the score has some weaknesses. Firstly, the amount of time required to complete the score in clinic is onerous. Secondly, the ADL and IADL components are subjective, either on the part of the patient reporting the score or in the interpretation by the doctor. Cook *et al* argue that a simpler score that predicts early treatment-related mortality and failure to tolerate medication is desirable for 'real world' myeloma management.[71]

Another criticism of the IMWG frailty score is the effect age has on a patient's classification. For example, a patient of 75 years of age, with no other issues (classified as fit), would automatically become 'intermediate-fit' on their 76th birthday without any change in other variables. This binary chronological age-related classification raises the question of whether age should be considered a continuous rather than discrete variable.

In an effort to overcome some of these perceived weaknesses, Cook *et al* devised a different prognostic system, which they report to be less subjective.[68] It requires baseline WHO Performance Status, ISS, CRP and age, which together form the 'UK Myeloma Alliance Risk Profile' or MRP. They undertook a retrospective analysis on non-transplant eligible patients in Myeloma IX (a cohort comprising of older, less fit patients deemed non-transplant eligible) and Myeloma XI patients (a mixed cohort containing both transplant eligible and non-eligible patients) and found that their groupings of low, intermediate and high risk were associated with progression free survival, early mortality and percentage of protocol drug dose delivered. Interestingly, there was also an association with baseline quality of life measures. These results have subsequently been validated in a real-world data set.[72]

The MRP model is an attractive model, given that the components are already standard of care. It does however not include cytogenetic risk, which is an independently validated predictor of survival. Furthermore, while the components of the score are easily measured, the MRP score is not easily calculated. The paper does not provide a simple methodology for calculating the individual patient's MRP score. Therefore, despite the constituent

components of the score being easy to assess, the score itself appears complicated to compute. This is perhaps a limiting factor in its clinical utility.

Another such example of an aggregate score, combining information of comorbidity and a prognostic score (ISS) it that proposed by Kleber *et al*.[56] A combination of these two variables allowed for subgroups to be further refined, with more significant differentiation of overall survival. This is an important paper as it aggregates a vulnerability score with a score that assesses disease burden and activity. Perhaps a more refined example of this, validated in a larger data set is the Revised Myeloma Comorbidity Index.[59] This composite assessment combines renal impairment, respiratory impairment, Karnofsky performance Status, frailty and age to subclassify patients into three groups, with statistically significant differences in overall survival of 10.1, 4.4 and 1.2 years respectively.

In summary there are a number of established scores that combine myeloma specific disease markers and vulnerability assessments to provide significant prognostic data. However, a common theme to all these scores is the high number of variables and the amount of time required to calculate them.

3.7: Are there any novel approaches to assessing vulnerability in myeloma patients?

Cook *et al* suggest a novel approach to the issue of fitness and frailty.[68] They hypothesise that data could be gained from 'wearable activity monitors'. They suggest that data collected by the individual, such as resting heart rate and step counters could provide interesting data to give insight into aspects of vulnerability. Razjouyan *et al* were able to identify a group of pre frail individuals using home pendant sensor technology and discriminate them from a healthy population.[73] While the use of technology is appealing it is not validated in myeloma and it is often younger generations that use the technology, rather than the frail or pre frail groups. As such there is as yet a lack of data for its use in the assessment of this patient cohort.

3.8: Summary of scores used to assess patients with multiple myeloma

In summary, there are numerous scoring systems available to assess patient characteristics in multiple myeloma and are summarised in Figure 1. The scoring systems can be categorised into those that assess one aspect of a patient's vulnerability or assess multiple aspects of vulnerability. There are others that assess a single aspect of myeloma disease burden, while others assess multiple parameters affecting burden of disease. Finally, there are composite scores that attempt to assimilate disease specific factors to vulnerability.

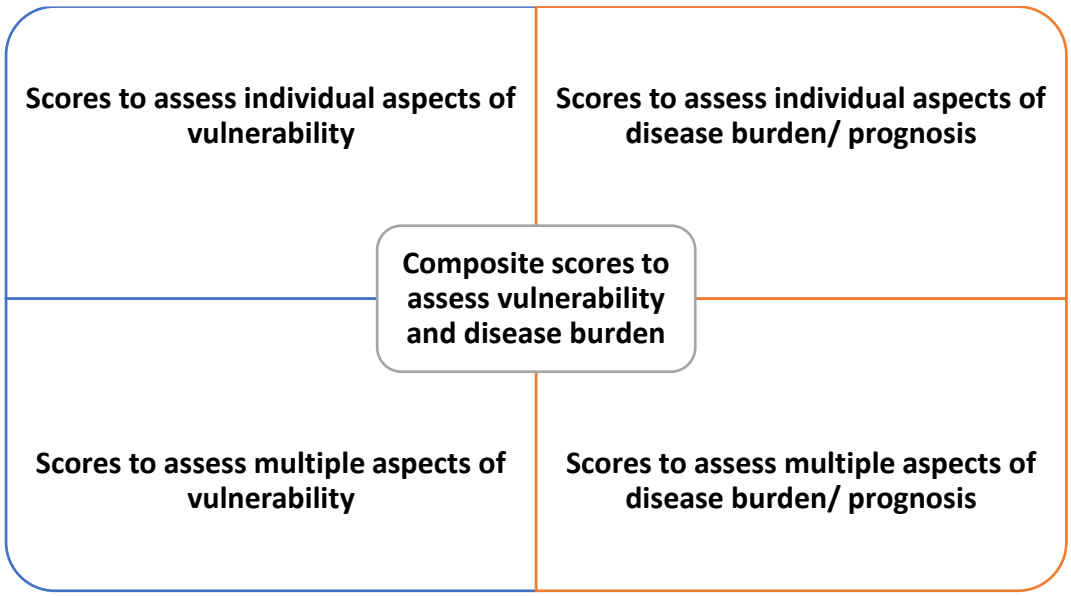


Figure 1: *The scoring systems used to assess disease and vulnerability in multiple myeloma*

This figure demonstrates that there are essentially five categories of scoring system that are used to assess disease and frailty in multiple myeloma. Some scores assess single aspects of disease or frailty, while others are composite scores, including a number of components.

Substantial insights into a patient's clinical condition can be gained from using the aforementioned scoring systems. Indeed, there is most definitely not a lack of options to assess disease burden or vulnerability. A strong argument can indeed be made that there are too many, each with subtle differences, but no uniformly accepted standard. The IMWG Frailty score and the R-ISS scoring systems are however the most widely accepted systems

used for assessing vulnerability and disease-burden respectively. They both provide evidence based, statistically significant and reproducible tool to help clinicians.

What these scores do not however do is differentiate patients who have a poor prognosis because of disease burden from those who are vulnerable due to disability, frailty and comorbidities. The unanswered question is: 'is there a group of patients who appear frail at diagnosis, in whom intense therapy would reverse the perceived vulnerability and lead to good overall survival outcomes?' If this is the case, can this group of individuals be easily identified in the clinical setting with easily available measurements? If this is possible, then there would be potential for a clinically significant change in management of newly diagnosed or relapsed myeloma.

4.0: Introduction Part 3: Big Data Sets

4.1: What is Big Data?

There is no universally accepted definition of big data. However, consensus is that it involves data sets with large amounts of separate data elements that require advanced analytics to search for meaningful patterns.[74] While large data sets (for example those with thousands of patients and data points included) have significant advantages over sets with limited variables (for example clinical, physiological and biochemical parameters in medical data sets), it is important that the correct statistical method should be used when exploring them.[75] Failure to do so could result in incorrect conclusions being drawn.

Historically, smaller numbers of variables have been assessed between different cohorts using statistical tests such as the T-test or ANOVA (Analysis of variance), looking for statistically significant differences between the cohorts.[76] However, this becomes challenging when simultaneously assessing cohorts with large numbers of variables which increase the potential for false discovery.

In this context, false identification of biomarkers can arise from the problem of multiple comparisons (Type 1 errors). Big data sets will compare large numbers of variables for associations. Statistically, if enough variables are tested for association, some false identifications of biomarkers will be made by chance (false positives). This problem can be addressed using single step methods for multiple testing correction for example Bonferroni correction.[77] The false discovery rate is another single step method that can be used to give the probability of an incorrect discovery. Sequential methods for multiple testing correction can also be used to limit the false identification and these include the Holm Correction and the Hochberg correction. Menyhart et al recommend that for exploratory analysis a single step method such as Bonferroni is suitable for analysis.[77]

4.2: What is Principal Component Analysis (PCA)?

Essentially PCA is a type of mathematical algorithm that assesses a large data set and simplifies it using a dimension reduction approach.[78] This then allows for a large data set with multiple variables to be represented by a small number of values that together identify

the major patterns of variability within the set, these are named Principal Components (PC), and allows exploration of the sources of variation by examining vectors and contributions to variability.[79] In this way the entire cohort can be assessed visually, with members of the cohort with similar characteristics being clustered together on a PC Plot. It has, for example, been used in the assessment of gene expression of over 25,000 genes in 105 breast tumour samples, to identify cohorts of patients with similar gene expression profiles.[80] It should also be noted that PCA is a type of unsupervised analysis. That is to say that there is no pre-selection or prioritisation of groupings prior to analysis.[79] It is therefore a useful tool to identify important variables that discriminate members of a cohort in large data sets where the researcher wishes to explore for potential patterns of variance.

4.3: What is Random Forest Analysis?

A common goal in the assessment of big data is the need to classify the cohort into subgroups based on shared properties or variance. Random forest (RF) is an example of a classification method used in this setting.[81]

RF is itself based on a decision tree model.[81] Essentially a decision tree is a method used to separate a cohort based on population characteristics. At the top of the tree, the cohort will be split by the variables that best separate data (e.g. have the best discriminatory power). Each subsequent decision or node in the decision tree will continue to split the cohort by variables, but with less discriminatory power. The end-result will be the generation of discreet groups with similar characteristics. An example of this would be if one were to create a decision tree to classify ball sports. The first decision in the tree might be 'hand sport' or 'foot sport'. The hand sport branch could then be separated into racket or bat sports and so on. Decision trees are a supervised approach but can be prone to overfitting data.[82] That is to say that they are good at classifying data into predetermined groups but will do so even if the data is weak.

A distinct advantage of RF is that it avoids overfitting data. It uses an algorithm to decide which variables in a large data set (where there are more variables that number of patients in a cohort) are most likely to predict which of the predefined grouping a patient is likely to belong.

RF does this by initially testing a proportion of the sample or cohort (named the test set), and attempts to classify them using randomly generated trees with a limited number of decision points or nodes.[83] It then uses only the most successful limited node trees and retests the process on a confirmatory data set (drawn from the original data). This allows for assessment of the selected decision trees and allows for parameters to be tweaked for incorrect classifications, known as out of box errors. It is the use of multiple decision trees, coupled with the discrete test and confirmatory steps that avoid the issue of overfitting found in simple decision tree models.[83] RF is also able to rank which variables are most important in the classification of a patient into their predefined cohort.[83] RF has an established use in medical research, for example in the identification of important proteins in the life cycle of cells and for genetic association studies.[84,85]

Finally, logistic regression can also be used to classify big data, and can be used to predict if a particular individual in a cohort is part of a particular binary outcome, for example fit or frail, dead or alive. The issue with logistic regression in the setting of this research is that it requires all the variables to be independent. This data set contains some highly correlated variables including eGFR and creatinine, immunoglobulin and paraproteins. It is therefore not a suitable method of classification in this analysis.

RF therefore has been selected as the most suitable form of supervised analysis to identify our patient cohorts in this analysis, and will be discussed in more detail in chapter 2.

4.4: Is there a novel statistical approach to using big data in this myeloma cohort?

It is therefore possible to combine these techniques in a stepwise fashion. Firstly, PCA can identify cohorts of patients with similar characteristics that can identify patients with shared clinical or biological features. Random forest can then be used as a supervised approach using these distinct groups to analyse of other measurements, such as protein biomarkers that may identify the groups. This step wise PCA-RF approach has previously been used in the field of sports, to identify classify complex patterns of movement using big data sets of thousands of recordable variables.[86]

5: Chapter 1- Unsupervised analysis of clinical data in a cohort of multiple myeloma patients was able to objectively identify clinically significant sub-groups with different prognoses

5.1: Introduction

There are a variety of scoring and classification systems with which to evaluate multiple myeloma. There are those that seek to define prognostic subgroups (ISS[87], R-ISS[88]), those that attempt to classify degree of fitness or frailty (IMWG classification[89], ECOG performance score[90]) and those that quantity and classify co-morbidities (HCT-CI[91] and Charlson Comorbidity index[92]). These scores are a mixture of myeloma specific and generic scores, validated in the setting of multiple myeloma.

The advantage of these scores is that they allow clinicians to classify individuals into clinical subcategories. Do they have high-risk (thus early relapse risk) disease that requires intensive treatment regimens, or standard risk disease associated with longer progression free survival and overall survival? Do they have significant frailty and co-morbidities that will affect their ability to tolerate intensive chemotherapy, or are they able to tolerate intensive regimens and an autologous stem cell transplant?

While the R-ISS is score is recommended as standard of care in UK guidelines and is now routinely performed,[97] the guideline fails to reach a consensus on which frailty score (if any) to use. Consequently, there is variation in assessment of frailty and comorbidities out with specific clinical trials. This results in significant variability of patient care.

The goal for the clinician however remains to achieve the deepest remissions, with the lowest possible treatment related morbidity and mortality. The National Institute for Clinical Excellence (NICE) have approved the combination of daratumumab, bortezomib, thalidomide and velcade (DVT-d) as the standard of care in fit, transplant eligible patients,[14] on the basis of the data from the CASSIOPIA trial.[93] The addition of daratumumab improved progression free survival (PFS), with median PFS not reached in the daratumumab arm compared to 46.7 months in the observation arm. The CASSIOPIA study therefore achieves the aim of deeper and more durable remissions, but at the cost of higher morbidity, for example an increased number of patients suffering serious adverse events in the

daratumumab treatment arm (23% compared to 19%). Bringhen *et al* have previously demonstrated that drug discontinuation due to adverse events leads to increased risk of death (HR 1.67).[58] This demonstrates the delicate balance clinicians must tread with regards to treatment selection. There is no current nationally agreed algorithm or scoring system to aid this decision, instead the decision remains subjective.

If a novel scoring system in multiple myeloma was produced, it would be important for it to have two important properties. First, it must be informative to the treating clinician. It should provide clinical information on prognosis, treatment related morbidity and mortality or likely-response to therapy. But it must also simple, not significantly time consuming and be affordable (i.e. use standard of care results and measurements). The currently available prognostic scoring system, the R-ISS provides statistically significant prognostic data,[42] and uses standard of care investigations. However, it relies also on cytogenetic testing of bone marrow aspirate, which not always feasible in patients. Currently used frailty and co-morbidity assessments are more time consuming, relying on subjective assessment or patient reporting. While they are not expensive to perform, their use is limited by time pressures in clinic. A recent survey of oncologists showed that only 52% performed a frailty assessment in the majority or all of their oncology patients,[94]with one-time pressures cited as one of the reasons for the low percentage.

This chapter therefore draws on the standard of care data collected for a real-world patient cohort at a tertiary myeloma centre, over a period of 35 months. It explores and validates the pre-existing scoring systems mentioned, both in newly diagnosed and relapsed patients. It also explores the data using unsupervised analysis, seeking to identify new clinical subsets of patients that can be identified to improve current standard of care prognostic and vulnerability assessments.

5.2: Aim

To identify patterns of clinical and laboratory tests in patients with newly diagnosed or relapsed multiple myeloma using baseline standard of care investigations, clinical prognostic tests and frailty scores. The aim was to identify clinical meaning from the data; in particular, we aimed to identify patterns of variation that might reflect fitness or frailty.

5.3: Objectives

1. To compare the characteristics of this cohort of patient with reported cohorts to determine how well they represented overall myeloma populations
2. To perform exploratory statistics to identify cohorts of patients with distinct shared clinical features
3. To assess those clinical variables that underlay the distinctive groups to assign them a potential clinical meaning
4. To determine whether the groups had distinctive features of survival

5.4: Methods

Criteria for Selection

Data collection commenced in March of 2017; the last patient was consented for data collection in February of 2020. The data collection was curtailed at the onset of the COVID-19 pandemic. Overall, data collection was for a total of 35 months.

Data collection

Primary data collection was undertaken by a haematology clinical-research nurse with experience in the field of plasma cell dyscrasias. Data collection was prospective using an Excel spreadsheet kept in a password-encoded electronic folder. Extensive data were collected at baseline, with interim data collected at two monthly intervals, until progression of disease, death, or withdrawal of consent. A more detailed set of data point were collected at six monthly intervals.

If a patient relapsed during the study period and they consented to remain in the project, they were re-screened for all baseline data and re-entered as a new analysis number. They were again followed up until progression, withdrawal of consent, death or censoring date. This dataset will be made available for peer review purposes on completion of the project.

Ethics and consent

The project was conducted under the governance framework of Manchester Foundation Trust (MFT).[95] Each patient was consented at the start of each cycle and written consent forms have been securely stored for each. ETHICS NUMBER IRAS Project ID 209727 (Study PIN R04443)

Characteristics of the data set

The data collected was a mixture of qualitative and quantitative (continuous and discrete) data points. Table 3 outlines all the data collected for each of the enrolled patients, at baseline and at subsequent time points. Some of the data points (e.g. R-ISS Score) consist of a combination of clinical parameters. These were calculated using the relevant scoring

system, using publicly available online calculators such as 'MedCalc'. Specific online tools used are detailed below.

Table 3: Data points collected for each patient in data collection phase of study at the three time points

Data collected at baseline	Data collected at 2 monthly intervals	Data collected at 6 monthly intervals
Qualitative data		
Quality of life questionnaire		Quality of life questionnaire
Patient Demographics		
Age at start of cycle of chemotherapy	Change in medical history	Change in medical history
Height		Height
Weight		Weight
BMI		BMI
Gender		
Date of diagnosis		
Age at diagnosis		
Comorbidities		
Clinical features of disease		
List of co-morbidities		
Number of comorbidities		
Cardiovascular comorbidities		
Pulmonary comorbidities		
Endocrine comorbidities		
Existence of previous cancer		
Medications (names and number)		
Scores		
Performance Status (ECOG)	Performance status (ECOG)	Performance status (ECOG)
Charlson Comorbidity Index		Charlson Comorbidity Index
ADL Score		ADL Score
IADL Score		IADL Score
IMWG Frailty Score		IMWG Frailty score
HCT-IC score		
Myeloma specific details		
Paraprotein level	Kappa and lambda light chains	Kappa and lambda light chains
Kappa and lambda light chains	Paraprotein level	Paraprotein level
Isotype of myeloma		
Percentage of plasma cells on bone marrow aspirate		
Percentage of plasma cells on trephine		
ISS Stage		
R-ISS Stage		
FISH results		
Standard of care test results		
Haemoglobin	Haemoglobin	Haemoglobin
Neutrophils	Neutrophils	Neutrophils

Platelets	Platelets	Platelets
Creatinine	Creatinine	Creatinine
eGFR	eGFR	eGFR
Corrected calcium	Corrected calcium	Corrected calcium
CRP	CRP	CRP
Urine protein/ creatinine ratio	Urine protein/ creatinine ratio	Urine protein/ creatinine ratio
LDH	LDH	LDH
D-dimer	D-dimer	D-dimer
NT-Pro BNP		NT-Pro BNP
HbA1c		HbA1c
Albumin		
B-2- microglobulin (B2M)		
Treatment details		
Date commenced treatment	Response to treatment	Response to treatment
Name of treatment regimen	Change in medications	Change in medications
Line of treatment	Change in chemotherapy dose	Change in chemotherapy dose
Functional assessments		
Gait speed x 3 (and median)		Gait speed x 3 (and median)
Grip strength x 3 (and median)		Grip strength x 3 (and median)

This table summaries the data collected from the cohort of patients included in the study. Data was collected at baseline, two months and six months. The data collected can be classified as either: qualitative, demographic, comorbidity related, myeloma specific, standard of care tests, calculated established scores and functional assessments. Subsequent analysis in this chapter will be based on these data

Data analysis approach- Descriptive

Descriptive data analysis used Excel and GraphPad Prism software. Baseline overall analysis of survival employed the log rank test to compare survival curves for statistically significant differences. The Log-rank test was the preferred method as it is a powerful statistical test and these data fulfil the assumption of proportional hazards. It also gives equal weighting to deaths at all time points. This contrasts with the Gehan-Breslow Wilcoxon Test, which gives extra weight to early deaths.

Data analysis approach- Exploratory

PCA was selected as the initial method of analysis for baseline data. This statistical methodology allows for analysis of large data sets and is an example of unsupervised data analysis. This is to say that the analysis will identify patients who have common

characteristics, with no pre-selecting of groups. One way ANOVA was used to detect statistically significant differences between PCA subgroups.

Preparation of data set Part 1: checking and completion

Following initial data collection, the dataset was reviewed to identify missing or potentially inaccurate information. These data were subsequently corrected where possible to ensure an accurate dataset. This was done by a systematic review of a variety electronic patient records and clinical documentation.

If an individual patient had a missing data point at baseline, data from the next available clinical interaction was the preferred method of data collection. If this data was also not available then the mean value for that individual IMWG cohort was used (fit, intermediate, frail). If an individual patient had more than 4 data points at baseline missing, they were excluded from the analysis.

Table 4 outlines the detailed methodology for how each of the variables included in the analysis were assessed. Specifically, it states how clinical information was obtained and checked, along with specific data handling principles. Where appropriate, it also references how and where the clinical scores were calculated. Furthermore, it highlights if any data is missing for each of the variables assessed.

Table 4: Overview of data handling and explanation of 'imputed' values in primary data collection

Variable measured	Data Handling
International Myeloma Working Group Frailty Score	All data available and calculated on all patients (Appendix 1) Calculated using: http://www.myelomafrailtyscorecalculator.net/Geriatric.aspx
Age of patient	All data available
Line of treatment	All data available- confirmed with electronic clinic records
Comorbidities	Data available and cross checked with clinic letters. Rules used: treated conditions that were fully resolved at the time of diagnosis and unlikely to contribute to frailty were not included. Examples include treated B12/ Vitamin D deficiency. These were listed as descriptive data but not included in the numerical co-morbidity count. If there was a possibility co-morbidity could have an impact on current health status, it was included.
Pulmonary comorbidites	All data available. Classified as none/ mild (CTCAE 0-1) or moderate/ severe (CTCAE 2-4).[96]
Haematopoietic cell transplantation comorbidity index (HCT-CI)	All data available. Calculated using: https://qxmd.com/calculate/calculator_108/hematopoietic-cell-transplantation-specific-comorbidity-index-hct-ci
Number of medications	All data available. Medication list on clinic/ discharge letters used. Where patients were newly diagnosed, supportive myeloma medications e.g. VTE prophylaxis, anti-viral medication not included.
Albumin	All data available and cross checked with electronic records
B2M	Data for 2 patients missing. • Both had subsequent B2M measurements. These data points used.
ISS	All data available and cross checked • ISS calculated for 2 patients using subsequent B2M data.
LDH	Data for 6 patients missing • 5 patients had subsequent LDH performed at time point +2 months. These data points used • 1 patient did not have subsequent LDH performed. Extrapolated LDH value used- Median score for patient in FIT IMWG population used (patient IMWG Fit)
Haemoglobin	All data available and cross checked
Neutrophils	All data available and cross checked
Platelets	All data available and cross checked
Creatinine	All data available and cross checked

eGFR	All data available and cross checked • Creatinine clearance would have been preferred metric but insufficient data available to calculate this parameter
Corrected calcium	All data available and cross checked
CRP	All data available and cross checked • Patients with CRP '<1' converted to CRP 1
NT-Pro BNP	Data for 4 patients missing • 3 patients had subsequent NT-Pro-BNP measurement. These data points used. • 1 patient had no subsequent NT-Pro-BNP performed- extrapolated value used. Median score for FIT IMWG population used (patient FIT IMWG)
Paraprotein level	All data available and cross checked. • If no paraprotein seen/ paraprotein too low to quantify/ paraprotein only seen on immunofixation- value of 0 given.
Pathological serum free light chain level	All data available and cross checked.
D-dimer	Data for 6 patients missing • All 6 patients had subsequent D-dimers performed. These data points used
ECOG Performance status	Data for 7 patients missing • Performance status estimated based on clinic letters, subsequent performance status and from ADL/ IADL scores • ECOG score outlined in Appendix 2
Average gait speed	Data for 6 patients missing • Imputed speeds used based on the median gait speed for the IMWG frailty group the patient belongs to. Performed using Excel software • If patient unable to perform test, speed of 0 given.
Average grip strength	Data for 2 patients missing • Imputed strength used based on the median grip strength for the IMWG frailty group the patient belongs to. Performed using Excel software • If patient unable to perform test, speed of 0 given. If patient unable to perform test, strength of 0 given.
Cytogenetics	Too many data points missing to include in PCA • Data available for 63 patients • Cytogenetics from diagnosis used • Risk stratified using BSH Guidelines on risk stratification[97]
R-ISS	Too many data points missing to include in PCA (used as a label) • Data available for 73 patients • R-ISS calculated without cytogenetics and or other missing data points and reported only if missing data would not change R-ISS score • Note: R-ISS score uses just del (17p), t(4;14) and t (14;16) as high risk abnormalities[102]

Charlson Comorbidity Index (CCI)	All data available and cross checked with electronic records Calculated using: http://www.myelomafrailtyscorecalculator.net/Default2.aspx
Activities of daily living (ADL)	Data for 1 patient missing • Estimated using subsequent data and clinical letters Calculated using: http://www.myelomafrailtyscorecalculator.net/Default2.aspx
Instrumental activities of daily living (IADL)	Data for 1 patient missing • Estimated using subsequent data and clinical letters Calculated using: http://www.myelomafrailtyscorecalculator.net/Default2.aspx

This table lays out how data was handled and calculated, including how missing values were managed. It sets out the rules that explain how and why some values were 'imputed' to allow subsequent analysis. If a score calculation is required, the source of the calculation is stated.

Preparation of Data Set Part 2: Limitations and Adjustments

PCA requires all data points for all variables to be available. Any specific variable had more than 10% of data points missing was excluded form PCA analysis. Variables excluded from PCA analysis include: cytogenetics, R-ISS, urine analysis, HbA1c, percentage plasma cells on aspirate and trephine and body mass index. Where less than 10% of data points was missing for an individual variable, a description as to how the 'imputed value' was produced is given in Table 4. The mean imputation involves using the mean of all known values and estimating the value of the missing data. While this method has drawbacks, it does not have a significant impact on the overall PCA results and is a recognised method of data management.[98]

For PCA, the clinical continuous variables were converted to $\log^2$ values. This was to reduce the impact of outliers and to normalise any skew of data. Log-transformed variables included: a: albumin, beta-2-microglobulin, LDH, Hb, neutrophils, platelets, creatinine, eGFR, corrected calcium, CRP, NT-Pro-BNP, paraprotein, pathological serum free light chain, D-dimer and gait speed. Scores using cumulative variables were included in the PCA. This is because these scores have been previously verified as prognostic in previous papers.[40,42,45]

PCA treats all variables as equal, without weighting them in terms of clinical importance. PCA does however normalise the data to a mean of 1 standard deviation +/- 0.5. This is to allow equal treatment of all values, irrespective of range.

Unfortunately, longitudinal follow-up data collection was suboptimal with many missing data points. The reason for this is likely to be multifactorial and includes patient factors such as mortality and withdrawal of consent, along with external factors such as limitations imposed by the COVI-19 pandemic. The lack of follow-up data precludes all but basic descriptive analysis.

5.5: Results

5.5.1: Baseline demographics of our cohort were similar to existing published studies

91 patients from the cohort met the criteria for the study. Five had two entries since they relapsed during the active follow-up time of the study. Of these, 57% were male and 43% being female. The population had a mean age of 67 years at enrolment, with a range from 38 to 87 years. The detailed breakdown of demographics of this cohort is given in Appendix 3. Table 5 compares the Manchester research cohort to historical cohorts and demonstrates that they are comparable in terms of disease, demographic and vulnerability status.

Table 5: Comparison of baseline characteristics of our cohort compared to historic cohorts

	Manchester Royal Infirmary Cohort (n= 92)		**Engelhardt UKF Cohort[100] (n=125)**		**Palumbo IMWG cohort[99] (n= 869)**	
	% of patients	**Median (IQR/range)**	**% of patients**	**Median (IQR/range)**	**% of patients**	**Median (IQR/range)**
Age (years)						
≤ 65	41	67	59	63 (56-71)	2	74 (70-78)
66-74	32		26		52	
75-79	13		15		46	
> 80	14		3		19	
Creatinine, mg/dl						
< 2 (< 177μmol/l)	79	1.06	85	1 (0.80-1.40)	92	0.98 (0.80-1.22)
> 2 (> 178 μmol/l)	21		15		5	
Missing					3	
ECOG PS						
0	36	1	22	1	30	1
1	40		50		46	
2	15		26		19	
3	8		2		2	
4	1		0		0	
ISS						
I	21	2	28	2	28	2
II	36		34		42	
III	43		38		31	
Chromosomal aberrations						
Standard	48		51		38	
Unfavourable	20		32		24	
Missing	32		17		17	

ADL		6 (5-6)		4 (4-5)		6 (5-6)
> 4	92		48		86	
≤ 4	8		52		14	
IADL		7		8 (6.4-8)		8 (6-8)
> 5	70		85		82	
≤ 5	30		15		18	
CCI		2		2 (1-3)		0 (0-1)
< 2	54		35		83	
≥ 2	46		65		17	
IMWG Classification						
Fit	38		18		39	
Fit-intermediate	26		34		31	
Frail	36		48		30	

This table compares my research cohort (MRI cohort) to two published cohorts of myeloma patients (Engelhardt et al[100] and Palumbo et al [99]) in terms of demographics, disease status and vulnerability.

Legend: ADL- Activities of Daily Living score; CCI- Charlson Comorbidity Index; ECOG-Eastern Cooperative Oncology Group Performance status; IADL- Instrumental Activities of Daily Living score; IMWG- International Myeloma Working Group Frailty Score; ISS- Multiple myeloma International Staging System

5.5.2: Our population of multiple myeloma patients were representative and using validated outcome scores

Preliminary overall survival analysis was performed using log rank testing of each of the subgroup variables. Statistically significant differences in the overall survival were demonstrated using the Multiple Myeloma International Staging System (ISS) groups (p =0.05), ECOG performance status (PS) groups (p= 0.05), Activities of daily Living (ADL) groups (p= 0.04), Revised ISS (R-ISS) groups (p=0.001) and cytogenetic risk groups (p= 0.03). The log rank analysis of survival curves of all the different assessments is shown in Table 6, while Figure 2 shows the Kaplan-Myer plots for selected subgroup survival analysis.

Table 6: Log rank analysis of overall survival using different baseline characteristics and clinical risk scores in our cohort

Score/ investigation	Groups Compared	Statistically significant difference in OS	p value
IMWG Frailty Score	Fit vs Intermediate Vs Frail	No	0.36
Age	< 75 vs 75- 80 vs > 80 years	No	0.55
Age	< 60 vs 60-69 vs > 70 years	No	0.24
ISS *	1 Vs 2 Vs 3	Yes	0.05
Renal function (eGFR)	>90, 60-90, >60	No	0.1
Respiratory disease	None/ mild Vs Mod/ severe (ECOG)	No	0.44
Charlson Comorbidity Index (CCI)	0-1 vs 2 vs 3 vs 4+	No	0.24
Charlson Comorbidity Index (CCI)	0-2 vs >3	No	0.22
ECOG performance status (PS) *	0 vs 1 vs 2 vs 3 or 4	Yes	0.05
Pro-BNP	Normal range vs above ULN	No	0.06
Activities of daily living (ADL) *	> 4 vs </= 4	Yes	0.04
Instrumental activities of daily living (IADL)	> 5 vs </= 5	No	0.29
HCT-CI	< 2 vs >/= 2	No	0.51
Plasma cell percentage in marrow	< 60% Vs >/=60%	No	0.11
R-ISS*	1 Vs 2 Vs 3	Yes	0.001
Cytogenetics/ FISH of CD138 selected cells*	Standard Vs Unfavourable (as defined by Palombo et al[42])	Yes	0.03

*This table demonstrates that there is a statistically significant overall survival in this cohort, when compared by ISS, ECOG performance status, ADL, R-ISS and cytogenetic risk group using T-testing. * Denotes significant difference between groups (p= ≤ 0.05)*

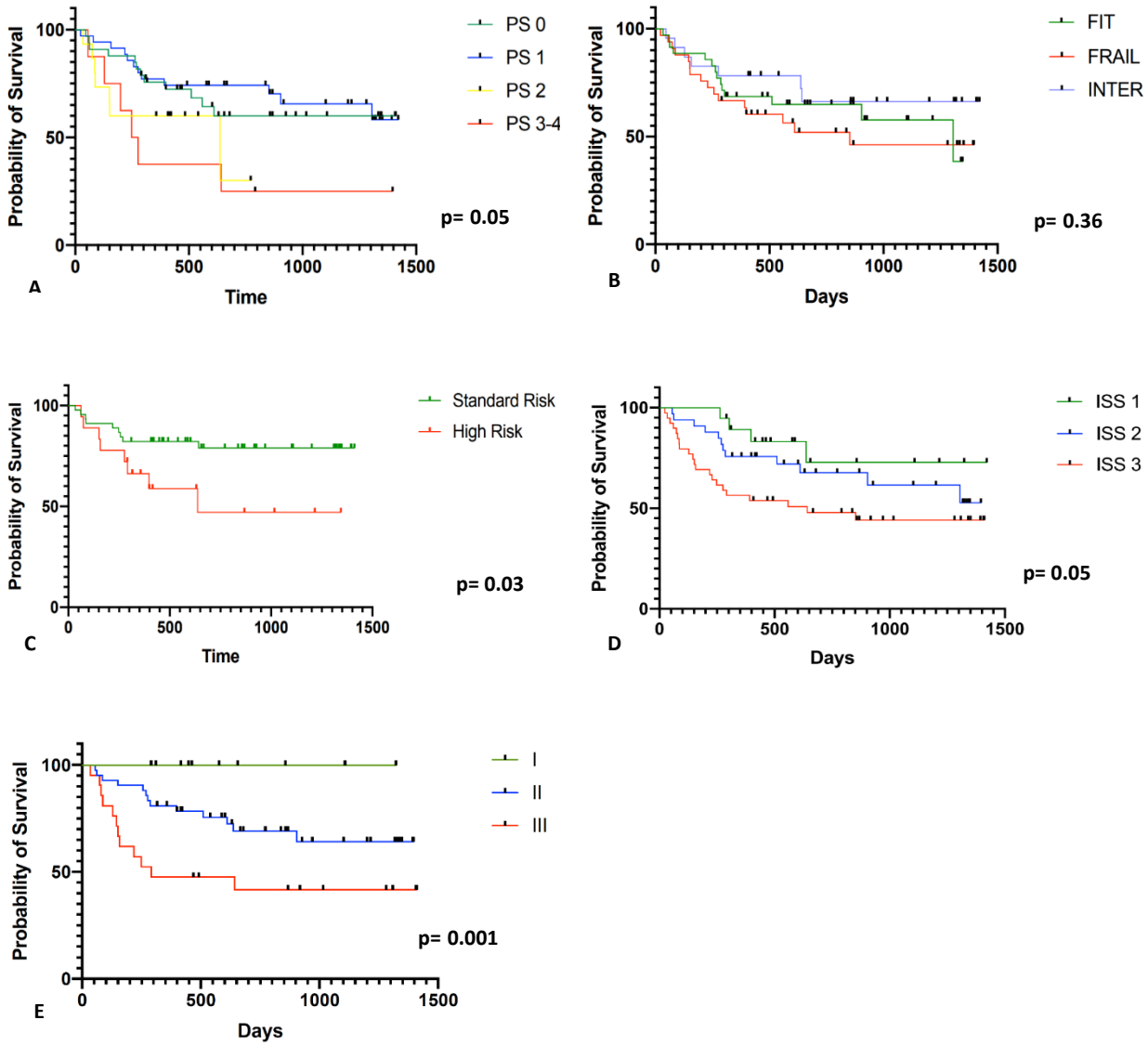


Figure 2: Overall survival curves comparing selected subgroups: (A) ECOG performance status groups, (B) IMWG Classification, (C) Cytogenetic risk groups (D) ISS groups (E) R-ISS groups

These survival curves use log rank analysis to test the null hypothesis that the difference in survival of the subgroups is identical, and that any differences in survival are due to chance. For those analyses where there are three or more groups, a p value of < 0.05 suggests that there is a significant difference in survival in the three groups. This figure shows that there is a difference in survival in the ECOG PS subgroups, the ISS groups, the R-ISS subgroups and the cytogenetic risk subgroups. There is however no difference in the survival between the IMWG subgroups.

5.5.3: PCA Section 1: Exploring Groupings produced by Unsupervised Analysis

Unsupervised PCA of the cohort identifies novel clinically significant subgroups with distinct clinical features

The clinical and laboratory datasets collected at baseline for each individual in the study reflect processes relating directly to the activity of the myeloma disease (myeloma disease burden), but also measurements relating to their fitness or frailty. The initial purpose of PCA was to produce PC plots to identify groups of patients with common characteristics. The eigenvectors were then used to infer which variables were most important in explaining the clustering on the PC plots.

Essentially, PCA uses the covariance of different components in the dataset to group individuals who have similar characteristics. A PCA separation applied to all patients in the data set (Figure 3), with each individual identified as a circle, and separated according to the two most important elements of shared variability within the dataset: PC1 and PC2. The contribution of each variable to these PCs is shown in the Eigenvector Plot (Figure 4), explaining the contribution of each variable to the separation of the data.

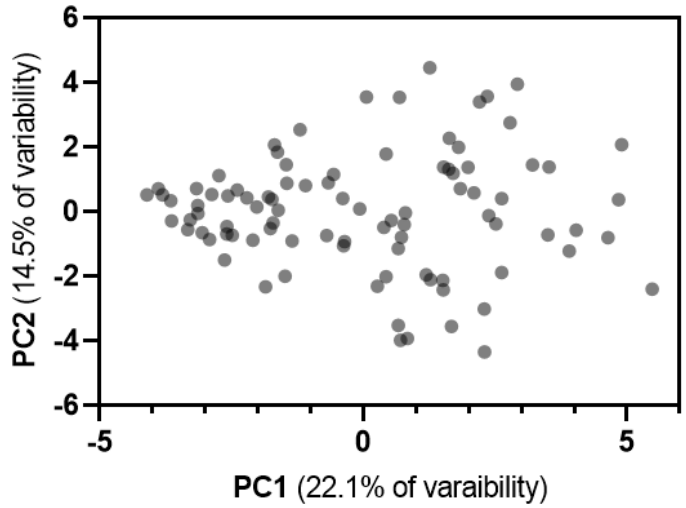


Figure 3: The Principal Component Analysis applied to all variables in the study

Each of the 91 cases is represented by a circle and is plotted according to its value for PC1 and PC2. 22.1% of the variability in the data set is explained by PC1, with 14.5% explained by PC2.

Figure 4 shows that cases are separated by PC1 according to the combined effects of every variable, but with different relative contributions indicated by the length of the vector line. In contrast, the vectors that contribute to PC2 appear to separate patients according to specific processes determined by their direction, with a direction above the central axis mainly comprising measurements relating to myeloma disease (labelled a-m), and those below the axis relating to intrinsic health or frailty (labelled n-z).

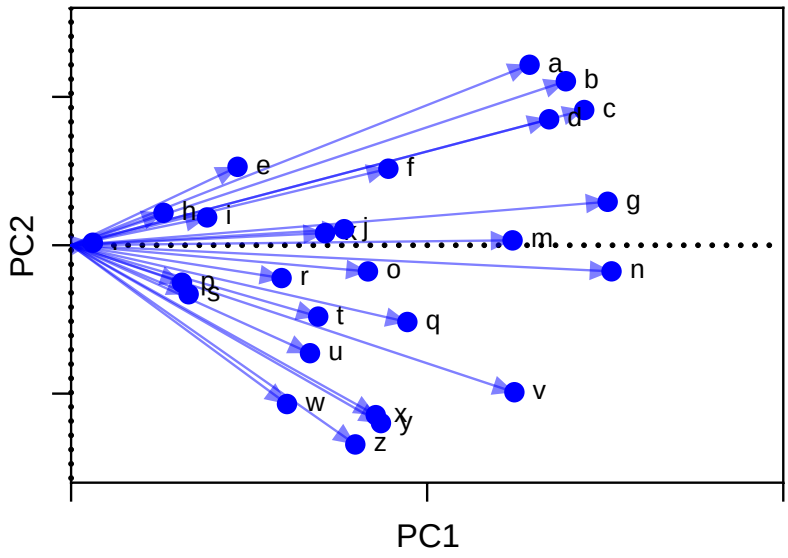


Figure Key: Variable associated with each of the letter in Figure 4			
Creatinine	a	D-dimer	o
eGFR	b	Corrected calcium	p
B2-microglobulin	c	Number of comorbidities	q
ISS Stage	d	Albumin	r
LDH	e	Line of treatment	s
pathological SFLC	f	Number of medications	t
NT-proBNP	g	Age	u
Platelets	h	IMWG frailty	v
Paraprotein	i	ADL Score	w
CRP	j	Gait speed	x
Haemoglobin	k	ECOG-PS	y
HCT-CI Score	m	IADL Score	z
Charlson Comorbidity Index	n		

Figure 4: Eigenvector plot showing how each different variable contributes to the separation by PC1 and PC2 (loadings)

This figure shows the relative importance of each of the listed variables to the separation by PC1 and PC2. The further the extension of the arrow along the x-axis, the more important the viable is to the separation by PC1. In this figure, NT-Pro BNP and Charlson Comorbidity Index are the variables with the most significant impact on PC1. The more divergent the axis from the x-axis (in either positive or direction), the more important the variable is to separation by PC2. In this figure, creatinine and IADL are the most important variables in the separation by PC2. The relative contribution to PC1 and PC2 of each variable is further assessed in figures 5 and 6.

Analysis of the vectors that underlie the separation reveals that the major component of variability, shown on the X-axis (PC1), separates cases according to their overall level of abnormality for each of the clinical and laboratory measures. Those patients who have relatively normal values ('myeloma: clinically well') lying on the left of the axis, and cases with mainly abnormal overall results lying to the right. This can be seen in Figure 5. The second most important component of variability (PC2), separates the cohort in a different way. The eigenvectors suggest this is based on different biological processes. Figure 5 is a series of heat maps that demonstrate how PCA in this data set can be useful to separate the cohort of patients into groups with similar co-variances, underpinned by measurable biological processes.

Figure 5A shows the PCA of the entire data set, but this time with individual patient points colour coded by platelet count. Patients with a low platelet count are cloured red and those with elevated platelet counts in light green, with those in between on a graded scale between the two. In this case there is no obvious seprartion of platelet count in each of the quadrants. There is no statistically significant difference in values between those patients with positive values for PC1 against those with negative values, or between those with positive vales for PC2 against those with negative values. This shows that biological variability of platelet count in newly diagnosed or relapsed myeloma is not a significant factor in the PCA undertaken.

However, a different and significant pattern can be seen in Figure 5B. This is again a PCA of the cohort, but this time with each patient colour coded by Charlson Comorbidity Index Score (CCI). Lower scores are associated with a reduced number of comorbidities. Visually there is a clustering of green to the left of the vertical bisector through the PCA 0 point, indicating lower CCI scores in patints to the left of this line (regardless of positivity or negativity on PC2). There is a statistically significant difference between the CCI of the patients in the area of the figure labelled 'well' compared to the area in the upper quadrant on the right of the figure labelled 'high disease burden'(p= 0.0013). There is also a statistically significant difference between the area of the graph labelled 'well' when compared to the area in the lower quadrant on the right of the figure labelled 'frailty' (p= 0.0004). This demonstrates the first pattern of separation by PCA: *patients who have*

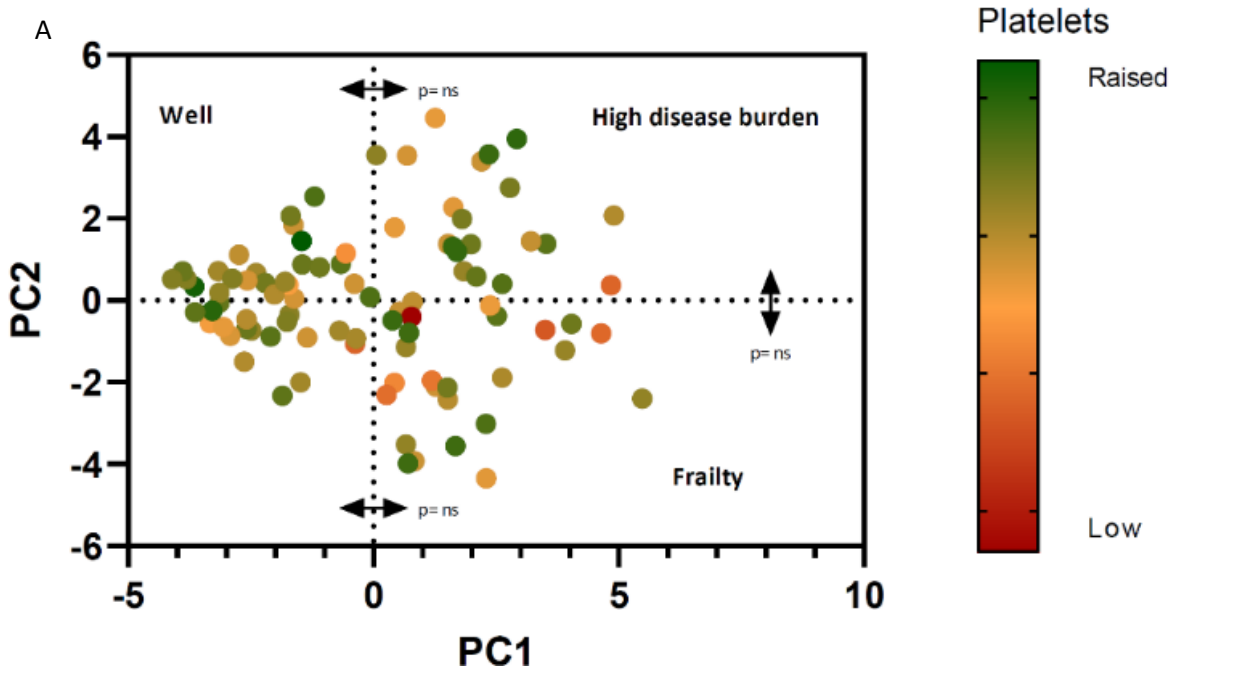
disease with minimal adverse features are statistically likely to have low values for PC1 and be found the the left of the figure, regardless of their value for PC2.

The second pattern observed can be demonstrated in Figure 5C, where patient are colour coded by gait speed. Visually patients with a slower gait speed (red) are clusted in the area of the figure labed 'frailty'. These patients have a positive value for PC1 and a negtive value for PC2 (bottom rihgt quadrant). Using gait speed as a surrogate marker for frailty, we can show that there is a statistically siginficant difference in the gait speed between the 'well' group and the 'frailty' group (p= 0.0155) and between the 'high disease burden' group and the 'frailty group' (p= 0.0349). The second pattern demostrated is: *patients with a high burden of frailty are statistically likely to be found in the right lower quadrant of the PCA figure, with a positive PC1 value, but a negative PC2 value.*

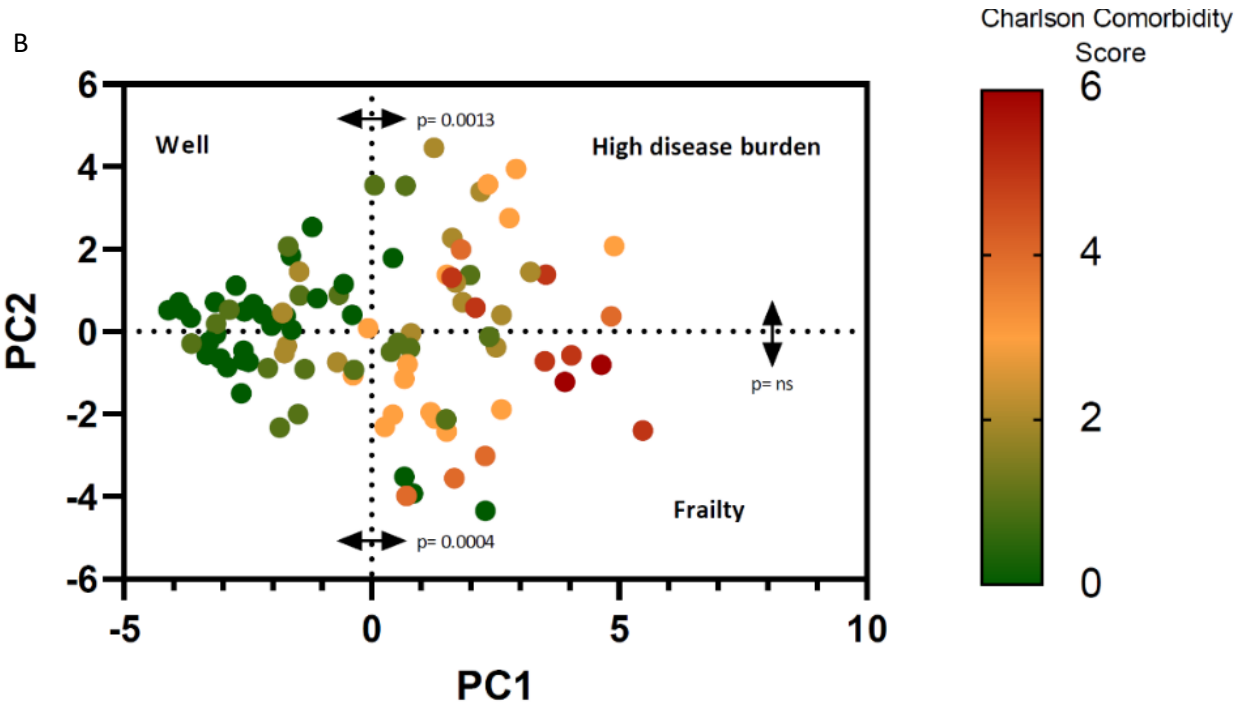
The third pattern that can be observed is with regards to disease burden. Here we use B2M as a surrogate marker of disease activity (Figure 5). In this figure there is a visual trend that patients with a high B2M (red) are located in the upper right quadrant, with a positive PC1 value, and a postiive PC2 value. There is a stastically significant difference between the B2M values for the patients in the 'high disease burden' area, when compared to either the 'frailty area' (p = <0.0001) and the 'well' area (p= 0.0001). The third pattern demonstrated is: *patients with a high burden of disease are statistically likely to be found in the upper right quadrant, with a positive PC1 value and a positive PC2 value.*

These four variables were chosen as illustrative examples for the different patterns observed when PCA was used. Similar patterns were found with other variables but are not shown here.

A


Platelets
Raised
Low
Well
High disease burden
Frailty
p= ns
p= ns
p= ns
PC1
PC2
6
4
2
0
-2
-4
-6
-5
0
5
10

B


Charlson Comorbidity Score
6
4
2
0
Well
High disease burden
Frailty
p= 0.0013
p= ns
p= 0.0004
PC1
PC2
6
4
2
0
-2
-4
-6
-5
0
5
10

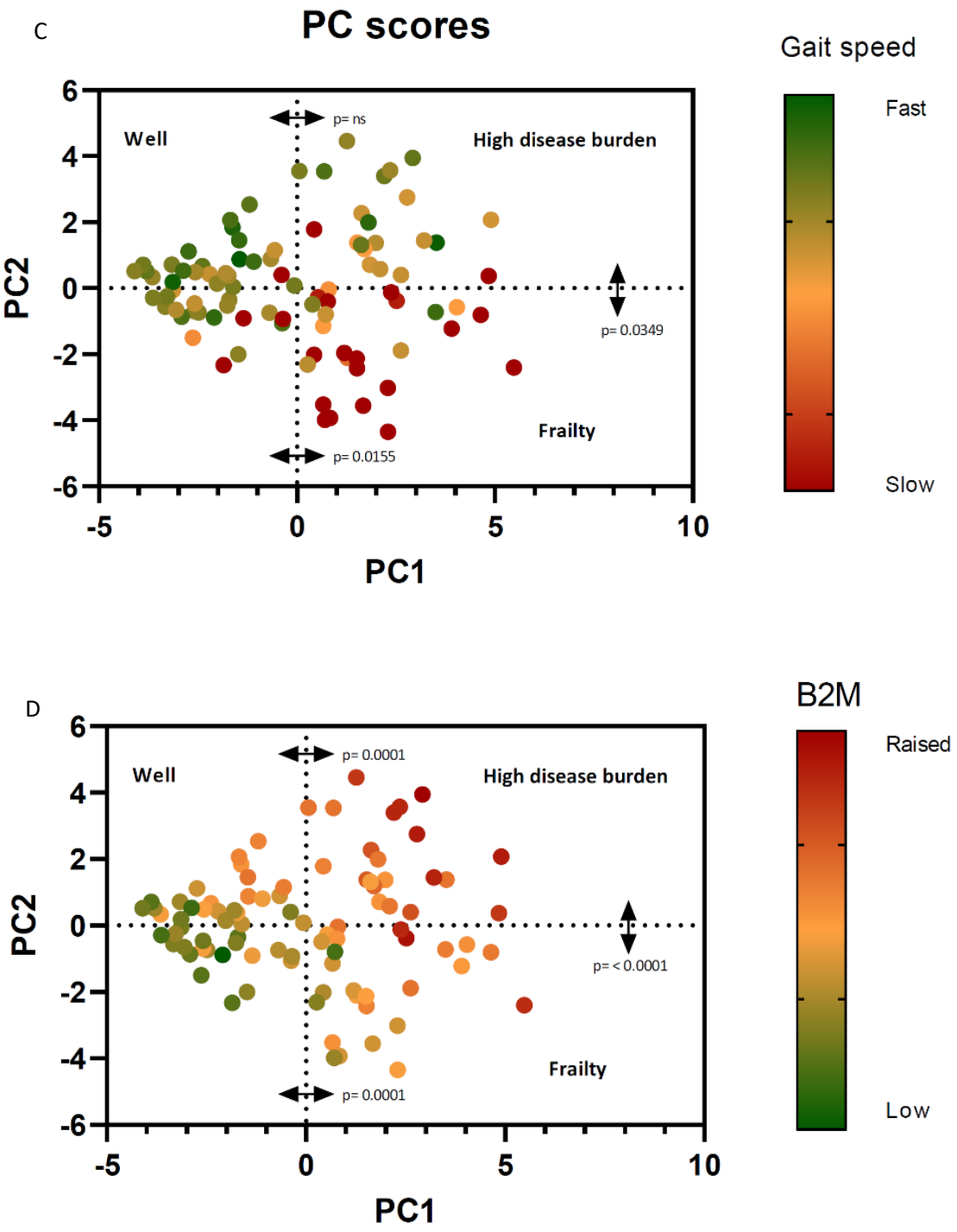


Figure 5: Distribution of clinical parameters by PC1 and PC2 (A) Distribution by platelet count; (B) Distribution by Charlson Comorbidity Index; (C) Distribution by gait speed; (D) Distribution by B2M

5A shows that there is no obvious pattern of distribution of platelet count by either PC1 or PC2, suggesting that platelet count is not a significant factor in the separation of patients.

5B shows a clustering of green dots (representing individual patients) to the left of the x-axis.

This visually demonstrates that patients with lower values for PC1 have a lower CCI score. In 5C there is a clustering of green dots (patients with fast gait speeds) at the lower end of the x-axis. This suggests that patients with more physiologically normal gait speed have lower values for PC1. In 5D there is a clustering of green dots, representing patients with low B2M levels to the left of the x-axis. This is a visual demonstration that patients with a low B2M (a physiologically normal level), have low values for PC1 and are clustered together.

Summary of PCA Part 1

Using surrogate markers for disease activity and frailty we have shown that PCA is able to statistically significantly separate our cohort of patinets into those who have minimal adverse features (well) or have a high burden of disease or frailty. This is summarised in the Figure 6.

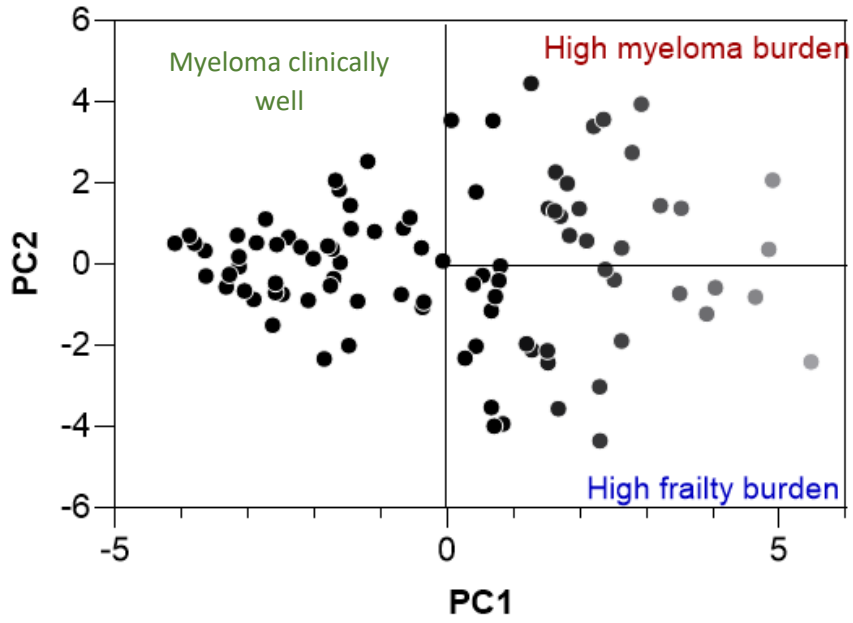


Figure 6: PCA analysis of whole cohort showing patients can be grouped into clinically meaningful categories using PC1 and PC2

This figure shows the PCA splitting the cohort into three distinct groups with simlar clinical characteristics. The group with low values for PC1 are a well group with low disease burden, the group with high PC1 and PC2 values are a cohort with high disease burden. The group with high PC1 values but low PC2 values are a clinically frail group according to conventioal assessment methods.

5.5.4: PCA Section 2: Validating and explaining groupings produced by unsupervised analysis

The major component of variability in the dataset (PC1) involves all data elements and correlates with disease outcome in myeloma

The clinical significance of the separation of cases by major component of variability (PC1) was assessed. Cases were separated into two groups according to their position on the PC1 axis (Figure 7). The PC1-Low group (n= 46) were those patients to the left of the bisector traveling through the zero value on the PCA graph, to the left of the axis. This patient group had relatively normal clinical and biological features. The PC1-HIGH group (n=45), have relatively abnormal clinical and biological features and are located to the right of the axis. The line that bisects zero on the x-axis was chosen as an arbitrary cut off, with the useful property that it splits the cohort into almost exactly in half numerically.

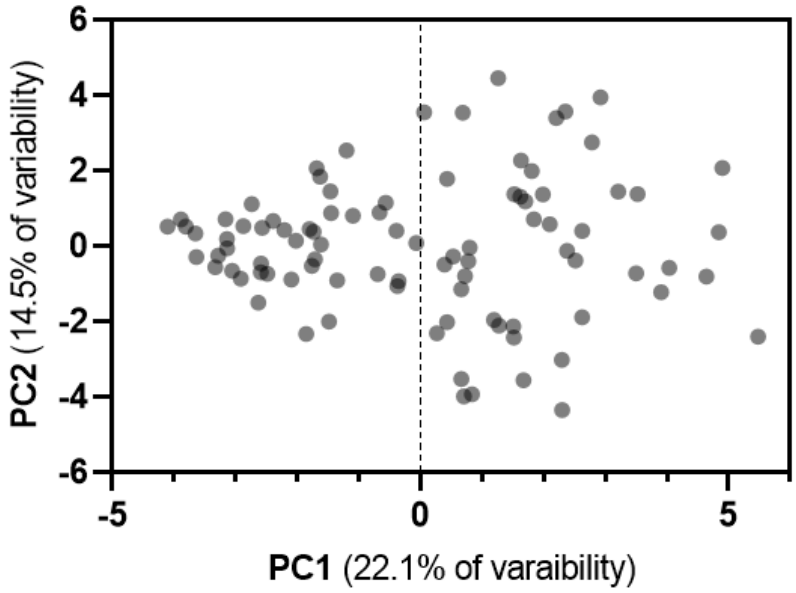


Figure 7: Separation of patients by PC1 into two subgroups

This figure separates the cohort of patients into two groups along the PCA axis- PCA-1-LOW and PCA-1 HIGH. Each dot represents an individual. There is no distinct grouping, but rather a spread of values for PC1. PCA-LOW (n=46), PCA-HIGH (n=45)

These two groups showed highly significant differences in survival. The PC1-HIGH group had a signinificantly worse overall survival (hazard ratio 3.6, p<0.0001) when compared to the

PC1-LOW group. The Kaplan Meier curve shows the survival of the two groups (Figure 8).

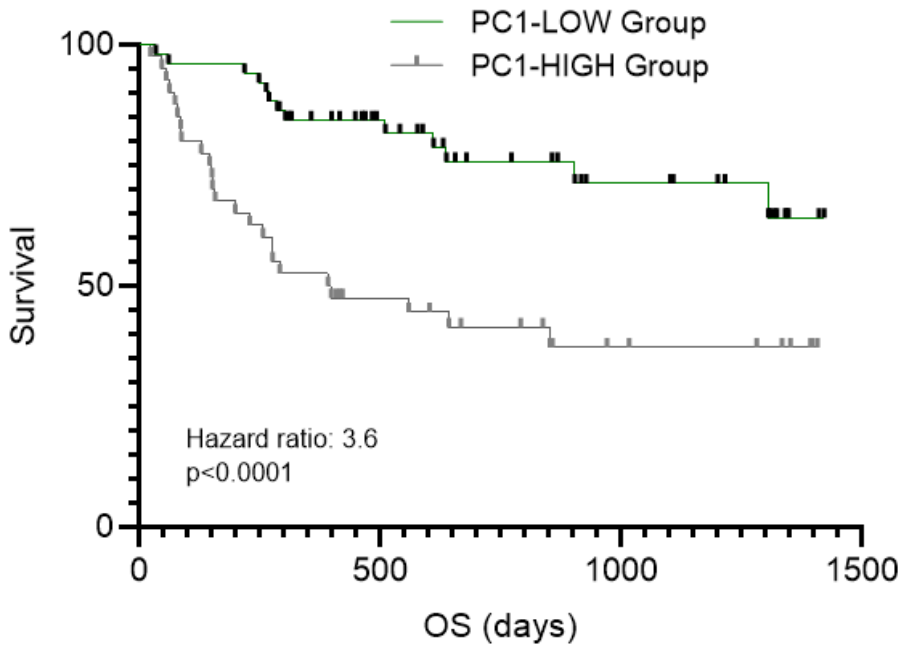


Figure 8: Kaplan-Meyer Analysis of overall survival of cases according to PC1

This figure shows the survival curves of the two PCA groups, based on their PC1 values (PC1-LOW and PC1-HIGH). Those in group PC1-LOW have the longest overall survival, with those in PC1-HIGH, having a shorter overall survival. The p value for difference in overall survival is highly significant at <0.0001. This demonstrates that patients with a high PC1 value have an inferior overall survival to those with a low PC1 value.

This separation was more significant than any other single element in this dataset (Table 6).

The relative contribution of each clinical and laboratory measure to the separation of PC1 showed that all elements in the dataset contributed to the separation, but with different relative weighting for their contribution (Figure 9). Importantly, those biological processes that contributed most significantly to PC1 (dark grey bars on the figure) appeared to reflect a range of disease processes including markers of myeloma disease burden (ISS stage,

β2microglobulin), organ function (creatinine, eGFR, pro-nitro BNP) or frailty processes (CCI, IMWG-frailty).

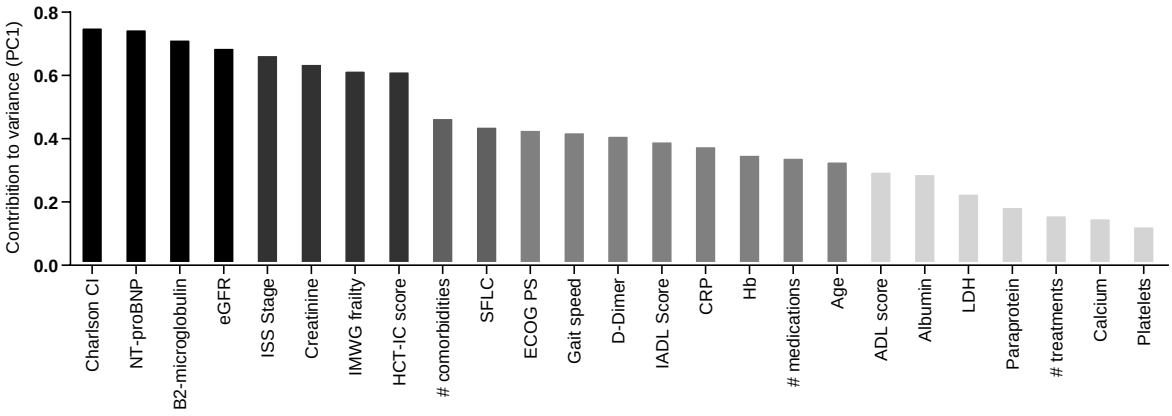


Figure 9: The contribution of each variable to the separation of patients by PC1

This figure shows the relative importance of each of the variables to PC1, derived from the Eigenvector values. Those variables with the black bars contribute most, with those in light grey least. There is no clear pattern as to which category of variable (frailty markers, disease markers, co-morbidities) causes most separation by PC1.

PC2 divides cases according to measures reflecting distinct processes related either to myeloma disease burden or to frailty. These carry a different prognostic implication

Separation by PC2 was then used to assess the contribution of parameters to distribution above the origin or below the origin. Below the origin, the processes are most often associated with performance, comorbidities and inflammation. Above the origin, processes appear to relate mainly to myeloma factors - shown by the relative contributions of the eigenvectors (loadings) in Figure 10.

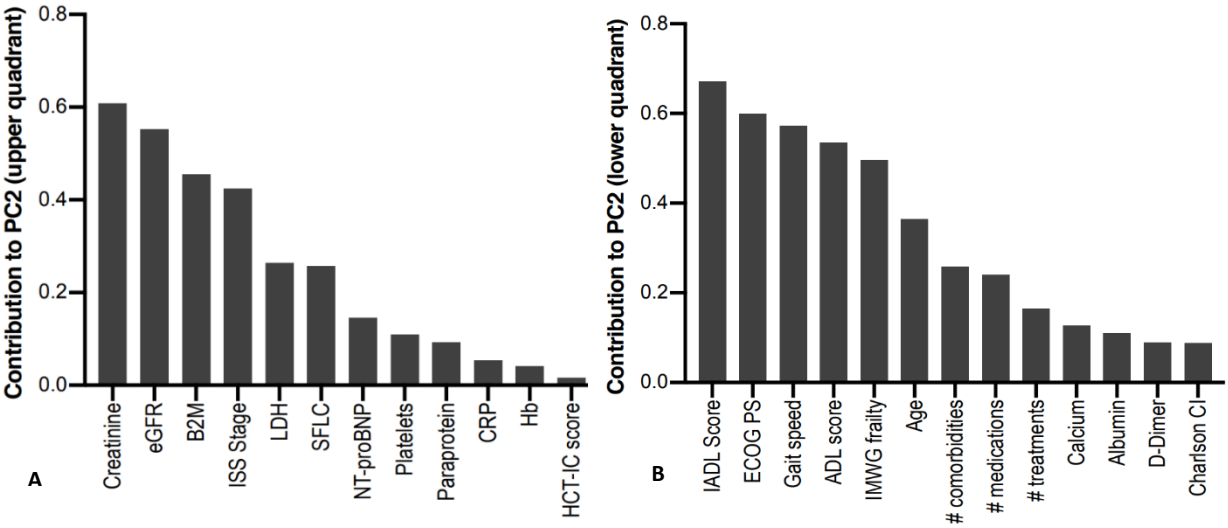


Figure 10: (A)The contribution of each of the variables to the upper quadrant of PC2; (B) The contribution of each of the variables to the lower quadrant of PC2

The contribution of each variable is derived from the Eigenvector value. 10A shows that variables associated with disease activity such as renal function, B2M and ISS classification are the most important in separating patients by PC2. These could together be described as variables that largely reflect a high disease burden. 10B shows that variables associated with frailty burden such as IADL, gait speed and ECOG PS are the most important in separating patients by PC2 in the lower quadrant. These could together be described as variables that largely reflect a high frailty burden.

Given that PC2 appears to separate patients into distinct groups: those with high disease burden and those with significant frailty, we combined these findings with the observation that PC1 separates patients into different prognostic groups. Those closer to the origin on PC1 had significantly higher overall survival rates. We therefore split our cohort into three groups: ‘myeloma clinically well’, ‘myeloma high disease burden’ and ‘myeloma frail’ (Figure 11)

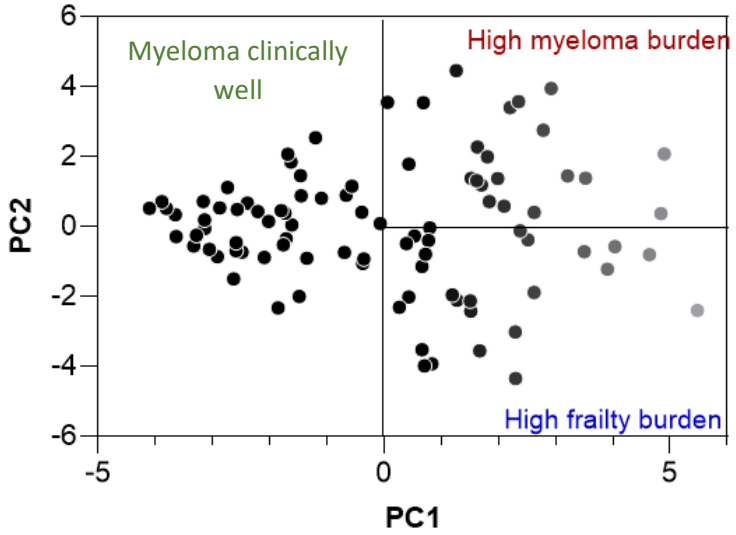


Figure 11: Separation of cohort based on PC1 and PC2 into three clinically distinct groups

This figure shows how the cohort of patients can be separated into distinct clinical groups. PC1 is a measure of overall fitness, while PC2 divides patient into those who are predominantly frail or those who have a high disease burden. Combining these two parameters, patients can be split into a no adverse features group, those with high disease burden or a high frailty burden. This therefore provides a meaningful separation of patients based on standard of care tests, with potentially useful clinical benefits.

These three groups were then compared with each other, aiming to identify any difference in overall survival. It was found that here was a statistically significant difference between the overall survival of the three groups (log rank p= 0.004). (Figure 12). The plots also show different survival patterns of the three groups. After 1-year of follow up (Point A on Figure 12) the patients with high frailty burden and those with high myeloma burden had a similar survival (curve comparison p=0.98) that was significantly inferior to the clinically well myeloma group (p<0.007, hazard ratio >3.4). However, at the end of the full follow up period (Point B on the Figure 12), the survival of those with high myeloma burden converged with the survival of the clinically well myeloma group. Indeed, these groups were

no longer statistically separable (p=0.07). However, the myeloma with frailty group experienced continued excess mortality, with an inferior survival when compared with the other groups (p = 0.0002, mortality 70%).

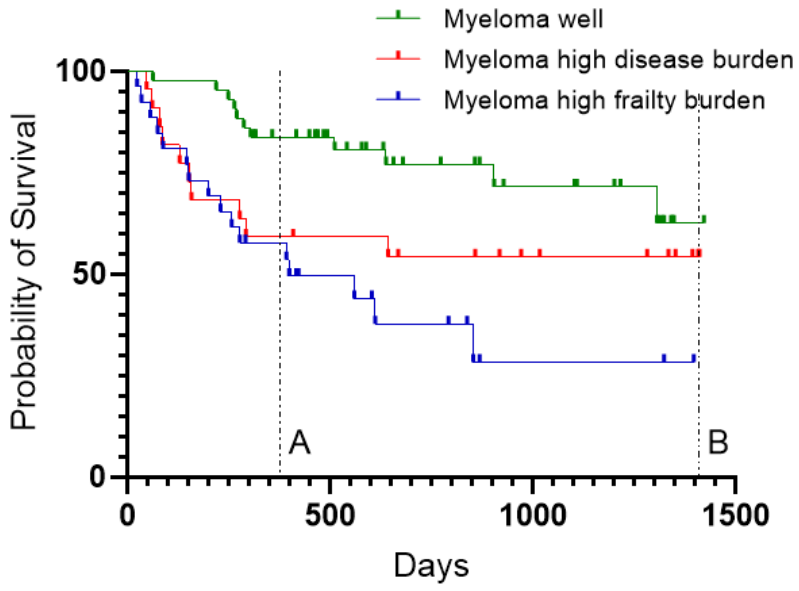


Figure 12: Overall survival of patient cohorts, separated according to PCA separation into groups: 'myeloma clinically well', 'myeloma with high disease burden' and 'myeloma with high frailty burden'.

This figure demonstrates a statistically significant difference (p= 0.004) in overall survival of the three groups, when split by PC1 and PC2 by log rank analysis. However, the groups display different patterns of survival. There is a higher mortality rate in the high disease burden group in the first year, but a significant plateau in survival thereafter, converging with the myeloma well group. The high frailty group experience excess mortality throughout.

5.6: Discussion

Our cohort are representative of a real-world myeloma cohort at a tertiary treatment centre, with baseline demographics comparable to other studies. Prior to drawing conclusions from our data, it is important to confirm the premise that this cohort is both comparable to previous similar papers and representative of a 'real world' multiple myeloma cohort at a tertiary treatment centre. There are two important studies with which our cohort was compared: the Palumbo *et al* study[99] (used to develop the IMWG Classification) and the Engelhardt *et al* study[100]. Engelhardt *et al* (UKF Cohort) produced a table comparing the baseline demographics of their cohort with the Palumbo paper.

Analysis of our patient cohort shows that our population are comparable to the previously analysed myeloma cohorts. All three of the cohorts (including our own) have a low representation of patients with ECOG performance status 3-4. This is an important limitation of all the studies that should be taken into consideration when drawing conclusions.

Palumbo *et al* [99] recruited patients from three prospective clinical trials of newly diagnosed myeloma patients (EMN01-NCT01093136 trial, 26866138MMY2069-NCT01190787 trial and IST-CAR-506-NCT01346787), ineligible for autologous stem cell transplant, in several European countries (Italy, Czech Republic and the Netherlands). This may explain the main difference between our cohort and this historic cohort; In their cohort, just 2% of patients were below the age of 65. This compares to 41% of patients in our cohort. This is likely due to our cohort including transplant eligible and ineligible patients, while transplant eligible are excluded from their study.

Another difference is renal dysfunction. In the Palumbo cohort only 5% of patients had a creatinine of > 2 mg/dl[99], compared to 21% in our cohort. This again likely reflects their recruitment of exclusively trial eligible patients, compared to our unselected cohort. The exclusive recruitment of trial patients in the Palumbo cohort also likely explains the higher CCI in our cohort (2 vs 0), as patients with significant comorbidities were likely excluded.

The Engelhardt cohort[100] is perhaps a more comparable group, in that they included both transplant eligible and ineligible patients in their cohort. They recruited 125 consecutive patients at diagnosis from their hospital in Freiburg, Germany. This does differ to our cohort,

in that we included both newly diagnosed and relapsed patients in our analysis. It is perhaps therefore surprising that we had fever patients with ADL score < 4, as one may assume a pre-treated cohort may have activities of daily living affected by previous lines of therapy. It is also surprising that our cohort contains more IMWG fit patients than the Engelhardt cohort (38% vs 18%) given a proportion of our patients were pre-treated. Unsurprisingly however, our cohort does have a worse median IADL score, perhaps reflecting disease and previous treatment side effects. Other characteristics including renal function, cytogenetics and ISS are similar between these cohorts.

In summary, despite minor differences, most notably our inclusion on transplant eligible and relapsed patients, the cohorts appear comparable in terms of baseline demographics.

Our myeloma cohort is representative when validated with published prognostic outcomes of other studies. When the overall survival of our cohort was assessed, five baseline factors were able to separate the patients into statistically significant cohorts, with different survivals: ECOG performance status (p=0.05), ADL (p=0.04), cytogenetic risk (p=0.03), ISS (p=0.05) and R-ISS (p=0.001). R-ISS was the most significant discriminator of overall survival.

These results are largely consistent with those of historic cohorts and appear logical given the nature of the disease and frailty. The Engelhardt cohort[100] also showed significant differences when comparing survival curves of patients with different cytogenetic risk and ADL score. They did however also show a difference in survival when separating patients by IADL and CCI. This finding is corroborated in the Palumbo cohort, where a multivariant Cox regression model showed worse outcomes for patients with IADL ≤ 5 and a CCI ≥ 2. While our cohort did show a trend to inferior overall survival with lower IADL and higher CCI, we were unable to reproduce the finding of shorter overall survival using a log rank test to compare survival curves. The reason for this may be that our cohort was too small to identify this difference. We also had fewer patients in the CCI ≥2 cohort (46% Vs 65%), possibly making it more difficult to identify a statistically significant difference.

Greipp at al[101] showed a statistically significant difference in overall survival when comparing ISS groups I vs 2 vs 3, with overall survival reported as 59 months (ISS 1) Vs 49

months (ISS 2) vs 26 months. Their cohort included over 10,000 patients, significantly more than our cohort. Despite this we were able to corroborate the findings of the difference in overall survival between the three groups (p= 0.05). Our significance is lower than the Greipp cohort due to a much smaller sample size. However, it is nonetheless important that our cohort does corroborate the finding in this paper.

The most significant discriminator of survival in our cohort was R-ISS (p=0.001). This prognostic scoring system, which built on the ISS score, includes B2M level, albumin, LDH and presence of high-risk cytogenetic aberrations. Palumbo *et al*[102] used a cohort of over 5000 patients and showed that 5 year overall was 82% in the R-ISS 1 group, 62% in the R-ISS 2 group and 40% in the R-ISS 3 group, with a significant difference between each. Again, it is important that our cohort corroborates the findings of previous studies. Similarly, high risk chromosomal abnormalities (del 17p +/- t(4;14) +/- t(14;16) have been shown to significantly reduce overall survival.[103] Again, we were able to corroborate this in our cohort, where cytogenetic risk alone produced two different groups with a significant difference in overall survival (p=0.03).

It is however also important to recognise that certain factors do not predict overall survival. Age for example did not produce significantly different survival curves. This is a finding also noted in the Engelhardt cohort.[100] Interestingly, in the Palumbo cohort, when the effect of age on overall survival was assessed using a multivariant Cox regression model, patients aged between 75-80 had worse overall survival (HR 1.35) than younger patients, with those over 80 have an even more significant reduction in overall survival (HR 2.68).[99] However, when assessed with simple linear regression, only a trend to worse outcomes with increased age was identified. This emphasises the issue that chronological age is itself not a sufficiently good discriminator of patient fitness and more sophisticated tools are required to classify patients into prognostic groups. It does however seem reasonable to include it in a composite assessment score.

Palombo *et al* were also able to show statistically different 3-year overall survival rates between the IMWG 'fit', 'fit-intermediate' and 'frail' groups.[99] This measurement is of clinical importance, as it is being used in the current UK wide Myeloma XIV (FiTNEss) Study.[104] The three-year overall survival in their fit cohort was 84%, 76% in the intermediate fit group and 57% in the frail group. The hazard ratios were all significant. The data

collection period of our cohort was shorter that the Palombo paper[89], thus it was not possible to calculate a 3-year progression free survival for these clinical groups. We were not able to find a difference in overall survival by comparison between the survival curves of each IMWG classification. This may reflect the limited follow-up in our cohort.

Finally, we were able to show a statistically significant difference in overall survival by log rank testing between patients with an ECOG score of 0-1 vs 2 vs 3-4 (p=0.05). This separation of ECOG scores was selected based on sample number rather than an underlying physiological difference. ECOG performance status was in fact removed from the IMWG classification as its inclusion resulted in less significant discrimination between the three groups. Our finding that it has a significant effect on survival is an interesting one and not without precedent. Afram *et al* showed the ECOG performance status was a significant factor in the overall survival of their replaced/ refractory cohort treated with the monoclonal antibody daratumumab (p= 0.001).[105] This study separated patients into ECOG 0-1 vs ECOG ≥ 2, rather than the three groups we used. It is therefore possible that the ECOG performance status is less significant in newly diagnosed patients (as recruited in the Palumbo cohort) compared to in the relapsed and refractory setting.

It is also worth noting that the ECOG performance status (Appendix 2) is a rapid, clinician assessed tool to assess performance status but appears to be a useful predictor of patient outcome. However, a study of patient-oncologist agreement in performance status showed that there was only 50% correlation between reported patient and clinician reported scores[106]. Despite this, both scores separated patients into distinct groups with different overall survivals.

In summary we have been able to reproduce findings that show overall survival in myeloma is significantly affected by ECOG performance status, ADL, cytogenetic risk, ISS and R-ISS. We were unable to show a difference in overall survival when classified by IMWG classification, the current UK standard score for patient classification. These findings support the hypothesis that there are multiple important predictive variables for overall survival in this patient group. These variables include disease specific factors, comorbidities, and frailty factors. It does however appear that different classification tools identify different patients with a poor prognosis (Figure 13). ISS, ECOG and CCI identify patients in our cohort with a high mortality, but seem to be identifying different patients. In fact, of the patients in the

poorest risk group for each classification, only one patient was identified by all three. This suggests that a unified score or aggregate of multiple variables may improve prognostication in this patient group.

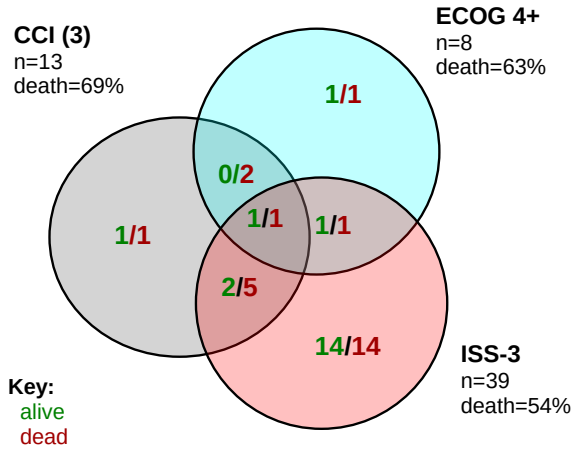


Figure 13: Venn diagram of patients identified by CCI, ECOG and ISS in the poorest prognostic groups

This Venn diagram shows all the patients identified as being in the poorest prognostic or most frail/ co-morbid group in three validated classification systems. While they are good at predicting mortality, they identify different patients with poor prognoses.

Unsupervised analysis (PCA) can be used to identify clinically significant groups within the cohort. Our initial findings demonstrated that our cohort was representative and comparable to published cohorts, and that a range of established biomarkers and scores could predict statistically different overall survival. We therefore proceeded to investigate this cohort further using exploratory statistical analysis.

As previously mentioned, there a number of scoring systems can be used to assess a patient with multiple myeloma. It is critically important however for the clinician to understand what each of the scores is designed to show, and therefore how to interpret them. Figure 14 is an attempt to classify these scores.

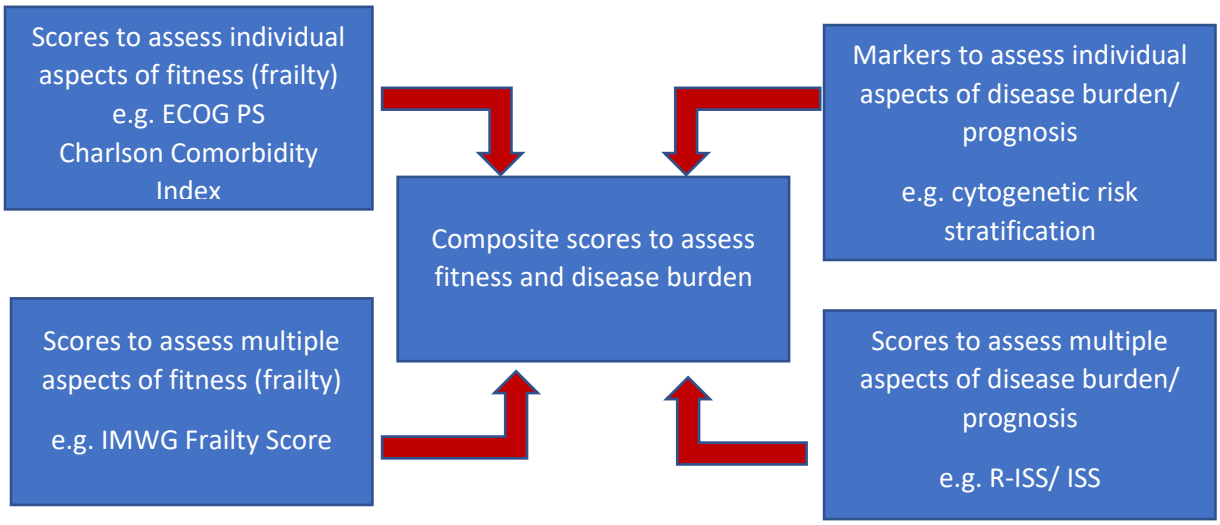


Figure 14: Classification of the various scoring systems used in multiple myeloma

These scores can use individual or composite measures of fitness or frailty. They may have some predictive properties (e.g. overall survival or toxicity of medications) but can also be used to classify patients into different treatment pathways (i.e. IMWG classification). Equally the individual markers or composite scores of disease burden can be used as predictive tools for overall-survival but may also be of some benefit in deciding treatment. The ideal goal would however be a simple, cost-efficient composite score, including both markers of disease burden and fitness that had predictive properties and could be used to assign the best possible therapy to the individual.

The fundamental issue is that composite scoring systems with improved predictive value often require more clinical data, making them difficult to use out with clinical trials. Clinicians are often significantly time pressured and are not able to perform complicated scores for each patient. However, baseline patient characteristics, blood investigations and standard of care tests provide extensive material with which to assess a patient's disease, fitness and ultimately outcome.

PCA is a statistical method used to assess data sets containing large numbers of variables. Essentially, PCA uses the covariance of different components in the dataset to group individuals who have similar characteristics. It is also an example of unsupervised analysis. That is to say that clusters of patients will be identified using the PCA algorithm purely based on common co-variances, with no preselected groupings. This is as opposed to

'supervised analysis' where a pre-exisitng cohort of patients are identified as being similar (e.g. fit Vs frail) and the question is 'what characteristics do these patients have in common?' The main advantage of PCA is its ability to assess the relative importance of a large number of variables and to group patients into clusters based on all of their baseline characteristics, rather than a select few statistically significant variables, identified as important by multivarient regression analysis.

The first finding of our PCA analysis was the loadings of each variable or contribution of the Eigenvectors. This essentially demonstrates which of the baseline variables for the entire cohort are most important in separating the patients into clusters of covarience. The eight facotrs most significant in this separation were: renal function, B2M, ISS score, LDH, ADL score, IADL score and gait speed. This is an intreaging finding as the first four variables are associated with disease burden and the later four with fitness or frailty. This initial finding would suggest that both disease burden and frailty are inportant contributors to patients with newly diagnosed or relapsed myeloma.

PC1 separates patients by 'good or poor health'. PCA provides several different separations of the data set, and these are numerically named PC1, PC2, PC3 etc.. PC1 will have the most significant separation of patients, with PC3 the least. Each PC can therefore be assessed alone, or in combination with other components. Each PC is ascribed a percentage variance. This value shows how much of the variability of the entire data set can be explained by that PC. In this case, PC1 explained 22.1% of the variability in the data set, with PC2 accounting for 14.5% of the total variability. Together they account for 36.6% of the total variability. The fact that this total figure is less than fifty percent may appear low but is unsurprising given the number of biological variables involved in a complex condition such as multiple myeloma.

Separation of PC1 (the major component of variability) separated patients into two groups- the PC1-LOW group (n=46) and the PC1-HIGH group (n=46) and were assessed using Kaplan Myer survival analysis and log rank comparison of survival curves. The PC1-HIGH group had inferior overall survival, with a hazard ratio of death of 3.6 (p= 0.0001) when compared to the PC1-LOW group. This separation of survival curves is the most statistically significant of

all survival curves in our data set, including IMWG frailty score, cytogenetic classification and ADL score. It is also possible to assess which of the baseline variables contribute most to separation by PC1 alone. The variables we found to most affect PC1 (thus survival) were CCI score, NT pro-BNP and B2M, renal function. The presence of the CCI in this list is expected, it is present in the IMWG frailty score (which predicts overall survival) and has been found to be significant my multivariant analysis.[99] B2M is the light chain of the HLA histocompatibility complex and has historically found to be significantly increased in myeloma.[107] It has subsequently been included in the ISS staging system for prognostication.[40]. The fact that PC1 found it to be an important contributor to overall survival is therefore consistent with expected findings. Indeed, the presence of renal impairment has also been shown to be an independent statistically significant risk factor for reduced overall survival in myeloma patient in a historical cohort.[108] This led to Englehardt *et al* suggesting it should be included in a revised IMWG frailty classification.

The presence of N-terminal fragments of pro-hormone brain natriuretic peptide (NTpro-BNP) in this list is perhaps more surprising. It has not been included in traditional predictive scoring systems for myeloma but is an established predictive factor in AL amyloidosis.[109] NT pro-BNP is released predominantly by the left ventricle, and largely in response to high ventricular filling pressures[110] and is associated with cardiac dysfunction. It is a largely renally excreted molecule. It may therefore be raised in myeloma due to the presence of cardiac amyloid, due to reduced renal excretion due to myeloma associated renal dysfunction or may simply represent antecedent cardiac dysfunction, present prior to the development of myeloma.

Abe *et al* have reported that patients with a NT pro-BNP of greater than 341µg/ml have shorter overall survival than those with values less than 341µg/ml, making it a potentially useful predictive parameter.[111] This finding was found to be significant, even when adjusted for renal dysfunction. Indeed NT-proBNP has recently been used in a novel myeloma frailty score produced by Milani *et al* which combined age (> 70) , ECOG performance status (≥2) and NT-proBNP (≥300µg/ml) to produce clinical subgroups with highly significant overall survivals.[112] We therefore conclude that NT-proBNP has an important role in determining overall survival in our cohort, and this is supported by recent studies.

It is also important to recognise those variables that did not significantly impact PC1 and thus overall survival. Age > 70 was only the 18th most important contributor to PC1. This is despite age previously being found to be an independent risk factor for overall survival in several cohorts[39, 99, 109]. It is however, worth noting the age cut-offs used in these studies vary, with some using three groups (< 75, 76-80 and > 80) while others use a binary > or < 70. Our data was analysed using the same cut offs as the IMWG cohort[99] and the UKF cohorts.[100] The UKF cohort also failed to show a statistically significant difference in overall survival between the age groups. It may therefore be that either our data set was too small to notice a difference or that differences in cohort population (e.g. comorbidities) were more important factors.

PC2 separates patients with high disease burden or high frailty burden. PC2 is differently affected by the baseline variables. It is the PC with the second most statistically significant separation of patients. When the contribution each of the variables to PC2 is assessed, it is different to PC1. Factors contributing to separation in the upper quadrant of PC2 were: renal function, B2M, ISS and LDH. These variables can largely be considered factors of 'significant myeloma disease burden'. In juxtaposition to this, factors contributing to separation in the lower quadrant of PC2 were: IADL, ECOG PS, gait speed and ADL. These factors are more associated with fitness/ frailty. We therefore suggest that PCA can separate patients into those who's phenotype is dominated by high disease burden, and those who's disease is mostly fitness/ frailty associated.

PC1 and PC2 combined produce 3 sub-groups with clinically distinct phenotypes: myeloma clinically well, myeloma with high disease burden and myeloma with high frailty burden. These groups have significantly different overall survivals. It is then possible to combine PC1 and PC2 together to produce a plot, containing each patient as a different data point, plotted against PC1 on the x axis, and PC2 on the y axis. This combines the observations that PC1 appears to be assessing disease prognosis in terms of survival, and PC2 separating patients by high disease burden or fitness/ frailty. This amalgamation of PC1 and PC2

produces 3 distinct clinical groups: ‘myeloma clinically well’, ‘myeloma with high disease burden’ and ‘myeloma with high frailty burden’.

When log rank survival analysis was undertaken, there was a statistically different survival when the three groups were compared to each other (p= 0.003). The clinically fit group had the longest overall survival, with the myeloma with high frailty burden having the shortest. The myeloma high disease burden group had an intermediate prognosis, with many early deaths. When one year survival analysis was undertaken, patients with high frailty burden and high disease burden had similar survival curves when compared directly (curve comparison p= 0.98, HR 0.99). Both groups had significantly inferior survival to the clinically well myeloma group (p= 0.007, HR >3.4). This is in contrast to the full follow-up period, where the survival of the high disease burden group began to converge with the myeloma clinically well group (p= 0.053, HR 2.2). However, the myeloma with high frailty burden continued to have excess mortality when compared to other groups (p=.0004, HR 4.5)

This observation is an important clinical one for a variety of reasons. Firstly, regarding the myeloma clinically well group, it would be beneficial for the clinician and patient to know that the prognosis and overall survival is good. However, there are multiple antecedent prognostic scoring systems that will also provide this information.

The potentially more useful clinical subdivision is between the myeloma with high disease burden and myeloma with high frailty burden groups. It is often difficult at diagnosis to differentiate clinically between an unwell patient in whom the primary aetiology is high disease burden and one who’s frailty and comorbidities are the dominant cause. The combination of PC1 and PC2 appears to allow separation of these groups.

Also important is the observation of the convergence of the myeloma clinically well and high disease burden survival curves. This appears to suggest that if patients with high disease burden survive the initial post diagnosis period, they have an improved prognosis. It would therefore be of significant clinical benefit to identify these patients from the frail cohort, who continue to experience a high death rate after the initial treatment period. On a practical level, it may be that the high disease burden patients may require a re-review of frailty/ performance status, consideration to intensification of therapy and potentially reconsideration for an autologous stem cell transplant.

Finally, the identification of a myeloma with high frailty burden cohort potentially allows clinicians to identify a cohort with a poor overall survival. This may allow improved decision making in terms of selecting regimens with less toxic side effect profiles and a focus on quality of life.

5.7: Conclusions

In summary it can be concluded that our cohort of patients are comparable to previously studied myeloma cohorts, with some of the same limitations. We have been able to corroborate historical findings demonstrating EOCG PS, ISS, R-ISS and cytogenetic risk group are independent risk factors for overall survival. We have also demonstrated our novel PCA groupings identified by unsupervised analysis are able to predict survival with more statistical significance than any of the pre-existing variables. Analysis of these novel groupings also demonstrate different biological processes and drivers for separation including the presence of active disease and frailty. These observations are of significant interest in terms of improved prognostication and insight into the biological processes in multiple myeloma.

6: Chapter 2- Analysis of serum protein biomarkers of frailty in Multiple Myeloma

6.1: Introduction

The results in the previous chapter demonstrated that our cohort of myeloma patients had comparable demographics and baseline standard of care results compared with published cohorts. We also showed expected patterns of overall survival when separated according the predictive markers and scores: ISS, R-ISS and cytogenetics. Our novel PCA groupings produced a significant overall survival curve separation (p =0.003) between the 'myeloma with no adverse features' group, the 'myeloma with high disease burden' group and the 'myeloma with high frailty burden' group. This is a comparable p value to that produced by the R-ISS predicative model (p= 0.001), but has the additional property of distinguishing patients with either frailty or high myeloma disease burden who may require different treatment.

However, our PCA groupings require the recording and assimilation of 25 different baseline tests or investigations. Although all are all standard of care tests incorporating them into a PCA would not be suitable for clinicians in a busy myeloma clinic. Indeed, a recent review of clinical scores has suggested a score or prognostic tool should 'include routinely recordable variables… easily applied to an algorithm', along with providing 'increasing efficiency and improving outcomes'.[113] It could be argued that although our novel groupings have the potential to improve outcomes, they do not meet the other criteria. Furthermore, due to the nature of the PCA, it is not possible to prospectively classify patients into these groups.

This chapter therefore explores whether protein biomarkers identified in serum can identify the same novel PCA groups, and so potentially meet the objectives of a useful prognostic tool. The use of biomarkers in multiple myeloma has precedent, with recent literature being published on serum and bone marrow samples in the condition.[114] There is also growing body of published evidence of proteomic biomarkers and their identification in aging and frailty.[115, 116, 117, 118] Proteins are commonly used as biomarkers in medicine as they are easily quantified and tested and are an objective measurement free from bias.[119] They are also significantly easier to measure than the complicated disease processes that cause the homeostatic imbalance, leading to changes in their abundance.[120]

Representative serum samples from patients belonging to the different PCA subgroups were therefore selected and subjected to SWATH proteomic analysis to look for potential biomarkers. This chapter explores whether specific serum proteins can be identified that have the potential as biomarkers to identify each of the three prognostic groups making them more easily identifiable for trial or future clinical use. Finally, the identification of proteins in the serum of myeloma patients might also offer further insight into the complicated interplay between biology of frailty and multiple myeloma.

6.2: Aims and objectives

Aim

To use quantitative serum proteomics using supervised analyses to determine whether we could prospectively identify the clinical subgroups identified in the previous chapter and to allow prognostic classification, and to see if the proteins identified could improve understanding of myeloma or frailty processes.

Objectives

1. To identify whether differences in the quantitative abundance of serum proteins could be used to identify the clinical subgroups prospectively
2. To assess whether these differences could indicate biological significance with reference to frailty or myeloma.
3. To identify a set of potential protein and clinical biomarkers that together could form a potential clinical and biomarker set for prognostic assessment in myeloma

A summary flow diagram of the process of analysis in this chapter can be found in Appendix 4.

6.3: Methods

Sample collection

1) 4ml Ethylenediaminetetraacetic acid (EDTA) sample was taken from each patient at entry to trial.
2) The sample was immediately centrifuged using a microfuge, then the supernatant (serum) was extracted.
3) The serum sample was labelled with unique patient identifier and stored at -80^{o}C until analysis.

Sample selection and transport

1) PCA has previously identified three discreet patient groups: 'myeloma with no adverse features', 'myeloma with high disease burden' and 'myeloma with high frailty burden'. 17 representative samples from each group were identified (n=51).

   Figure 15 (below) shows the PCA analysis of these 51 samples, demonstrating the samples to be representative of the three previously identified groups. It should be noted that the majority of the serum samples were closely clustered in the respective quadrants, though some others were more loosely clustered, particularly between the high disease burden and high frailty burden group.

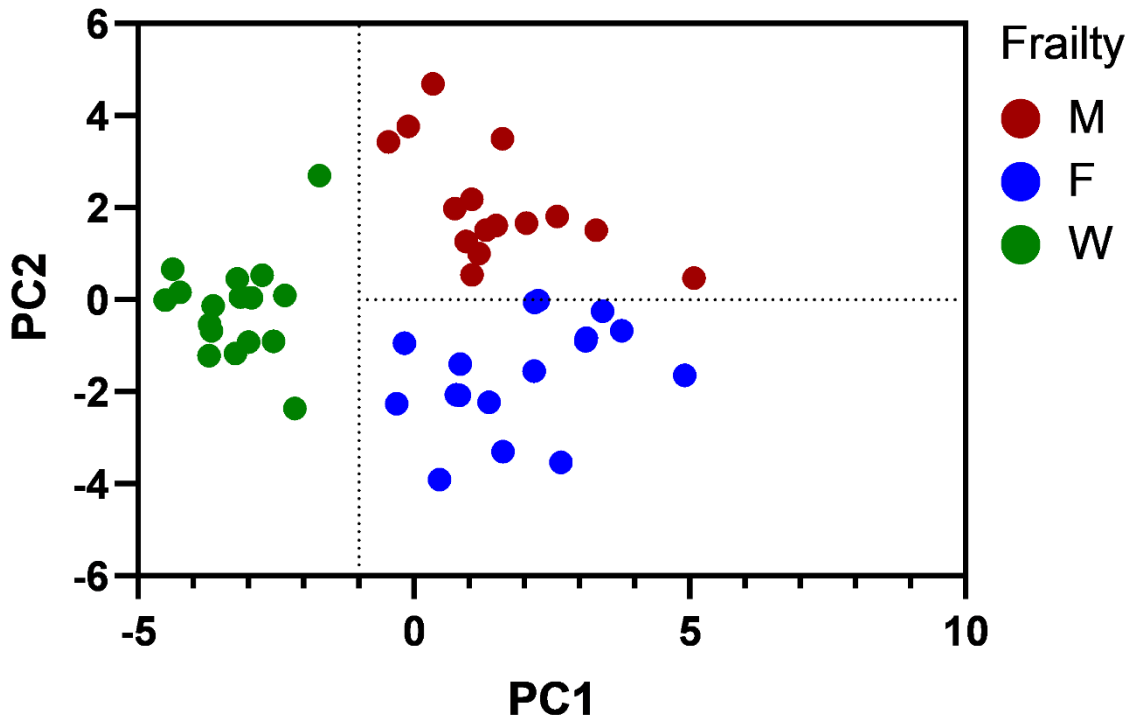


Figure 15: PCA analysis of the 25 baseline characteristics/ tests of the 51 patient samples selected for proteomic analysis

This figure shows the PCA analysis of the 25 baseline characteristics/ tests of the 51 patient samples sent for proteomic analysis. (W): well with no adverse features, (M): myeloma with high disease burden, (F) myeloma with high frailty burden. It demonstrates that the samples selected are representative of the original cohort

2) Once the samples were identified they were transported to the Stoller Biomarker Discovery Centre (SBDC) in Manchester for proteomic analysis.

Sample processing

Note: this stage of the research was performed by staff at the SBDC and I was not involved in the physical processing of the samples. The SBDC then provided the results which were subsequently analysed for the MD Project. In brief, the SWATH (Sequential Window Acquisition of All Theoretical Mass Spectra) analysis was performed as follows:

1. Prior to processing the mass spectrometer was calibrated, and internal quality control performed using a consistent control sample supplied by the instrument manufacturer.
2. Samples for SWATH-MS were prepared in the following steps:
    a. Immunodepletion was used to remove high abundance proteins that would reduce identification and quantification of lower abundance proteins in human serum, these were albumin, IgG, IgA, IgM, IgD, IgE, kappa and lambda light chains, alpha-1-acidglycoprotein, alpha-1-antitrypsin, alpha-2-macroglobulin, apolipoprotein A1, fibrinogen, haptoglobin, and transferrin
    b. Initial protein quantification was performed to check quality and allow normalisation of results
    c. All samples were subjected to denaturation, reduction, alkylation, digestion, and acidification to generate stable peptides suitable for SWATH
    d. Samples were lyophilised and stored, the re-suspended at the time of analysis

3. Samples were then processed in four separate batches, alongside a batch control and a control normal serum to monitor the instrument performance over time.
    a. Control serum contained commercially available pooled serum
    b. Batch control contains a pool of all the samples in the study.
4. Gel electrophoresis was performed on the batch control and the control plasma to assess the effectiveness of depletion/digestion prior samples being run on the mass spectrometer.

Sample Quality Control

17 frozen serum samples were sent to the Stoller Biomarker Discovery Centre (SBDC) from each of the PCA selected groups ‘myeloma with no adverse features’, ‘myeloma with high disease burden’ and ‘myeloma with high frailty burden’. Table 7 shows a breakdown of the samples sent to the SBDC. Two samples were not processed, the first due to lack of clear labelling and the second due to sample haemolysis. All remaining samples passed the quality control QC stage.

Table 7: Quality control analysis of samples submitted to SBDC

	Total	Myeloma with no adverse features (Group 1)	Myeloma with high disease burden (Group 2)	Myeloma with high frailty burden (Group 3)
Number of samples provided	51	18	17	17
Number of samples failing QC	2	1	1	0
Number of samples with analysis completed	49	16	16	17

This table shows that 49 of the samples (of 51) passed the quality control step, allowing 16 or 17 samples to be analysed in each of the novel groups.

SWATH Analysis

SWATH-MS identifies ions formed from peptides of the proteins in the prepared sample, without the need for prior knowledge of the protein constituents of the sample. The technique employs tandem mass spectrometry techniques by which peptides are selected in an in initial MS chamber (MS1) then individually subjected to collision ionisation in the second chamber (MS2) to allow ions formed from the peptides to be sequenced then compared to a virtual library. The peptide from which an ion arises is then identified from the library and quantified based on ion abundance.[121] The result is a quantitative and qualitative result for the protein constituents in a sample. Each protein is potentially represented by multiple peptides so "false discovery" is reduced by accepting only protein identification based on sufficient unique peptide identifications.

Subsequent initial analysis is as follows

1) SWATH maps are analysed using OpenSWATH software, with MSStats software then used to provide quantitative and qualitative data of the proteins identified in the patient trial samples, batch control serum and control serum.
2) PCA analysis is performed to compare control healthy serum to our cohort of myeloma patients' serum.

3) PCA analysis (unsupervised analysis- discussed in Chapter 1) performed on patient serum samples to determine whether groups based on co-variance of protein abundance could be identified from overall protein expression
4) Random Forest analysis is used as a supervised analysis technique (using R Statistical Computing Software to identify candidate biomarkers) that separate our pre-defined groups.
5) For a protein to be considered as a candidate biomarker, it had to be measurable in at least 30% of the patient trial samples and have passed the quality control steps.
6) At the end of this process, the SBDC provided a list of proteins found in the myeloma patient sera that were identified as potential candidate biomarkers.

Post SWATH-MS Analysis

Subsequent analysis was performed as part of the MD and was not provided by the SBDC.

1) The data from the SWATH-MS was provided to me in an EXCEL spreadsheet. The data included those proteins identified, a quantitative measurement of the protein abundance including FDR and peptide number. Additional characteristics included the number of patients in whom a protein was identified. The data highlighted the Random Forest and waterfall data showing which proteins were most differently identified in each of our three clinically-defined groups: myeloma with no adverse features, myeloma with high disease burden and myeloma with high frailty burden.

2) The data of protein abundance was then re-analysed to exclude those proteins that were not present in at least 80% of samples were removed. A 80% threshold was chosen ensure robustness of statistics and to reflect that these proteins must be useful in practical diagnostic testing (will be commented on further in the discussion).

3) Further statistical analysis employed Graphpad Prism Software, using only those proteins selected for analysis in step 2. This used a structured analysis to identify the most important proteins that separated the groups:

a. ANOVA to identify only those proteins with significant difference within the selected protein group.
b. T-tests were used to refine the analysis of those protein showing a significant difference by ANOVA. This allowed a paired assessment to see which diagnoses were distinguished by the makers (myeloma with no adverse features against myeloma with high disease burden; myeloma with high disease burden against myeloma with high frailty burden; myeloma with no adverse features against myeloma with high frailty burden)
c. Receiver operating characteristic (ROC) for confirmation of the pattern of difference between the three groups and add to the knowledge of potential diagnostic utility (in a pairwise fashion).

4) The proteins that fulfilled the following criteria were then taken forward to assessment of their ability to differentiate between the three groupings. A similar approach was used by Grissa *et al* in their search for early predictive biomarkers.[122]
    a. RF: Identified in the RF analysis (of novel classification groupings) and identified in at least 80% of patient samples
    b. ANOVA: $p < 0.05$ for difference level of protein abundance between three groups
    c. ROC: 0.7 providing a cut off for sensitivity and specificity
    d. T-test: $p < 0.05$ for difference level of protein abundance between each paired group
5) Assessment of biological criteria employed a literature review. The literature review was performed using an online database search (Pubmed) and the search words were: name of protein, ageing, frailty, myeloma and cancer. Articles not relevant were excluded. Proteins were also reviewed on 'Uniprot'[123], 'Gene'[124] and 'String'[125] for their biological function and tissue expression profile. A table of significance for each protein in relation to myeloma, disease and frailty was constructed from that data.

6.4.1: Results- Section One

Data presented to me by Stoller Biomarker Discovery Centre for interpretation

This data is included in the results as it is essential to the understanding of the subsequent data analysis and conclusions. The data is reviewed and interpreted in the section below.

6.4.1.1: Protein abundance was found to be different between healthy controls and myeloma patients, as well as between the novel PCA groups

The samples were first compared to a commercially available 'control serum'. That control serum (CS) (BIOVT Product No. HUMANSRMPNN) was prepared from healthy patients with no active multiple myeloma. This was compared to the total pooled serum (TOT) from each of the 49 patient samples in the study. TOT samples were tested as part of each SWATH-MS batch (n=4). The results were presented in the form of a PCA analysis of all proteins quantified (Figure 16). The PCA shows that the TOT serum (serum from patients with multiple myeloma) clustered together, showing a similar co-variance, while the CS (healthy serum) samples also clustered but in a separate area consistent with them recognising a distinct disorder. This shows that no instrument related batch effects influenced the data.

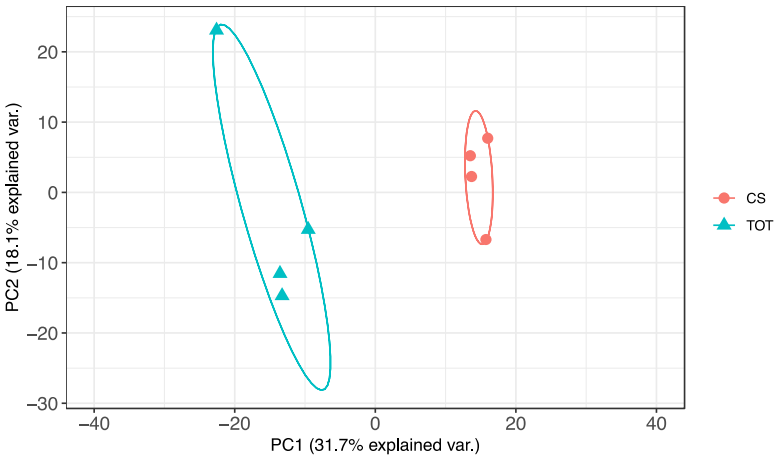


Figure 16: PCA of all proteins quantified in commercially available control serum and pooled serum from the myeloma trial cohort

This figure shows that the proteins identified by mass spectrometry in the control serum (CS) were closely clustered and separate to the serum from the myeloma patients. This shows

that there is a genuine difference in protein content of the samples, rather the difference being caused by technical or instrumental issues

PCA was then used to test CS, TOT serum and all separate study samples (n=49), with the results shown in Figure 17. This analysis shows that TOT control samples cluster much more tightly than study samples, indicating that variation across study samples represents biological rather than technical variation.

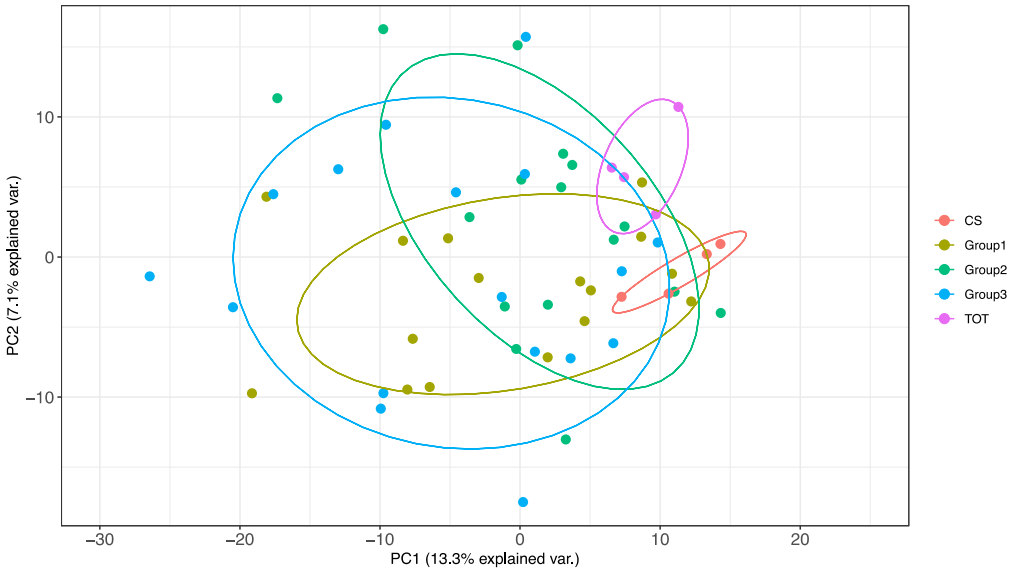


Figure 17: PCA of control samples (CS-red), Group 1 samples (Myeloma with no adverse features- yellow), Group 2 samples (myeloma with high disease burden- green), Group 3 (myeloma with high frailty burden- blue), Total pooled serum (TOT- pink).

This plot shows that the control samples cluster tightly, while the samples from the myeloma cohort are much more loosely clustered. This demonstrates that variation across study samples represents biological rather than technical variation.

PCA was then performed again excluding the CS and TOT samples. For this assessment only proteins present in at least 30% of samples were included in the analysis (Figure 18). The 30% cut off here was arbitrarily chosen by the SBDC and will be discussed in more detail in the results section. Although this analysis did not attempt to refine the protein set and so included many proteins not reflecting relevant biological or pathological processes, there is a suggestion that the “myeloma with no adverse features” and “myeloma with high disease

burden" groups show some loose clustering (as demonstrated by the coloured circles), with less clear clustering amongst those with a high frailty burden. There are however significant outliers identified and patients that appear to have similar grouping of their co-variance to patients in another group. This finding is unsurprising as this analysis reflects variation in all proteins, thus unlikely to give tight clusters as there are multiple biological processes involved in influencing the serum proteosome of an individual.

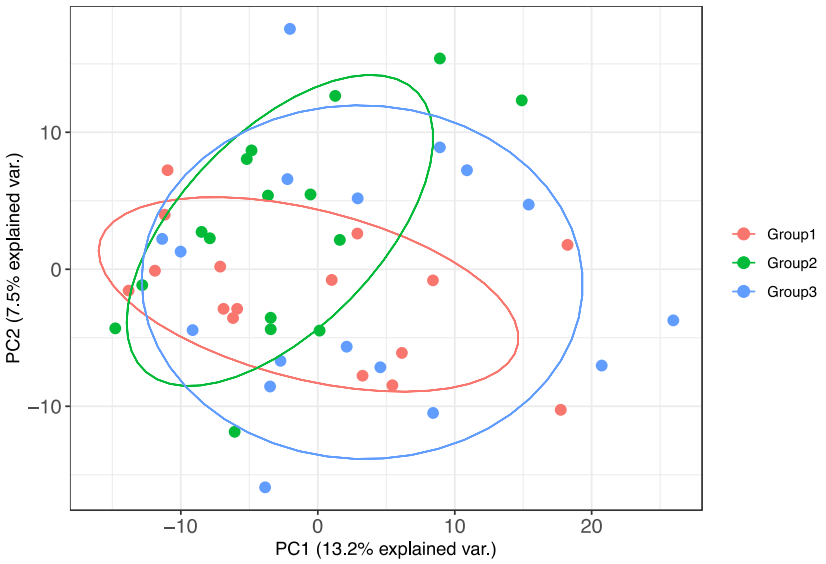


Figure 18: PCA of serum of patients in each of the three groups: Group 1 samples (Myeloma with no adverse features- red), Group 2 samples (myeloma with high disease burden- green), Group 3 (myeloma with high frailty burden- blue)

This PCA plot shows there to be some loose clustering of the myeloma with no adverse features group and the myeloma with high disease burden group, but minimal clustering of the high frailty burden group. This suggests that PCA of all proteins present in more than 30% of samples provides only limited separation of the pre-defined cohorts.

The three originally defined PCA groups were then compared to each other with the objective of identifying protein biomarkers with statistically different quantitative levels within the serum. A Bayes approach was used to identify these proteins. The criteria used to identify a significant candidate biomarker was for there to be >1 logFC difference between the mean abundance in each group and a p value of > 0.05. These criteria were chosen to

provide a pragmatic balance between a false discovery rate and a false rejection rate. Figure 19 is a visualisation of the above parameters.

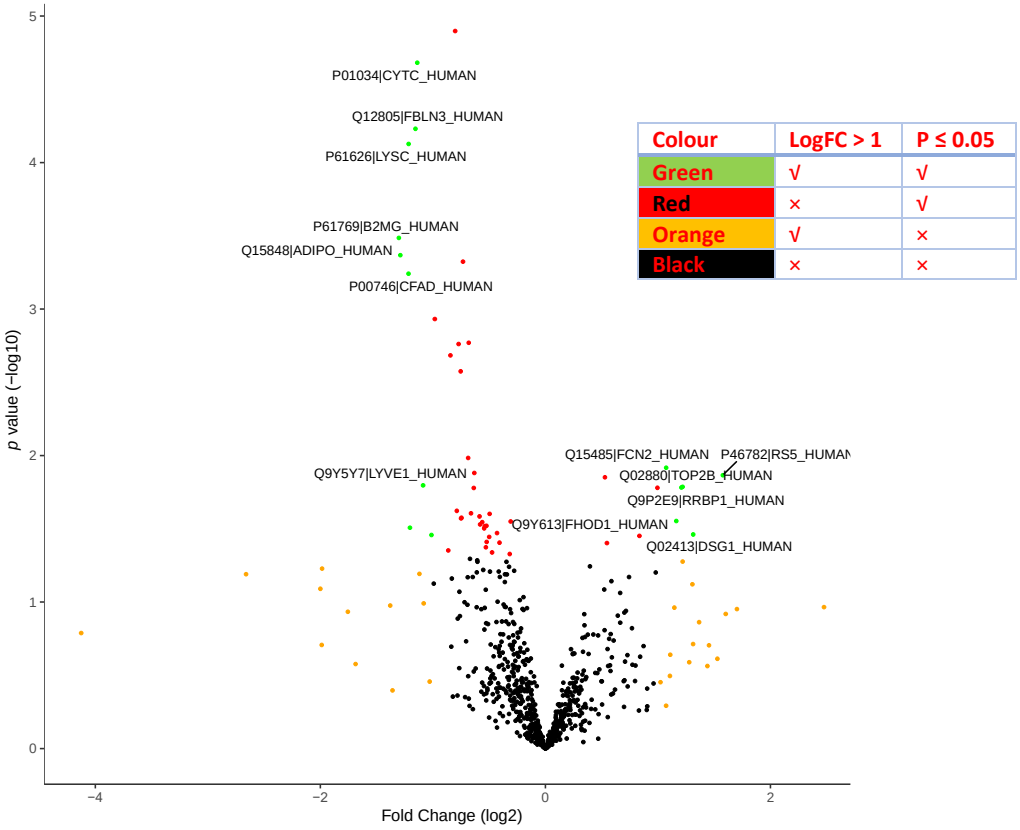


Figure 19: Volcano plot to show the log fold change in protein abundance and significance of this change between the myeloma with no adverse features group (Group 1) and myeloma with high disease burden group (Group 2).

This plot demonstrates that a select group of protein biomarkers discovered in the serum samples of the myeloma patients are able to distinguish between the no adverse features and high disease burden cohorts. These can be considered biomarkers of interest to separate the novel cohorts.

Overall, 49 proteins showed significant differential quantitative values between the two groups. The ten proteins with the most significant are listed the Appendix 5.

The same approach was used to test serum samples from Group 2 (high disease burden) and Group 3 (high frailty burden). Figure 20 shows the Volcano plot difference between the

two groups. 56 proteins were identified as candidate biomarkers. The ten proteins with the most significant differential abundance between the groups are listed in Appendix 5.

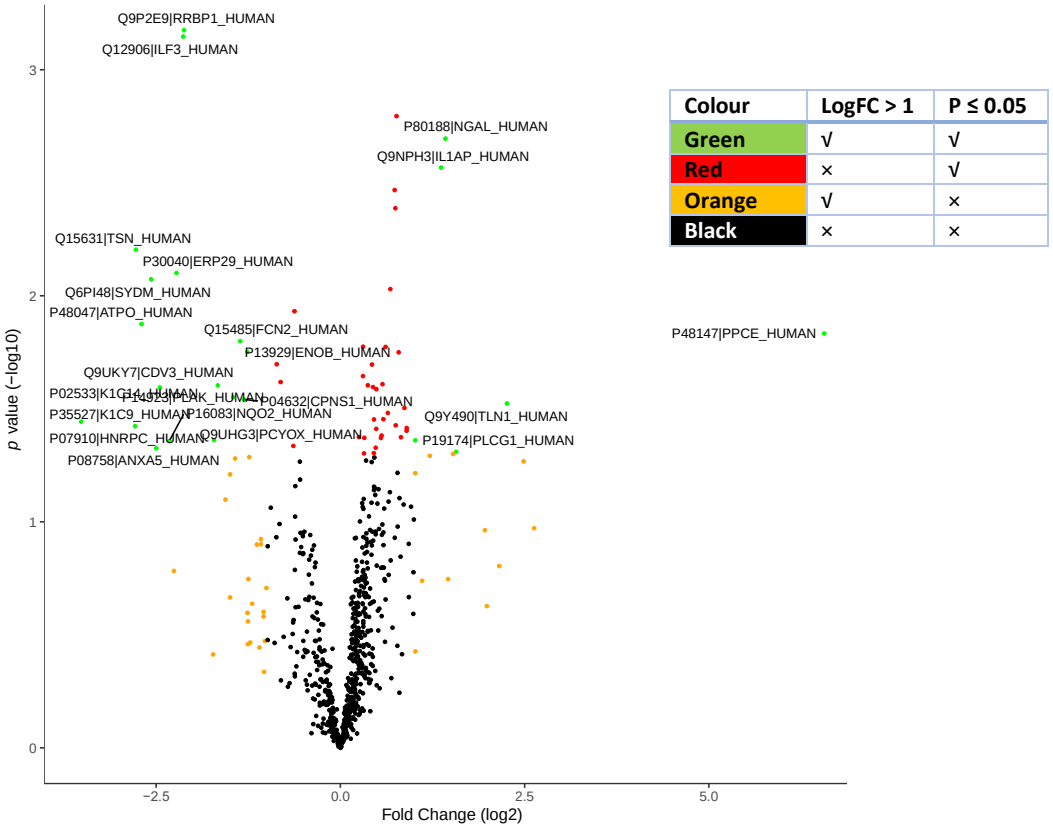


Figure 20: Volcano plot to show the log fold change in protein expression and significance of this change between the myeloma with high disease burden group (Group 2) and myeloma with high frailty burden group (Group 3).

This plot demonstrates that a select group of protein biomarkers discovered in the serum samples of the myeloma patients are able to distinguish between the high frailty burden and high disease burden cohorts. These can be considered biomarkers of interest to separate the novel cohorts.

The same approach was applied to compare Group 1 (no adverse features) and Group 3 (high frailty burden). Figure 21 shows the Volcano plot difference between the two groups. 25 proteins were identified as candidate biomarkers. Appendix 5 shows the top ten proteins with the most significant different quantitative levels between the groups.

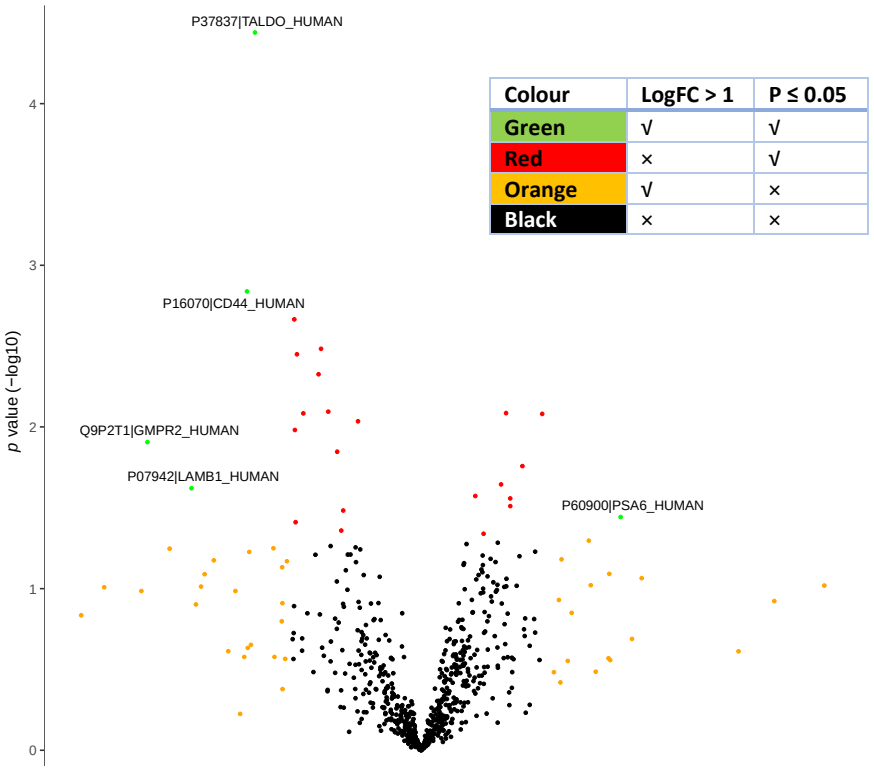


Figure 21: Volcano plot to show the log fold change in protein expression and significance of this change between the myeloma with no adverse features group (Group 1) and myeloma with high frailty burden group (Group 3)

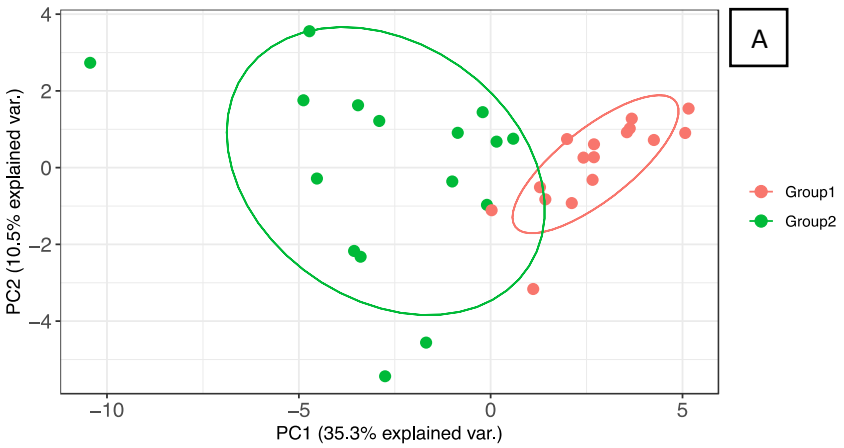
This plot demonstrates that a select group of protein biomarkers discovered in the serum samples of the myeloma patients are able to distinguish between the high frailty burden and the myeloma with no adverse features cohorts. These can be considered biomarkers of interest to separate the novel cohorts.

6.4.1.2: Random Forest modelling identifies candidate biomarkers that differentiate the three novel clinical groups

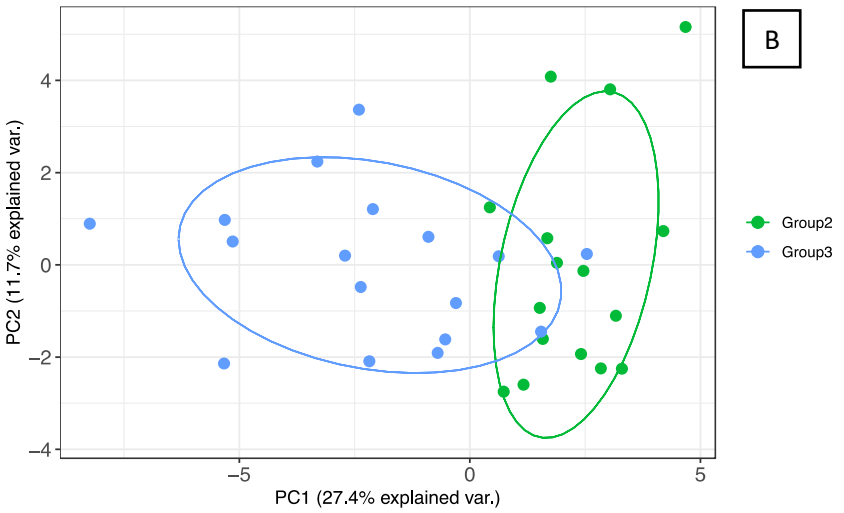
Unsupervised (PCA) did not identify clear candidate proteins to separate the groups within the overall "noise" of the large dataset. In contrast the volcano plots suggested that a supervised assessment could identify sets of proteins that had discriminatory value. The next step was to prospectively use a supervised machine learning protocol (RF) to establish those proteins that had best discriminatory value, to enable separation of the three groups: myeloma with no adverse features, myeloma with high disease burden and myeloma with high frailty burden. The benefits of using the RF approach have been discussed in the introduction section.

A RF model built with 1000 trees (three decision points in each tree) was run through 100 iterations to identify proteins which consistently came out among the most important for the RF classification model. Proteins had to be present in at least 30% of samples to be included in this analysis.

The RF analysis distinguished between the groups in pair wise analysis (Group 1 against Group 2, Group 2 against Group 3, Group 1 against Group 3) and produced a list of proteins that best distinguished between groups (Appendix 6). Myeloma with no adverse features vs myeloma with high disease burden (n=35); high disease burden vs high frailty burden groups (n=35), and no adverse features vs high frailty burden groups (n=34). PCA plots using these proteins suggested a good potential for some or all of these proteins as a clinical test: very few patients could not be classified or 'overlapped' between the two groups. while only three patients were re-classified to other groups (Figure 22).



Group 1: no adverse features
Group 2: high disease burden



Group 2: high disease burden
Group 3: high frailty burden

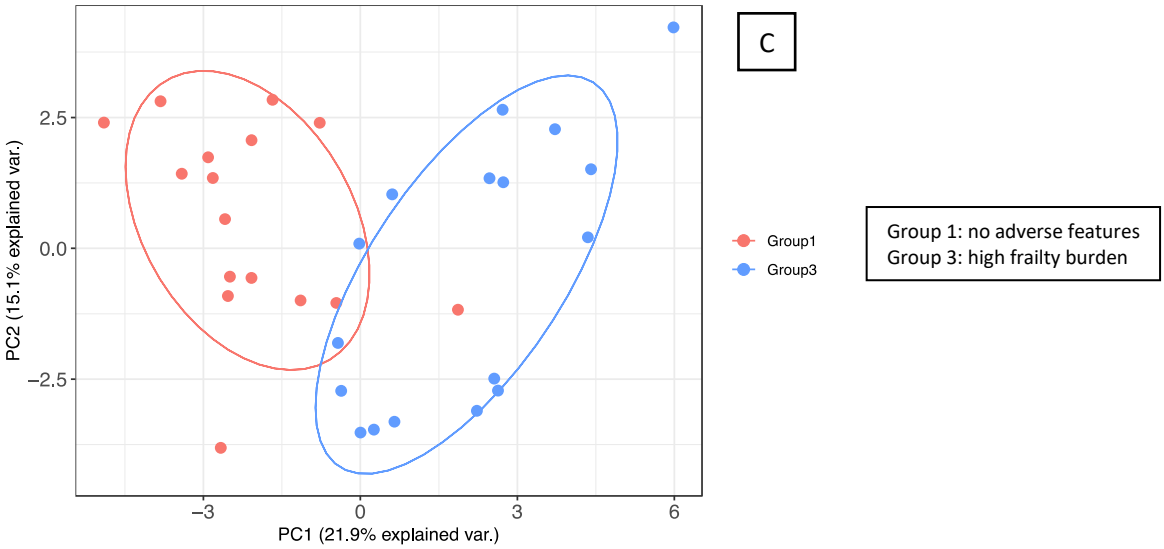


Figure 22: PCA analysis showing the separation of patients into the originally defined classification system using the important proteins selected by Random Forest Analysis

(A) Group 1- 'no adverse features' (red) against Group 2 'high disease burden' (green); (B) Group 2 'high disease burden' (green) against Group 3 'high frailty burden' (blue); (C) Group 1 'no adverse features' (red) against Group 3 'high frailty burden' (blue)

These PCA plots show that the candidate biomarkers selected by Random Forest are able to separate the patients into the three novel groups identified in chapter one. The separation is significantly improved compared to the separation on the PCA plots when using all proteins identified by mass spectrometry.

Overall, the RF analysis provides a limited protein set that can separate the three cohorts and discriminate between them although the initial analysis used relatively large numbers of proteins, some of which were represented in relatively few cases, to achieve this.

6.4.2: Results Section Two (Primary research)

6.4.2.1: Identification of candidate biomarkers found in > 80% of patients allows for improved clinical utility of a potential marker

In the preliminary analyses, RF was performed on proteins that were found in at least 30% of patient serum samples. However, to improve the clinical utility of a candidate biomarker, only proteins found in 80% of serum samples were taken forward for the next stage of analysis.

Table ***8*** shows the candidate biomarkers present in 80% of trial patient serum samples that were identified as important discriminants in the RF analysis.

To be an effective biomarker a protein needs to be readily detectable in all samples. While MS may not truly reflect the outcomes of antibody testing likely to be used in any diagnostic test, we elected to use cut-offs that allowed identification of more abundant proteins. Second, to allow statistical distinction between groups we required proteins to be represented in as many samples as possible. Therefore, the 80% cut off was chosen as a pragmatic balance between potential utility of a protein as a biomarker in this cohort (I.e. a useful biomarker must be present in a majority of patients in the cohort) and avoiding the issue of discarding potentially useful proteins that may challenging to identify if at low abundance when using the technique of mass spectrometry.

RF is a "decision tree" approach that aims to determine the best path (i.e. most useful proteins) to reach a defined outcome, using multiple iterations of the tree from a training set then re-examined using a test set to avoid overfitting. Once the decision tree is constructed, RF then selects the 'most important proteins' to discriminate between the three groups in its classification report. It combines the properties of precision (correct positive predictions relative to total positive predictions) and recall (correct positive predictions relative to actual positives) into an F1 score.[126] A protein biomarker with a high value for F1 will be useful to classify a given patient into one of the three novel groups: myeloma with no adverse features, myeloma with high disease burden and myeloma with high frailty burden.

Table 8: Candidate biomarkers derived from RF and present in >80% of trial patient serum samples

(A) Candidate biomarkers to distinguish between myeloma with no adverse features (Group 1) and myeloma with high disease burden (Group 2)
CD44 antigen
Lymphatic vessel endothelial hyaluronic acid receptor 1
6-phosphogluconate dehydrogenase, decarboxylating
Apolipoprotein D
Protein AMBP
Apolipoprotein E
Gelsolin
Inter-alpha-trypsin inhibitor heavy chain H3
Pigment epithelium-derived factor
Endothelial protein C receptor
EGF-containing fibulin-like extracellular matrix protein 1
Complement factor D
Lysozyme C
Beta-2-microglobulin

(B) Candidate biomarkers to distinguish between myeloma with high disease burden (Group 2) and myeloma with high frailty burden (Group 3)
Ribosome- binding protein 1
Interleukin enhancer-binding factor 3
60S ribosomal protein L10a
Kinesin-1 heavy chain
Ubiquitin-like modifier-activating enzyme 1
CD44 antigen
6-phosphogluconate dehydrogenase, decarboxylating
Apolipoprotein D

Apolipoprotein E
Inter-alpha-trypsin inhibitor heavy chain H3
Pigment epithelium-derived factor
Endothelial protein C receptor
EGF-containing fibulin-like extracellular matrix protein 1
Complement factor D
Lysozyme C
Beta-2-microglobulin
Gelsolin
Pigment epithelium-derived factor

(C) Candidate biomarkers to distinguish between myeloma with no adverse features (Group 1) and myeloma with high frailty burden (Group 3)
Beta-enolase
Kinesin-1 heavy chain
CD44 antigen
Transaldolase
Inter-alpha-trypsin inhibitor heavy chain H3
Beta-2-microglobulin
Pigment epithelium-derived factor
Endothelial protein C receptor
EGF-containing fibulin-like extracellular matrix protein 1
Complement factor D
Lysozyme C

The above tables above list the proteins identified by RF that are most powerful at discriminating our cohort into the novel patient groups, but only the proteins present in >80% of the patient serum are listed. This criterion has been added as it is important for a serum biomarker to be measurable in the majority of the cohort being assessed. Previously a threshold of just 30% was used.

To reduce those candidate biomarkers to those with the greatest discriminatory value we performed further discriminatory testing:

- ANOVA was performed to test the significance of the discriminatory proteins in group comparisons corrected for multiple sampling.
- T-tests were also performed to give an estimate of the individual discriminatory value of each marker.
- ROC analyses were performed between each of the groups for each protein

From this analysis the 'most significant' biomarkers were selected (Table 9 and Table 10), fulfilling the following criteria:

- ROC of > 0.7 for differentiating between at least two of the groups
- ANOVA p value < 0.05

Example ROC curves for Beta-2-microglobulin and complement factor D are also shown (Figure 22).

Table 9: Significance analysis of candidate biomarkers identified by RF found in > 90% of patient serum samples, most associated with identification of frail patients from other groups

Candidate Biomarkers		**ANOVA** (difference between 3 groups)	**T-test** myeloma with **no adverse features** Vs myeloma with **high disease burden**)	**T-test** myeloma with **no adverse features** Vs myeloma with **high frailty** burden	**T-test** myeloma with **high disease burden** Vs myeloma with **high frailty** burden	**ROC** myeloma with **no adverse features** Vs myeloma with **high disease burden**	**ROC** myeloma with **no adverse features** Vs myeloma with **high frailty** burden	**ROC** myeloma with **high disease burden** Vs myeloma with **high frailty** burden
Ribosome-binding protein 1		**	ns	ns	**	0.62	0.69	0.83
Beta-enolase		*	ns	**	*	0.52	0.83	0.74
Interleukin enhancer-binding factor 3		*	ns	ns	**	0.6	0.69	0.87
60S ribosomal protein L10a		*	ns	*	*	0.59	0.69	0.73
Kinesin-1 heavy chain		**	ns	**	**	0.53	0.75	0.79
Ubiquitin-like modifier-activating enzyme 1		*	ns	**	*	0.55	0.78	0.73
CD44 antigen		**	*	**	ns	0.79	0.85	0.73
Transaldolase		****	**	****	ns	0.78	0.91	0.62

* < 0.05
** <0.005
*** < 0.0005
**** < 0.00005

Table 10: Significance analysis of candidate biomarkers identified by RF found in > 90% of patient serum samples, most associated with identification of high disease burden patients from other groups

Candidate Biomarkers	**ANOVA** (difference between 3 groups)	**T-test** myeloma with **no adverse features** Vs myeloma with **high disease burden**)	**T-test** myeloma with **no adverse features** Vs myeloma with **high frailty** burden	**T-test** myeloma with **high disease burden** Vs myeloma with **high frailty** burden	**ROC** myeloma with **no adverse features** Vs myeloma with **high disease burden**	**ROC** myeloma with **no adverse features** Vs myeloma with **high frailty** burden	**ROC** myeloma with **high disease burden** Vs myeloma with **high frailty** burden
LYVE1	*	*	ns	ns	0.83	0.53	0.69
6-PDG	**	**	ns	*	0.82	0.72	0.76
Apolipoprotein D	***	***	ns	**	0.87	0.51	0.71
Protein AMB precursor	***	**	ns	*	0.86	0.58	0.78
Apolipoprotein E	***	*	ns	**	0.76	0.58	0.77
Gelsolin	**	*	ns	*	0.8	0.5	0.75
PEDF	****	****	*	ns	0.9	0.73	0.72
EPCR	*	**	ns	ns	0.86	0.71	0.59
EFEMP1	***	***	**	ns	0.88	0.73	0.66
Complement factor D	*	**	*	ns	0.85	0.74	0.56
Lysozyme C	**	***	**	ns	0.85	0.74	0.58
Beta-2-microglobulin	**	**	*	ns	0.85	0.73	0.63

*	< 0.05
**	<0.005
***	< 0.0005
****	< 0.00005

Legend: LYVE1: Lymphatic vessel endothelial hyaluronic acid receptor 1; 6-PDG: 6-phosphogluconate dehydrogenase, decarboxylating; PEDF: Pigment epithelium-derived factor**;** EPCR: Endothelial protein C receptor; EFEMP1: EGF-containing fibulin-like extracellular matrix protein 1

These tables list the proteins identified by RF as the most important proteins to identify the patients in each of the three novel cohorts. They then assess how significant the differences between the concentration of the identified proteins were in each of the groups using standard groupwise comparisons and ROC analysis.

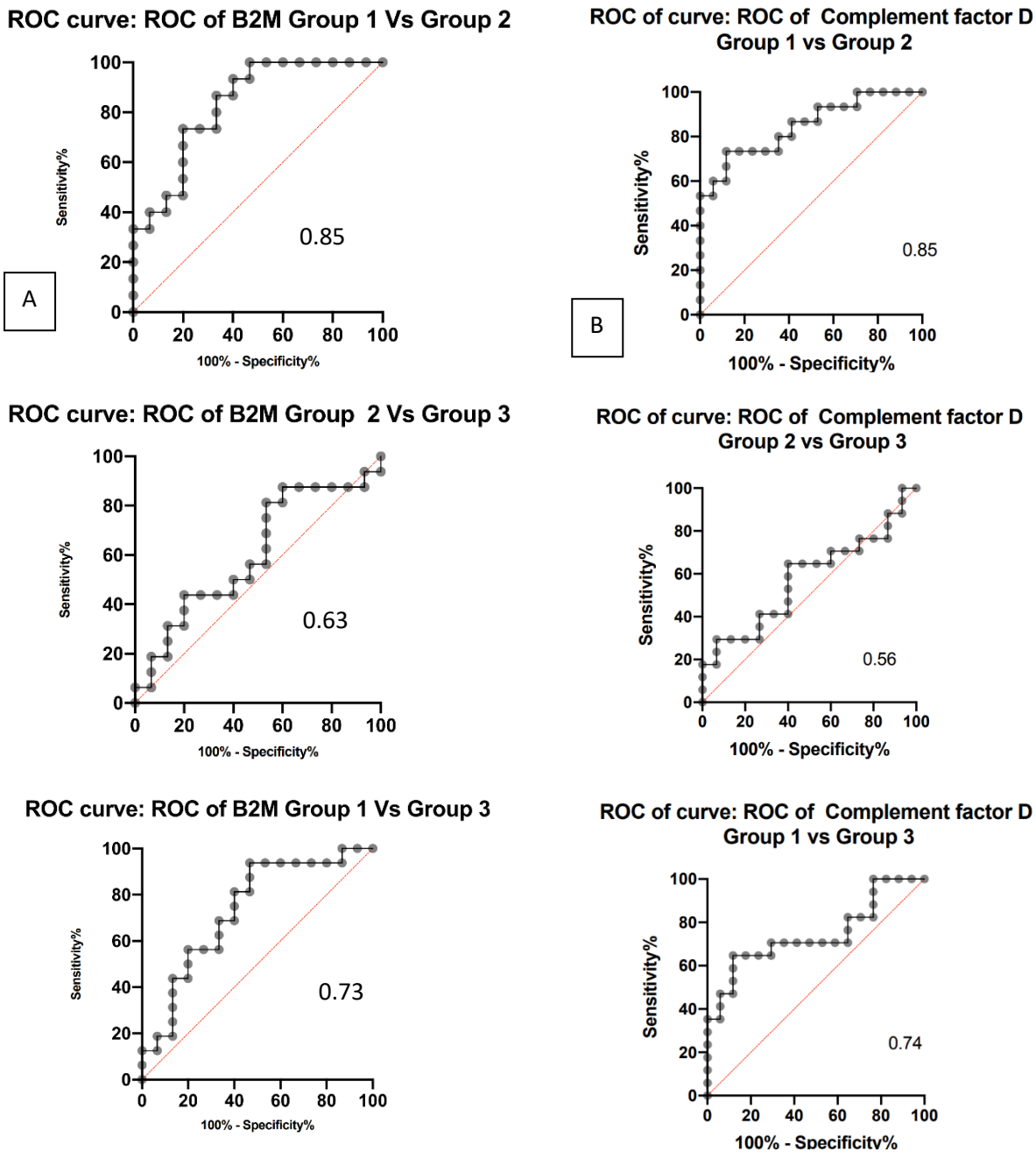


Figure 22A and B: (A) ROC Curves comparing the sensitivity and specificity of beta-2-microglobulin for classification of patients into the three novel PCA groups: Group 1 (myeloma with no adverse features), Group 2 (myeloma with high disease burden), Group 3 (myeloma with high frailty burden). **(B)** ROC Curves comparing the sensitivity and specificity of Complement factor D for classification of patients into the three novel PCA groups: Group 1 (myeloma with no adverse features), Group 2 (myeloma with high disease burden), Group 3 (myeloma with high frailty burden).

A value of > 0.7 for each ROC is considered as the cut off for performance of the sensitivity and specificity of the candidate biomarker.[181]

6.4.2.2: Biological Significance of Candidate Biomarkers

Having identified the candidate biomarkers that are most significant in classifying the patients into the three PCA cohorts, a literature review was then performed for each of the proteins. Table 11, Table 12, Table 13 and Table 14 (below) are the summary of this review.

Table 11: Summary of literature review of biology and function of proteins best able to discriminate the 'myeloma with high frailty burden' group from other individuals

Biological Function	Location of expression	Association with frailty	Association with myeloma/ cancer
Ribosome binding protein-1			
UniProt[123] • RNA binding • Signalling receptor activity • Osteoblast differentiation • Protein transport • Translation **Gene**[124] • ER proliferation • Secretory cell differentiation • Microtubule binding • No known disease associations • Expressed most: thyroid, stomach, colon, bone marrow (less) **String**[125] • Ribosome receptor	• Endoplasmic reticulum (ER) • Ribosome	• No known association	Myeloma • No known association Cancer • Over-expression in metastatic breast cancer[127] • Over-expression in lung cancer[128] • Over-expression associated with poor prognosis in bladder cancer[129] • Over-expression associated with unfavourable prognosis in colon cancer[130]
Beta-enolase (ENO3)			
UniProt	• Cytoplasm	• No known association	Myeloma • No known association

• Glycolytic enzyme/ glycolytic process • Function in striated muscle development/ regeneration **Gene** • Disease association with glycogen storage disease • Expressed most: heart, oesophagus, prostate, liver			Cancer • Low expression associated with longer OS in colorectal cancer[131] • Over-expression predicted reduced OS in DLBCL[132]
Interleukin enhancer-binding factor 3 (ILF3)			
UniProt • Biogenesis of circular RNA • Accumulates in viral infection-innate anti-viral response • Protein phosphorylation **Gene** • Binding protein • Regulates expression and stabilises mRNA • Required for T cell expression of IL-2 • Knockout models retard growth • Expressed most: bone marrow and testes **String**	• Nucleus/ nucleolus • Cytoplasm (during viral infection)	• No known association	Myeloma • No known association Cancer • ILF3 and HOXC8 co-activate cadherin 11 to promote breast cancer proliferation[133] •

• Back splicing circular RNA • Role in transcriptional/ post transcriptional processes			
60S ribosomal protein L10a			
UniProt • Cytoplasmic translation **Gene** • Protein component of 60s subunit of ribosome • Expressed most: ovaries, bone marrow	• Cytoplasm • Extracellular exosome • Membrane (less)	• No known association	Myeloma • No known association Cancer • No known association
Kinesin-1 heavy chain (KIF5B)			
UniProt • Regulates centrosome and nuclear positioning during mitosis • Axonal protein transport • Cytoplasm organisation • Cellular response to interferon gamma **Gene** • Protein binding activity • Microtubule binding and motor activity • Lysosome localisation	• Cytoplasm-cytoskeleton • Intracellular vesicles	• No known association	Myeloma • No known association Cancer • Role in tumourgenesis of breast cancer[134] • Depletion affects lysosomal distribution and accumulation of autophagososomes in cancer cells[135]

• NK mediated cytotoxicity • Positive regulation of protein localisation to membrane • Expressed: gallbladder, testis and bone marrow (less) **String** • Drives separation of nuclei and centrosomes in G2			
Ubiquitin-like modifier-activating enzyme 1			
UniProt • Involved in ubiquitin conjugation (protein modification) to mark cellular proteins for degradation. • Recruits TP53 and BRCA1 at sites of DNA damage **Gene** • Expressed most: thyroid and brain. Some expression from bone marrow **String** • Role in DNA repair	• Cytoplasm • Mitochondria • Nucleus	• No known association	Myeloma • No known association Cancer • Over expression associated with poor survival in hepatocellular carcinoma[136]
CD44 antigen			
Uniprot	• Cell membrane	• No known association	Myeloma

• Located on cell surface (extra-cellular and intracellular component) • Role in: cell adhesion, cell-cell interaction and migration • Activates T-cells • Role in haematopoiesis, inflammation • Immune response to bacterial infection • Platform for signal transduction: collagen, growth factors, cytokines **Gene** • Receptor for hyaluronic acid • Role in tumour metastasis • Expressed most: skin, appendix. Some expression in bone marrow.			• High expression of CD44 associated with increased resistance to dexamathasone[137] • High expression of CD44 associated with increased resistance to lenalidomide[138] • CD44 antigen expression upregulated on extra-medullary plasma cells[139] • IL-6 regulates CD44 cell surface expression in myeloma cells (increasing expression)[140] Cancer • Implicated in tumourgenesis of multiple types of cancer. Some discrepancy but increased expression thought to promote metastasis through multiple mechanisms[141]
Transaldolase			
UniProt • Role in pentoase phosphate pathway (involved in carbohydrate degradation) **Gene**	• Cytoplasm	• No known association	Myeloma • No known associations Cancer • Inactivation of transaldolase (affecting glucose metabolism) implicated in

• Most expressed in: bone marrow and oesophagus			oxidative stress, inflammation and carcinogenesis[142] • Overexpression in gastric cancer compared to controls (by proteomic analysis)[143]

Table 12: Summary of literature review of biology and function of proteins best able to discriminate the 'myeloma with high disease burden' group from other individuals

Biological Function	Location of expression	Association with frailty	Association with myeloma/ cancer
Lymphatic vessel endothelial hyaluronic acid receptor 1			
Uniprot • Transports ligands between intra cellular organelles and plasma membrane • Autocrine regulation of cell growth • May transport hyaluronanic acid (HA) • Binds to HA in extracellular matrices and role in cell adhesion and migration of cells through lymphatic system. • Role in wound healing **Gene** • Binds to soluble and fixed HA • Most expressed: adrenal gland and spleen. Very little marrow expression.	• Cell membrane • Cytoplasm	• No known association	Myeloma • No known association Cancer • Associated with reduced levels in serum in metastatic NSCLC than non-metastatic NSCLC[144] • Increased levels in oral squamous cell carcinoma (on histopath) associated with poor overall survival[145] • Homologue for CD44 (implicated in metastasis)[146]

6-phosphogluconate dehydrogenase, decarboxylating			
Uniprot • Oxidative decarboxylation- results NADP → NADPH • Part of pentose phosphate pathway **Gene** • Most expressed: bone marrow and oespohagus • Protects cells against oxidative damage	• Cytoplasm	• Decreased expression in frailty in plasma proteomic study[147]	Myeloma • No known association Cancer • No known association
Apolipoprotein D			
Uniprot • Ligand binding • Cholesterol binding • Lipid transporter • Mostly found in HDL > VHDL > LDL • Negative regulation of cytokines in inflammation • Negative regulation of T cell migration **Gene** • Mostly expressed: fat. No marrow expression	• Secreted	• Levels rise with age and neuropathologies (e.g. stroke, Parkinson's[148]	Myeloma • No known association Cancer • Upregulation in breast cancer, down regulation in prostate cancer. General pattern is that high expression and good prognosis- but not universal. Possible ant-tumoral function[149] • P53 family members regulate Apoliporotein D expression. Addition of recombinant ApoD in vitro was associated with inhibition of cancer cells[150]

• Expression controlled by oestrogen, glucocorticoids, progesterone, vitamin A and D			
Protein AMB precursor (alpha 1 microglobulin/Bikunin precursor)			
Uniport • Alpha-1-microglobulin-anti-oxidant, red cell protection from oxidation, anti-inflammatory • Bikunin: role in cell adhesion, extracellular remodelling, some interaction with hyaluronan **Gene** • Complex glycoprotein- secreted • It is a precursor of- alpha-1-microglobulin (transports proteins and regulation of inflammation) and bikunin (urinary trypsin inhibitor) **String** • Inhibits trypsin, plasmin and lysosomal granulocytic elastase • Inhibits calcium oxalase crystallisation.	• Secreted • Plasma membrane • ER • Nucleus • Cytoplasm	• No known association	Myeloma • No known association Cancer • Under expression associated with poor prognosis in oral squamous cell carcinoma[151]

Apolipoprotein E			
Uniprot • Lipid associated. • Role in lipid transport between organs via plasma/ interstitial fluid • Binds to many cellular receptors involved with lipids (LDL, VLDL receptors) **Gene** • Needed for catabolism of triglyceride rich lipoprotein constituents • Mutations lead to familial dysbetalipoproteinaemia • Most expressed: liver and kidney. Very little in bone marrow	• Secreted • Extracellular space/ matrix	• No association with APOE level and frailty[152], [153]	Myeloma • No known association Cancer • Increased serum levels in poor prognosis breast cancer[154] • Increased in tissue of NSCLC, ovarian, prostate, bladder cancer, colorectal cancer[155]
Gelsolin			
Uniprot • Actin modulating. Promotes monomer assembly into filaments • Role in ciliogenesis • Calcium regulated **Gene**	• Cytoskeleton (cytoplasm) • Secreted	• Higher levels associated with lower frailty in 80+ year olds[156]	Myeloma • No known association Cancer • Lower levels in head and neck cancer patients serum vs controls- candidate biomarker[157]

• Prevents monomer exchange between actin and filaments • Defects in gene cause a familial amyloidosis. • Most expressed: fat, heart. Small amount from bone marrow			• Lower levels in serum of colorectal cancer patients Vs controls[158] • Increased levels in pancreatic cancer patient serum vs control[159]

Table 13: Summary of literature review of biology and function of proteins best able to discriminate the 'myeloma with no adverse features' group from other individuals with either high frailty burden or high disease burden

Biological Function	Location of expression	Association with frailty	Association with myeloma/ cancer
Inter-alpha-trypsin inhibitor heavy chain H3			
Uniprot • Carrier of hyaluronan in serum • Role in synthesis/ degradation and binding of hyaluronan **Gene** • Role in stabilisation of extracellular matrix • Expressed in: liver only	• Secreted	• No known association	Myeloma • No known association Cancer • Reduced serum levels in colorectal patients compared to controls[160] • Down regulation/ expression in tissue of breast cancer, colon, uterus, ovary, lung, rectum, prostate[161]
Pigment epithelium-derived factor			
Uniprot • Induces differentiation in retinoblastoma cells • Inhibits angiogenesis **Gene** • Mutations lead to osteogenesis inperfecta • Most expressed: gallbladder, liver, fat. Very little from marrow **String**	• Secreted	• No known association	Myeloma • In vitro PEDF shown to inhibit survival and proliferation of VEGF- exposed myeloma cells[162] Cancer • Supresses angiogenesis and directly inhibits cancer cell proliferation[163]

• Potent inhibitor of angiogenesis			• Low PEGF associated with advanced grade and poor survival in a variety of cancers (meta-analysis)[163]
Endothelial protein C receptor			
Uniprot • Role in coagulation cascade • Binds activated protein C and enhances activation via interaction with thrombomodulin **Gene** • Mutations associated with VTE and late foetal loss • Most expressed: spleen, colon, bladder	• Membrane	• No known association	Myeloma • No known association Cancer • EPCR and PAR-1 interaction promote angiogenesis, vascular protective tube formation and tube formation. Some evidence that EPRC activation inhibits mets- conflicting data[164]
EGF-containing fibulin-like extracellular matrix protein 1			
Uniprot • Binds EGFR leading to downstream activation of signalling pathways. • Role in cell adhesion/ migration **Gene** • Upregulated in malignant gliomas • Most expressed: fat, gallbladder, bladder. Not in bone marrow **String**	• Extracellular space/ matrix	• No known association	Myeloma • No known association Cancer • Expression down regulated in HCC[165] • Gene may function as a tumour suppressor in gastric cancer[166]

• Role in chondrocyte differentiation			
Complement factor D			
• Homologous to factor C1s in classical complement pathway • Role in complement activation in the alternate pathway (cleavage of factor B). becomes C3 convertase • Immunity **Gene** • Role as an adipokine- secreted by adipocytes • Mutations underlie complement factor D deficiency- associated with recurrent bacterial meningitis • Most expressed: fat and colon	• Secreted	• No known association	Myeloma • No known association Cancer • Increased expression in cutaneous squamous cell carcinoma[167]
Lysozyme C			
Uniprot • Bacteriolytic function • Associated with monocyte/ macrophage system	• Secreted	• No known associations	Myeloma • No known association Cancer • No known association

• Enhance immune response- innate **Gene** • Target bacterial cell wall • Missense mutations associated with renal amyloid • Most expressed: stomach, salivary gland, bone marrow			
Beta-2-microglobulin			
Uniprot • Part of MHC I molecule • Presents peptide antigens to immune system **Gene** • Almost ubiquitous on cell surface of nucleated cells • Can form amyloid fibrils in disease • Most expressed: spleen, lymph node, lung, bone marrow	• Secreted • Cell surface	• Increased B2M in serum associated with higher frailty in elderly population[168], but only weakly associated as a predictor of frailty[169]	Myeloma • Established predictive biomarker in myeloma. • Raised B2M- worse prognosis, especially when combined with other factors in ISS[170]/ R-ISS[171]

Table 14: Uniprot ID of coding gene of candidate protein biomarkers

Protein	Uniprot ID of coding gene	Protein	Uniprot ID of coding gene
Ubiquitin-like modifier-activating enzyme 1	UBA1	Alpha-1 microglobulin/bikunin precursor	AMB1
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	Lymphatic vessel endothelial hyaluronic acid receptor 1	LYVE1
Complement factor D	CFD	EGF-containing fibulin like extracellular matrix protein 1	EFEMP1
Pigment epithelium derived factor	SERPINF1	60s ribosomal protein L10a	RPL10A
6-phosphogluconate dehydrogenase	PGD	Apolipoprotein E	APOE
CD44 antigen	CD44	Beta-enolase	ENO3
Transaldolase 1	TALDO1	Beta-2-microglobulin	B2M
Endothelial protein receptor 3	PROCR	Interleukin enhancer-binding factor 3	ILF3
Ribosome-binding protein 1	RRBP1	Kinesin family member 5b	KIF5B
Lysozyme C	LYZ	Apolipoprotein D	APOD
Gelsoin	GSN		

6.4.2.3: Candidate biomarkers can be separated into subcategories depending on location of expression

The previously mentioned candidate biomarkers can be subcategorised based on the site of expression, as demonstrated in Figure 23. The proteins most associated with high frailty burden are all cellular proteins, either found within the cytoplasm or on the surface membrane. Proteins associated with high disease burden are more heterogeneously expressed but are predominantly extracellular proteins.

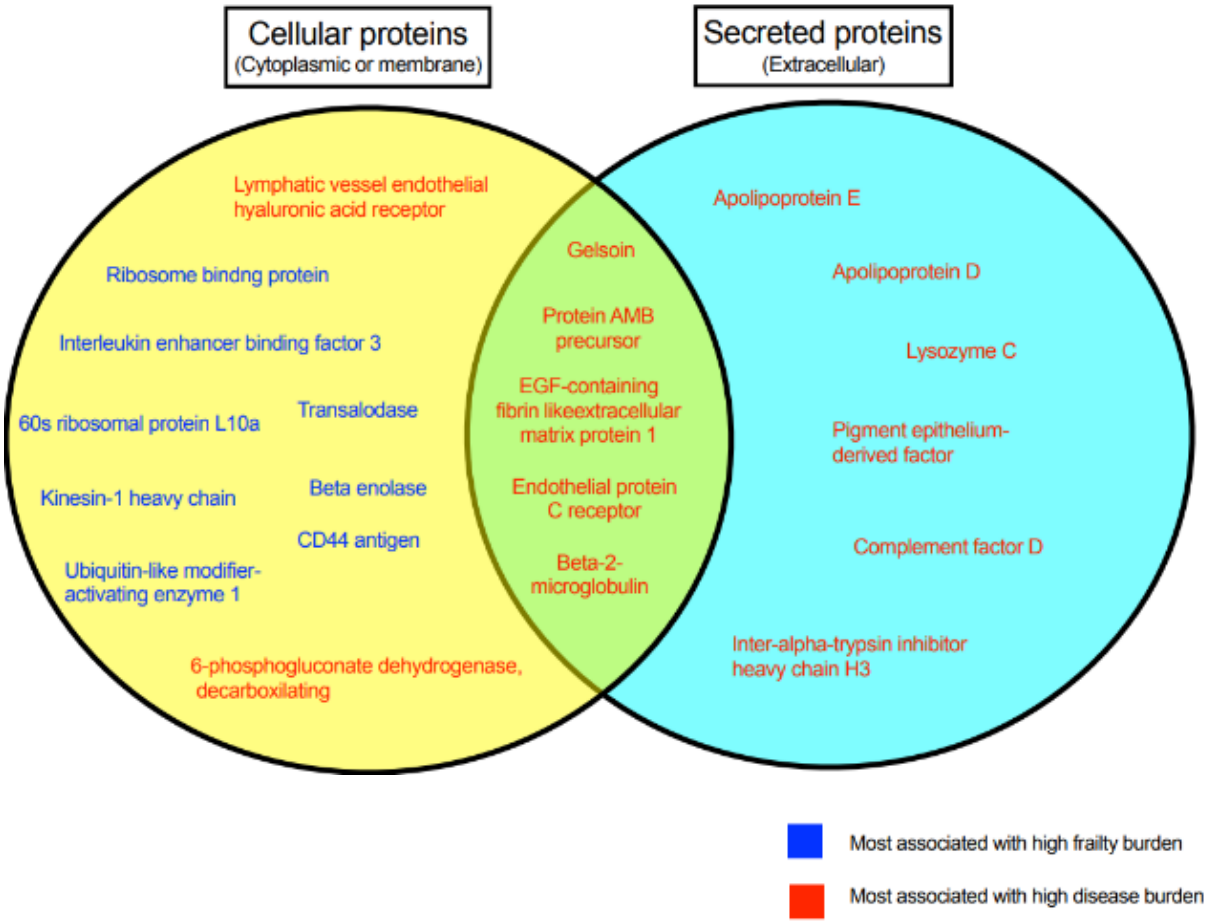


Figure 23: Location of expression of proteins most associated with frailty (blue) and location of expression of proteins most associated with high disease burden (red)

This Venn diagram demonstrates that proteins associated with frailty are almost exclusively cellular proteins but proteins associated with disease burden have a more heterogeneous expression profile

The candidate biomarkers can also be categorised based on biological function, as shown in Figure 24. The biomarkers most associated with high frailty burden are almost exclusively associated with protein and carbohydrate metabolism (with the exception of CD44). The

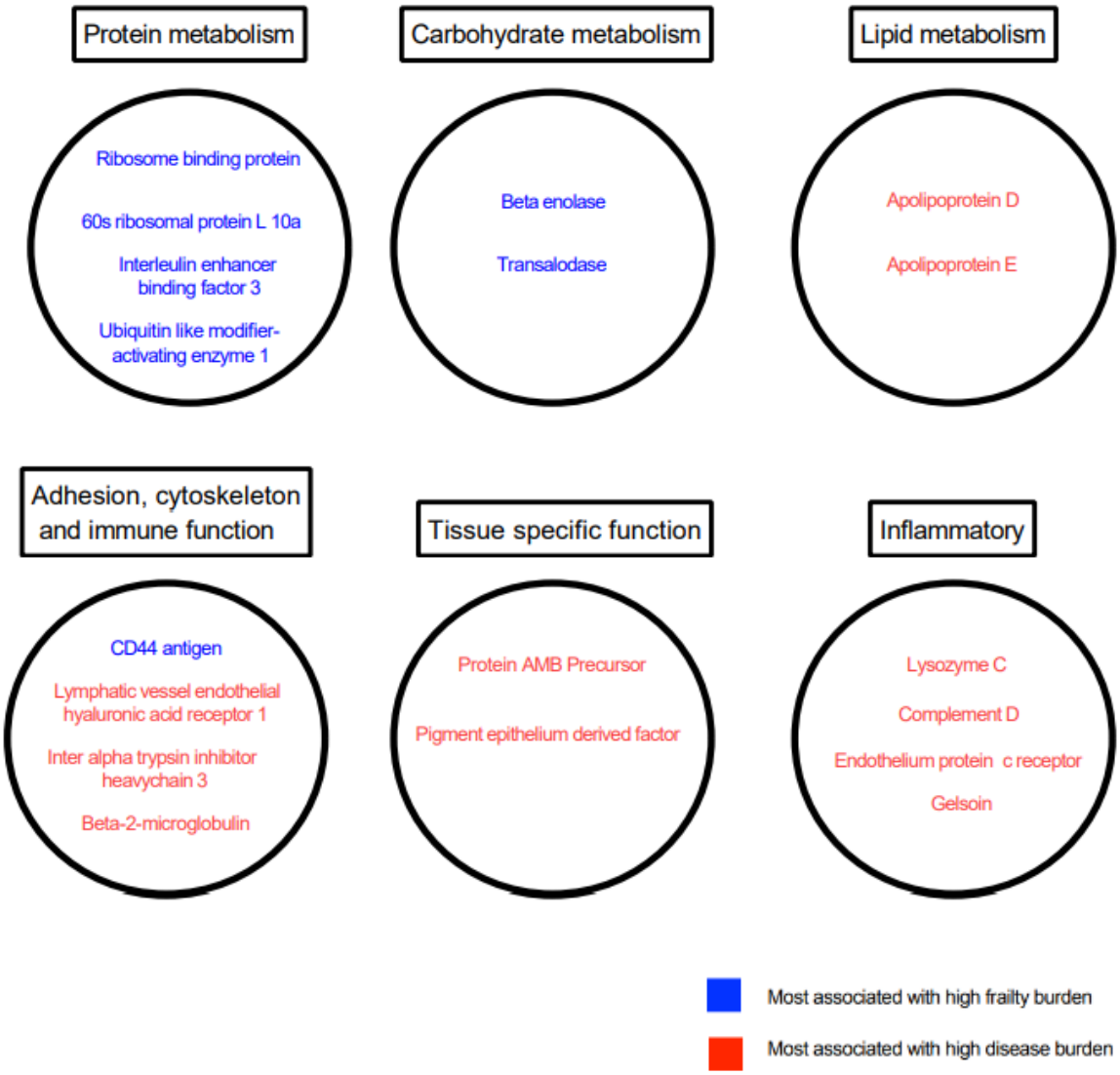


Figure 24: Summary of function of proteins most associated with frailty (blue) and high disease burden (red)

This figure shows that almost all of the biomarkers associated with frailty are involved in protein and carbohydrate metabolism but biomarkers of disease have a wider range of biological function.

6.4.2.4: Candidate biomarkers lack strong biological associations with each other

String Bioinformatics software[125] was used to assess for biological association between the proteins identified as candidate biomarkers. This software assesses a variety of associations including proximity of gene loci, disease associations, presence in related biochemical pathways and concurrent presence of proteins in literature. A 'high confidence interval' with a minimum required interaction score of 0.700 was used. The 'protein map' produced is shown in Figure 25.

The only significant interactions are the proteins involved in carbohydrate metabolism (TALO2, ENO3 and PGD), those proteins involved in cell adhesion (hyaluronic acid associated) (CD44 and LYVE1) and two ribosomal associated proteins (ILF3 and RPL10A). The majority of the proteins are not associated and there is no specific association of the proteins identified to differentiate each of the candidate biomarker groups. This is an expected finding as the search was investigation for associations between secreted proteins, rather than those is a specific pathway.

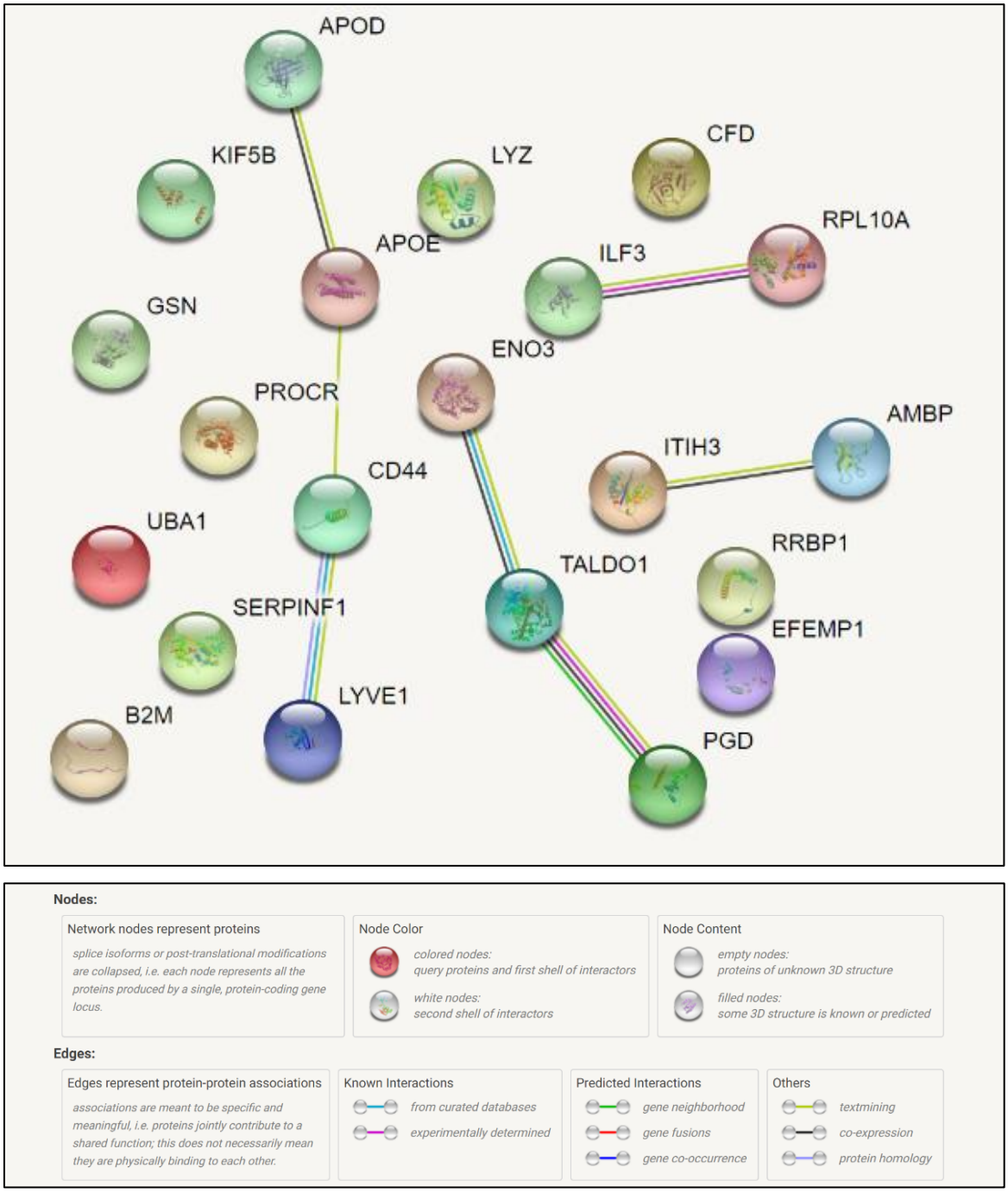


Figure 25: String Bioinformatics map of associations between proteins identified to be significantly different between novel groups identified by PCA

String bioinformatics software failed to demonstrate any significant associations between the proteins identified as candidate biomarkers.

6.5: Discussion

Protein abundance in the serum of myeloma patients differed from that of healthy controls: We sought to investigate serum biomarkers in multiple myeloma patients, aiming to identify whether subset of patients with different prognosis would have a qualitative and quantitative difference in proteins identified. This aim however assumes that serum protein levels in myeloma patients differs from that of the general 'healthy' population.

Chanukuppa *et al* performed proteomic analysis on the serum of myeloma patients using three proteomic approaches, including SWATH-MS (used in our analysis)[114]. They identified 116 'significantly differently abundant' proteins in their myeloma cohort compared with their control group. Our data also supports that the serum proteome of myeloma patients is different with initial comparison of the serum of health controls compared to four pooled serum samples of the myeloma cohort having a significantly different variance than healthy control cohorts.

Proteins identified in the serum of the myeloma cohort have some differences when compared to those previously reported in the literature: There are limited data regarding the serum proteome of myeloma patients available, with Chanukuppa *et al* being the most recent and comprehensive[114]. Their study identified five proteins: haptoglobin, kininogen 1, transferrin, albumin and apolipoprotein A1 as the biomarkers most differentially expressed between control and myeloma cohorts. There are however differences between their methodology and that used for the Manchester cohort. Of note, they assessed both bone marrow interstitial fluid (BMIF) and serum using three proteomic methodologies (2D-DIGE, iTRAQ and SWATH-MS), using healthy volunteers and patients with non-haematological malignancies as controls and did not immunodeplete their samples of the most abundant proteins prior to analysis. In contrast, we used only 'healthy controls' (with no known malignancy), assessed only serum, and immunodepleted the serum of the most abundant proteins prior to analysis.

The immunodepletion stage of our method was to remove the most abundant proteins in human serum. The logic was that due to their high concentration in normal human serum, would mask the lower abundance proteins and reduce the discrimination between our

novel PCA groups. Of the five proteins Chanukuppa *et al* identified as 'significant', only kininogen 1 was not removed in our immunodepletion stage. Kininogen 1 was not identified as a candidate biomarker to discriminate between our three groups.

In summary, the aim of this study was not to differentiate healthy 'control' plasma from myeloma plasma. We did however show that the serum proteome was significantly different in these cohorts. Methodological differences likely account for the differences in the proteins identified in our study and in other reports.

SWATH-MS was the best suited approach to identifying a biomarker in myeloma: SWATH-mass spectrometry has been shown to be an accurate, sensitive and specific method of undertaking targeted proteomics[172] and has previously been shown to be an effective method assess protein biomarkers in myeloma[114]. Furthermore, it was a pragmatic selection as it was the most readily available technique for this research. Tandem mass spectrometry was favoured over conventional mass spectrometry in this research. Conventional mass spectrometry uses a single chamber and analyses molecules based on their mass to charge ratio.[173] The advantage of tandem mass spectrometry is that it uses a two chamber technique. The first chamber allows for selection of a specific set of ions in the first chamber, before they are analysed in the second chamber. This allows for superior specificity of identified ions and 'superior analytical accuracy'.[174]

Proteomics has also been used by others to assess disease progression or treatment response. The clonotypic method uses liquid chromatography and mass spectrometry to monitor patient specific peptides on the immunoglobulin heavy chain region[175]. This offers a patient specific MRD assessment. While this approach, and others such as monoclonal immunoglobulin rapid accurate molecular mass (miRAMM) have many advantages, they are suitable only for disease monitoring and assessment. They do not appear to have a role in prognostication or be useful in the subclassification of patients based on disease activity or frailty.

Another potential technique for protein identification is using DIGE software. This is a gel electrophoresis model[176] but it requires fluorescent labelling of proteins during their

synthesis. Given that this research was using stored serum samples, this cannot be considered a viable technique.

It could also be argued that an enzyme linked immunosorbent assay (ELISA) technique could have been employed. It is an established, sensitive and affordable technique used to detect and quantify proteins[177]. However, ELISA requires a knowledge of the proteins that are likely to be present in a sample. This research had the aim of identifying proteins, thus mass spectrometry was the more appropriate test.

SWATH-MS the best approach to identify possible biomarkers of frailty: There is some precedent to using SWATH-MS in the assessment of the serum proteome in ageing. Bjelosevic *et al* used the technique on the serum of healthy neonates, children and adults.[178] They used PCA to demonstrate that the serum of each age group had different co-variances and split into distinct groups. Other studies have shown a trend to increasing concentrations of proteins involved iron transport, apoptosis, haemostasis and immune response[179]. Santos-Lozano *et al* used liquid chromatography and tandem mass spectrometry to demonstrate different serum protein levels in healthy elderly patients compared to those with significant frailty.[180] While there is limited data on both, this would suggest that proteomic analysis with SWATH-MS has utility in identifying frailty and age-related changes when assessing human serum. These studies have not however sought to disambiguate proteomic changes in a cohort of myeloma patients with different burdens of frailty.

In summary there is limited data support for the use of proteomic analysis in assessment of ageing, frailty and myeloma, and a small number of studies have used SWATH-MS technology in this setting. However, no reports have used SWATH-MS in a myeloma cohort to identify those with either a high disease burden or a high frailty burden. This seems a suitable area of research and will discussed further.

There a subset of proteins that can reliably classify myeloma patients into the novel classification system: RF analysis has provided a shortlist of proteins that are useful in the classification of our cohort. However, this still equates to 104 proteins required to classify

the cohort into one of the three groups. All but one of these proteins (beta-2-microglobulin) are not currently standard of care tests in the diagnosis and management of myeloma. Thus, if this classification is going to be of clinical use (rather than purely for research purposes), it must be possible to identify the three groups using a smaller number of measurable proteins.

In order to identify the proteins that have the most 'discriminatory power', further statistical analysis was undertaken. Firstly, it was decided that the proteins must be present in greater than 80% of the samples analysed. This decision was taken because it is important for a potential biomarker to be present and measurable in the vast majority of patients for the test to be clinically useful.

Having applied the above criteria, we identified the following:

- 8 proteins that best discriminate the 'high frailty burden group' from the other patients
- 12 proteins that best discriminate the 'high disease burden group' from the other patients

These proteins are listed in Table 9 and Table 10.

Subsequently, in the interests of statistical rigor, an ANOVA was performed to assess for a statistically significant difference between the values of each of the proteins in the three groups, with a cut off of 0.05 used for significance.

T-tests were the performed between the vales for each of the proteins in the different groups in turn, to assess for a statistically significant difference. It could be argued that the addition of the T-test to the ANOVA is unnecessary, but it was performed to demonstrate the significance of the difference in concentration of each of the proteins between each of the groups in turn. For example, a candidate biomarker may strongly differentiate the frail from the no adverse features group but be less good at discriminating the frail from the high disease burden groups.

Finally, ROC analysis was performed between each of the groups in turn, with a ROC value of 0.7 being used as the cut off. The value of 0.7 was chosen as an arbitrary value that provides an acceptable balance between sensitivity and specificity.[181] It would be hoped

that by combining several biomarkers, the sensitivity and specificity would be improved in a future model.

Some of the candidate biomarkers have been previously identified in literature in the context of myeloma and frailty; others are novel findings: The candidate biomarkers identified (listed in Table 9 and Table 10) are a heterogeneous group. B2M was the only protein identified as a candidate biomarker that has a well-established prognostic significance in multiple myeloma.[170] As such it is already incorporated into the ISS and R-ISS prognostic scoring systems. It has also been studied in the context of frailty and has been found to be elevated in frail inpatients.[168] Given that it has been shown to be raised in both frailty and myeloma, it is thus logical that RF has identified it as a candidate biomarker that best differentiates the 'no adverse features' cohort from the other patients.

B2M is a component of the major histocompatibility complex I molecule, found on the surface of all nucleated human cells. B2M levels rise, either if renal excretion is reduced, or due to increased production. The hypothesis is that in inflammatory conditions, (including cancer and myeloma) there is increased B2M shedding into the serum of affected patients.[182] However, its functional role as an antigen presentation molecule in the immune system would not intuitively lead to the conclusion that it is such a powerful prognostic marker in myeloma.

It is therefore rather unsurprising that the proteins identified do not have an obvious or direct association to multiple myeloma. It is also worth noting that we are assessing the serum proteosome, rather than intracellular proteomic changes. The consequence of this is that we have measured only exosomal proteins, cell membrane proteins or those released during apoptosis.

Figure 23 and Figure 24 are summary figures that identify broad trends in the structure and function of the candidate biomarkers. It should be noted that these observations are merely trends of provisional data, rather that statistically significant observations. It appears that the majority of the biomarkers that most discriminate the frail group from the other cohorts are predominantly intracellular proteins (7 out of 8), with just one being found on the membrane. However, the proteins that best discriminate the no adverse features group and

the high disease burden group are more commonly secreted. There is no intuitive or evidence-based reason for this pattern at this stage of analysis or from the literature review.

There was a wide range of functions attributed to the protein biomarkers identified. These functions included DNA/ RNA/ protein metabolism and transport, carbohydrate metabolism, lipid metabolism and transport, immune function and specific tissue function. Five of the eight proteins that best discriminate the frail group were involved in protein metabolism, but other trends were hard to identify. Once the most important candidate biomarkers are identified, it will be important to review the pathophysiology and cause of the variation of the protein concentration at the cellular level. An overview of the functions and associations for the high potential candidate biomarkers can be found in the literature review in Table 11, Table 12 and Table 13.

Overall, there is no strong or established biological reason why the 'high potential biomarkers' should be significant in the context of myeloma and frailty. Further research would be established to define a causal relationship of the best candidate biomarkers.

6.6: Conclusion

In the first chapter, we set out the paradigm of current clinical stratification of myeloma patients. The diagram has been reproduced in Figure 26.

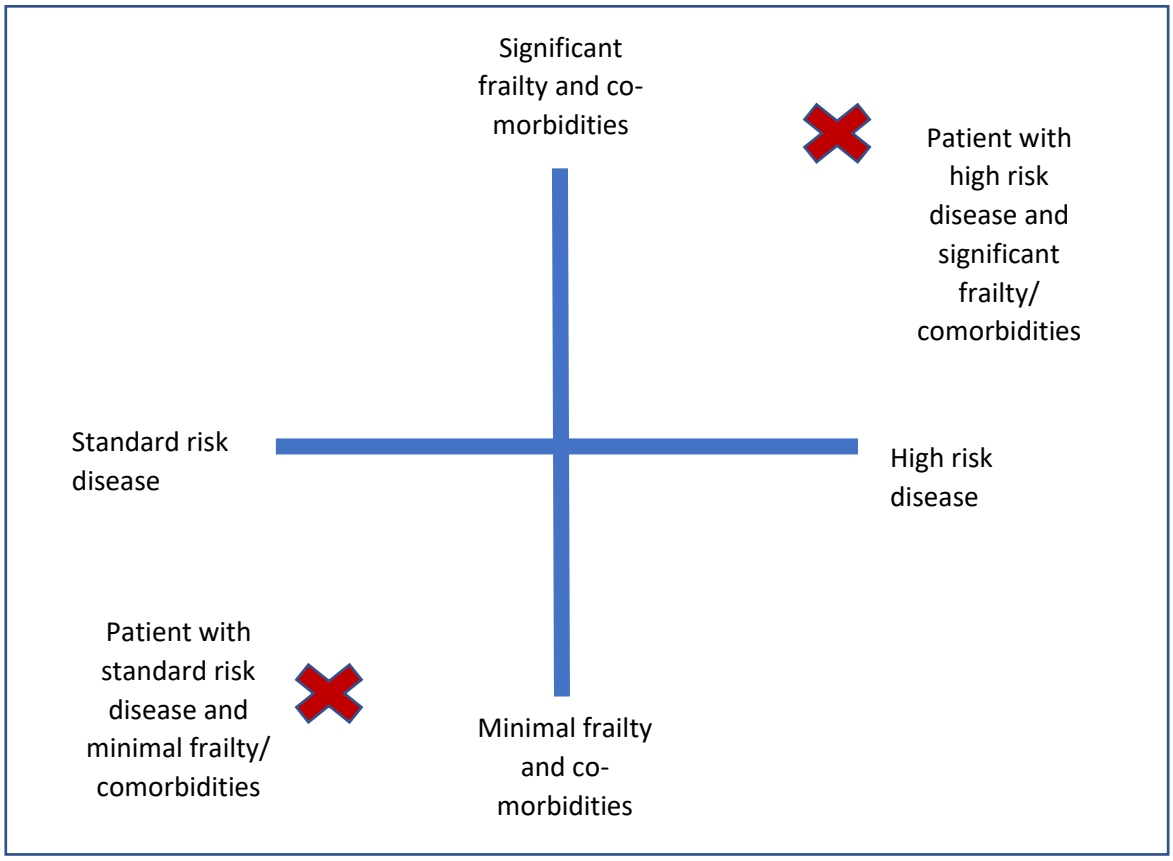


Figure 26: Visual representation of the utility of different scoring systems in multiple myeloma

This figure demonstrates that conventional scoring systems in myeloma are able to identify the combination of frailty and co-morbidities, as well as disease activity, though they seldom combine all these characteristics in an easy to use score or algorithm

We stated that an individual presenting with a new diagnosis or a relapse of myeloma would have a combination of factors that contribute to the clinical course of their disease. At the extremes, there are those with low vulnerability-no comorbidities, no frailty and standard risk disease. At the other end of the spectrum there are individuals with high vulnerability-significant co-morbidities and frailty with high-risk disease. The biomarkers identified in this research have gone some way to identifying and subcategorising this difficult to treat cohort of patients, with distinctive differences in overall survival. There is scope for further

assessment and validation of these biomarkers to produce a clinically feasible and useful prognostic model, that incorporates the effects of disease and frailty on the individual.

If a prospective study were to corroborate the clinical significance of the novel groupings (in terms of early mortality and overall survival), the question of how these biomarkers could be used to guide management becomes pertinent. The biomarker derived novel groups would offer potential option of informing upfront treatment allocation. They could be used to define the transplant eligible and transplant ineligible populations. They could also permit upfront dose modifications/ attenuations in frail patients, or those with high early mortality. What cannot however be extrapolated from this data is if the biomarkers could offer a dynamic assessment of frailty phenotype as patients progress through their treatment pathway although this may represent a potential application and would be a powerful tool in Clinical Trial evaluation of potential interventions.

7: Chapter 3- Quality of Life in Multiple Myeloma

7.1: Introduction

Assessment of outcome is an important constituent of evidence-based medicine in the evaluation of cancer targeted therapies. Common outcome measures include clinical, radiological or laboratory parameters, along with progression free and overall survival.[183]

Indeed, recent advances in management of multiple myeloma have resulted in improved overall survival rates.[184] However, these measures fail to account for patient factors such as quality of life. This is of importance in the management of multiple myeloma, as treatment is not curative but extended survival is now anticipated. Furthermore, previous studies have demonstrated that myeloma patients experience significant disease and treatment specific side effects, affecting their quality of life.[185] Their quality of life will also be affected by their response to therapy.[186]

A 'Health Technology Assessment' commissioned by the NHS recommended that outcomes considering 'the patient's perspective' should be included in clinical trials.[187] The rational for this statement is that quality of life data gives an insight into the 'physical function, social and emotional wellbeing of patients and help the clinician decide which patients will derive the most overall benefit from a treatment.[188]

If such outcome measures are to be included in clinical trials, they must be validated for the disease, have internal and external consistency, and provide clinically useful information that will impact patient care. These properties are however challenging in the context of patient derived quality of life outcomes. Perhaps because of the complexity of collecting this data, there is some scepticism amongst oncologists regarding the use of quality-of-life surveys, with some believing clinical judgement is a less time consuming and acceptable surrogate measure.[189]

While there is no universally accepted definition, physical, emotional, social and occupational wellbeing are commonly accepted constituents or 'domains' of quality of life.[188] The number of quality-of-life papers published has increased significantly in recent years.[183] The problem is therefore not a lack of tools or instruments to assess quality of life, rather which one to pick. The European Organisation for Research and Treatment of Cancer

(EORTC) have produced two quality of life questionnaires validated in myeloma: the QLQ-C30[190] and QLQ-MY20[191] questionnaires. The initial iteration of the QLQ-C30 questionnaire was written in 1987, with the most recent version (version 3.0) published in 2000. It is a 'cancer focussed' rather than myeloma specific questionnaire but has been validated in multiple myeloma[192]. The QLQ-MY-20 is a multiple myeloma specific questionnaire, validated for the assessment of this patient group in 2007[193]. The range of possible scores in the MY-20 questionnaire is 19-80, with higher scores recorded by patients with more severe symptomatology. This chapter will explore the clinical utility of quality-of-life questionnaires, specifically in the context of multiple myeloma.

7.2: Aims and Objectives

Aim

To determine whether the QoL could be effectively assessed in our cohort and to explore the potential relationship between QoL and clinical parameters of our patient cohort, adding knowledge the relationships of QoL and clinical parameters in myeloma

Objectives

1. To assess the feasibility of performing standard QOL assessments (EORTC QLQ C-30 and MY-20) in our real-world myeloma population
2. To compare our patient responses with those of historical studies of trial patients to examine their performance in a 'real world cohort'
3. To evaluate the relationship between quality-of-life data, clinical parameters, and survival

   To interrogate whether the 'symptom scales' (question groupings) in the MY-20 questionnaire using unsupervised analysis to determine which measure similar biological processes

7.3: Methods

The study dataset also studied clinical and quality of life data of patients who consented to enter this observational arm of the study, either at the time of diagnosis with multiple myeloma, or at relapse. Patients were invited to fill in two previously validated questionnaires: the European Organisation for Research and Treatment of Cancer (EORTC) QLQ C-30 and Myeloma-20 (MY-20) questionnaires, consisting of 50 questions in total. Patients were asked to complete the questionnaires during/ after clinic and were assisted by a trained research nurse if requiring assistance.

The strength of the data set was that it provided a large amount of contemporaneous data, allowing for retrospective exploration of relationships and associations between quality of life and a variety of clinical parameters and outcome measures for the same patients (as discussed in Chapter 1).

All patients who continued to attend clinic at the six-month timepoint (after diagnosis/ relapse) were invited to fill in the same two questionnaires. All the above data was collected prospectively, then analysed retrospectively on conclusion of the study.

EORTC MY-20 Questionnaire Questions and 'Scales'

Table 15 (below) is an abbreviated version of the MY-20 questionnaire used in the assessment of quality of life in this study.

Table 15: Abbreviated version of EORTC MY-20 Questionnaire

Scale	Question	
Disease symptoms (*Pain symptoms)	Have you had bone aches/ pains?	1
	Have you had pain in your back?	2
	Have you had pain in your hip?	3
	Have you had pain in your arms or shoulder?	4
	Have you had pain in your chest?	5
	If you had pain, did it increase with activity?	6
Side effects of treatment (*Systemic symptoms)	Did you feel drowsy?	7
	Did you feel thirsty?	8
	Did you feel ill?	9
	Have you had a dry mouth?	10
	Have you lost any hair?	11
	Did you have tingling in your hands or feet?	13
	Did you feel restless or agitated?	14
	Have you had indigestion or heartburn?	15
	Have you had burning or sore eyes?	16
Body Image	Have you felt physically less attractive as a result of your disease or treatment?	17
Future perspective	Have you been thinking about your illness?	18
	Have you been worried about dying?	19
	Have you worried about your health in the future?	20

***** For the purposes of this study these 'symptom scales have been renamed' but the components of the scales are unchanged from the original MY-20 questionnaire.

The nomenclature of the 'scales' was changed since the questionnaires asked patients to report a domain called 'side effect of treatment' symptoms prior to receiving any therapy. By definition, any symptoms reported at this stage did not reflect treatment. The decision was therefore made to rename 'disease symptoms' as 'Pain symptoms' and 'side effect symptoms' with 'Systemic symptoms'

The scores patients were asked to give were based on 'the extent to which (they) have experienced these problems or symptoms during the last week', on a scale of 1-4.

- 1 point Not at all
- 2 points A little
- 3 points Quite a bit
- 4 points very much

Statistical Analysis

Descriptive data analysis was performed using Excel software. GraphPad Prism software was used for other statistical analysis.

Mann-Whitney testing was used to compare the unpaired non-parametric data where there were two analysis groups. An example of this was when comparing the quality of life of patients with either 'normal' or 'abnormal' beta-2-microglobulins.

Kruskal Wallis testing was used to compare unpaired non-parametric data where there were three analysis groups. This method was used for example when comparing the quality of life of the three IMWG Frailty groups: 'fit', 'intermediate-fit' and 'frail'.

The Wilcoxon test was used to compare non-parametric paired data at baseline and at 6 months. Log rank (Mantel-Cox) test was used to compare survival curves.

7.4: Results

7.4.1: Collection of quality of life data is feasible but has significant challenges in our cohort

The database contained 91 data sets for patients either at diagnosis or relapse who were part of the clinical study and were confirmed to have sufficient baseline data to perform meaningful analysis. Of these 91 patients, 72 (79%) patients completed the MY-20 questionnaire at baseline, with 31 (34%) completing the MY-30 database. Of this cohort only 44 patients the questionnaires at the six-month timepoint. Patients unable to complete the questionnaire at the second timepoint included those who had died or transferred their care to a different centre. Of these 28 (64%) of patients completed the MY-20 survey, with 22 (50%) completing the MY-30 questionnaire. These data are summarised in

Table 16.

Table 16: Number of completed MY-20 and MY-30 Questionnaires

Group	Number (n)	Percentage (%)
Number of patients in cohort	91	
Number of these patients with complete MY20 at baseline	72	79%
Number of patients with completed MY30 at baseline	31	34%
Number of patients available to complete survey at 6 months	44	48%
Number of patients with completed MY20 at 6 months	28	31% (of original cohort) 64% (of possible responses)
Number of patients with completed MY30 at 6 months	22	24% (of original) 50% (of possible responses)

This table shows that there was an acceptable return rate for the MY20 questionnaire, but a much inferior return rate for the longer MY30 questionnaire.

There were no significant differences in the demographics of the patients who completed the survey compared to those who did not in terms of age, performance status or line of therapy (Table 17).

Table 17: Demographics of patients who completed compared to those who did not complete MY-20 Questionnaire

	Completed survey	Failed to complete survey
Mean Age	68	65
Median Performance Status	1	1
Median line of therapy	2	2

This table demonstrates that there was no large difference in demographic of those who completed the questionnaire and those who failed to return it.

7.4.2: Data collected for the MY-20 Questionnaire was of good quality

Only two data points were missing from the questionnaires in total. This excludes question 12 which was a subsidiary question to question 11. A number of patients answered this question despite a negative response to the previous question. The responses to this question were removed from analysis.

7.4.3: Results of MY-20 Questionnaire showed myeloma has a significant impact on self reported QoL

The questions in the MY-20 questionnaire can be found in the methods section and in the introduction and Appendix 7. Table 18 summarises the data collected in the MY-20 questionnaire at baseline.

The data has been separated into the 'scales' suggested in the Cocks *et al* validation paper: Symptom Complex 1 (SC1), Symptom Complex 2 (SC2), body image and future perspective.[193] The presence of 'bone aches and pains' (mean score 2.6) and 'thinking about (their) illness' (mean score 2.6) were the questions with the highest mean score, while the presence of hair loss was the least troubling symptom (mean score 1.2). Every possible score on the scale (1-4) was used by the cohort for each of the questionnaires. The mean overall

questionnaire score for all patients was 35.33 (SD 11.87).

Table 18: Responses to MY-20 Questionnaire at baseline

Scale	Question	Responses	Mean	Median (Range)	SD
Pain symptoms	1	72	2.6	3 (1-4)	1.1
	2	72	2.4	2 (1-4)	1.1
	3	72	1.9	2 (1-4)	1.0
	4	72	1.7	1 (1-4)	0.9
	5	72	1.5	1 (1-4)	0.8
	6	72	2.0	2 (1-4)	1.0
Systemic symptoms	7	72	1.9	2 (1-4)	1.0
	8	72	1.9	2 (1-4)	1.0
	9	72	1.8	1 (1-4)	1.0
	10	72	2.0	2 (1-4)	1.1
	11	72	1.2	1 (1-4)	0.5
	13	72	1.5	1 (1-4)	0.9
	14	72	1.7	2 (1-4)	0.9
	15	72	1.4	1 (1-4)	0.7
	16	72	1.3	1 (1-4)	0.5
Body Image	17	72	1.7	1 (1-4)	1.0
Future perspective	18	72	2.6	2 (1-4)	1.0
	19	72	1.8	1 (1-4)	1.0
	20	72	2.5	2 (1-4)	1.0

The Cocks et al[193] validation paper proceeded to undertake a linear transformation of mean scores for each scale. The data for this cohort is shown in Table 19. This was to allow for all scores to be compared on the same scale (0-100). This linear transformation however, it has the confusing property that 'higher scores' for Pain Symptoms and Systemic symptoms

indicate worse symptoms, while for body image and future perspective scales these scores indicate a better view of outcome.

Table 19: Linear transformation of mean scores for each Scale

Scale	Transformed mean	Standard deviation	Range
Pain symptoms [a]	33.10	25.36	0- 77.88
Systemic Symptoms [a]	21.24	19.56	0- 70.37
Body image [b]	75.46	34.48	0- 100
Future perspective [b]	56.64	29.91	0- 100

a Higher scores suggests more significant/ worse symptoms

b Higher scores suggest better function

This table shows that patients with myeloma reported a different relative burden of symptomatology on each of the different scales used in the questionnaire.

This information could perhaps be better displayed in Figure 27 (below). The 'body image' and 'future perspective scales have been inverted (100-score) to allow the means and standard deviations to be compared on the same figure. In this figure, a high score represents more significant symptomology for all scales. It shows there to be a wide spread of symptoms in all of the scales, with large standard deviations. The mean scores are highest in the 'pain symptom' scale (thought to be related to disease specific symptoms) and in the 'future perspective' scale (relating to anxieties about diagnosis and treatment).

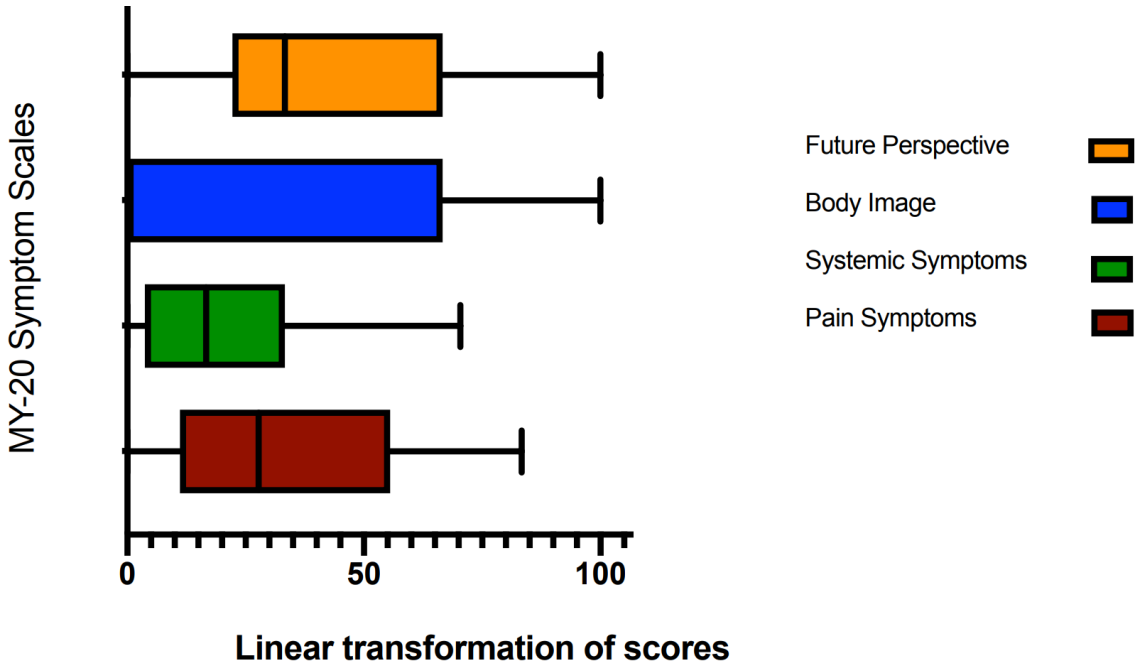


Figure 27: Linear transformation of MY-20 scores to allow comparison between each of the 'symptom scales'

This figure shows the relative burden of symptoms reported by patients in the MY-20 questionnaire, with concerns over future perspective leading to the highest burden on quality of life.

7.4.4: Association between MY-20 and clinical parameters/ scores

The aggregate scores from each of the scales (pain symptoms, systemic symptoms, body image, future perspective) were compared to baseline patient variables for a statistically significant difference between the groups. These clinical variables were selected to be representative of different aspects of the disease.

ECOG performance status was selected as a proxy for general overall health and function. Table 20 shows the overall MY-20 scores of patients, separated by performance status. There is a sequential increase in total MY-20 score as the performance status declines.

Table 20: ECOG Performance Status Score and MY-20 Score

EOCG performance Status Score	Average total MY-20 Score (MRI cohort)	Number
0	30	29
1	36	26
2	40	10
3	46	6
4	52	1

This table shows a trend of more significant reporting of symptomatology related to myeloma as the performance status declines. There are however only small numbers in the ECOG 3 and 4 groups

To facilitate the analysis, patients were grouped by ECOG performance score into Performance Score 0/1/2 (Good PS Group) or Performance Status 3/4 (Poor PS Group) for analysis. There was a statistically significant difference between the Good PS and Poor PS group in the 'Pain symptom scale' (p= 0.0430), in the 'Systemic symptom Scale' (p= 0.002) and 'body image scale' (p= 0.0161), with all symptoms being worse in the Poor PS group (Figure 28). There was a non-significant trend to increased concern regarding future perspective in the poor PS Group (p= 0.0607).

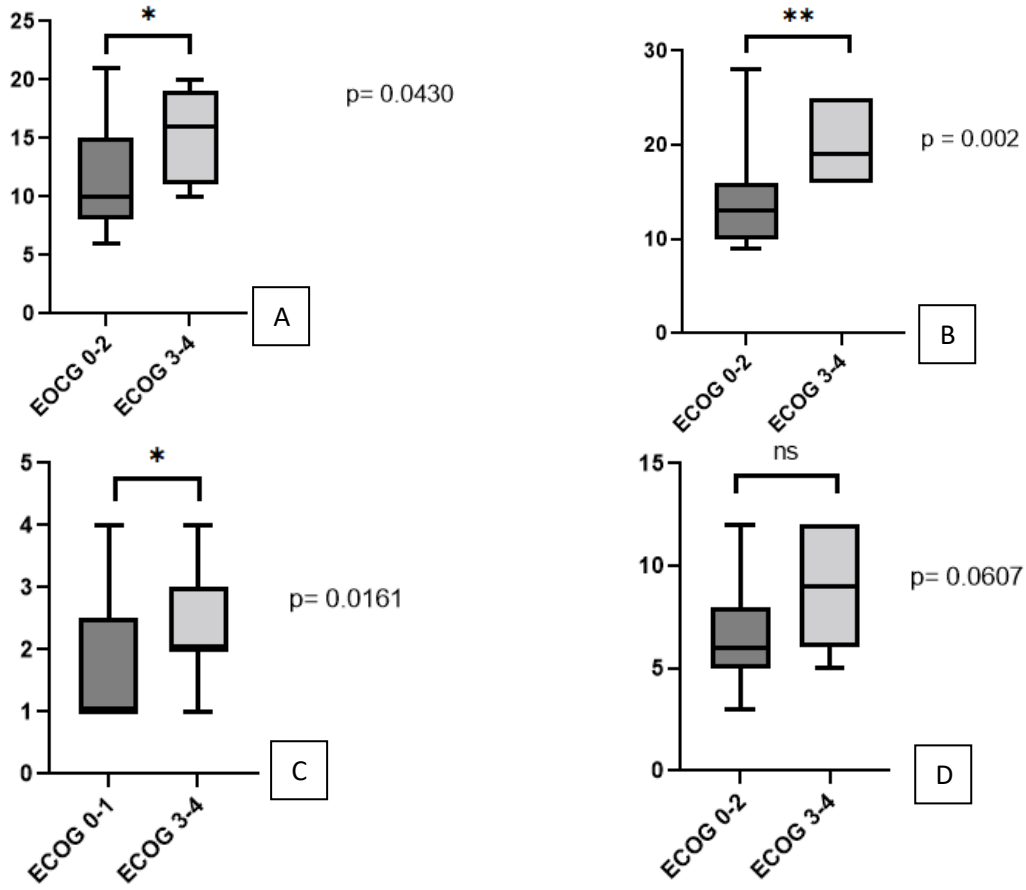


Figure 28: Associations between ECOG performance Score and MY-20 'Symptom Scales': (A) Pain symptoms (B) Systemic symptoms (C) Body Image (D) Future Perspective

This figure demonstrates that there was a significant difference between the ECOG Groups 1/2 compared with the ECOG 3/4 group in terms of burden of symptoms in the pain symptom, systemic symptoms and body image scales, suggesting an increased symptom burden as performance status declines

Baseline beta-2-microglobulin (B2M), a proxy marker for disease activity, was interrogated for a potential association with the MY-20 disease scales (thus quality of life). Patients were grouped into B2M < 4 g/l or ≥ 4 g/l. The only statistically significant association was with the 'Systemic symptom scale' (p= 0.0020). Patients in the low B2M group had lower symptoms in the 'Systemic symptoms scale'. This was confirmed by the fact that the multiple myeloma International Staging System (ISS) (another surrogate marker for disease activity) also showed a statistically significant association between ISS and quality of life on the 'Systemic

symptom' side effect scale (p= 0.001). No other statistically significant associations with ISS were identified.

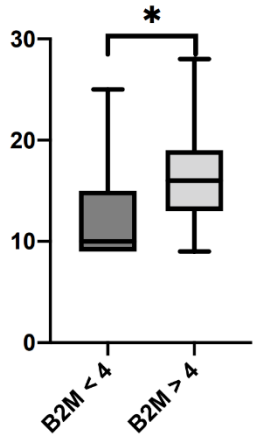


Figure 29: Association between B2M level and 'Systemic Symptom Scale'

This figure demonstrates that there was a statistical difference in the systemic symptom scale between patients when classified by B2M (with a cut off of 4g/l)

The IMWG Classification, selected as a measure of frailty burden in the cohort, was also used to assess for association. The only statistically significant association between IMWG classification at baseline and MY-20 symptom scale was with the 'Systemic symptom scale' (Kruskall Wallis p= 0.0025). There was a significant difference between the scores between the fit and frail groups (Mann Whitney p= 0.0014) and the intermediate and frail groups (Mann Whitney p= 0.0160), with an inferior quality of life reported as frailty increased. (Figure 30).

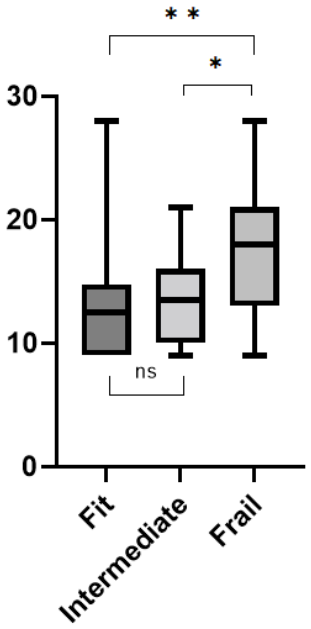


Figure 30: Association between IMWG Classification and 'Systemic symptom Scale'

This figure demonstrates that there was a statistical difference in the systemic symptom scale between patients when classified by the IMWG frailty classification, with more burdensome symptoms in the most frail group

Our novel PCA groups were also interrogated for differences in quality of life. Figure 31 (below) is a heatmap of the patients who had quality of life data available. Each individual patient is represented as a point of the graph. Lower overall MY-20 scores (indicating a good self-reported quality of life) are coloured green, and those with higher overall MY-20 scores (indicating worse self-reported quality of life) coloured red, with others in between. There is a significant difference in the MY-20 scores between our novel clinically fit group and the frailty group (p=0.003), as well as between the high disease burden group and the frailty group (p= 0.003). No difference in quality of life was found between the clinically well group and the high disease burden group (p= 0.35).

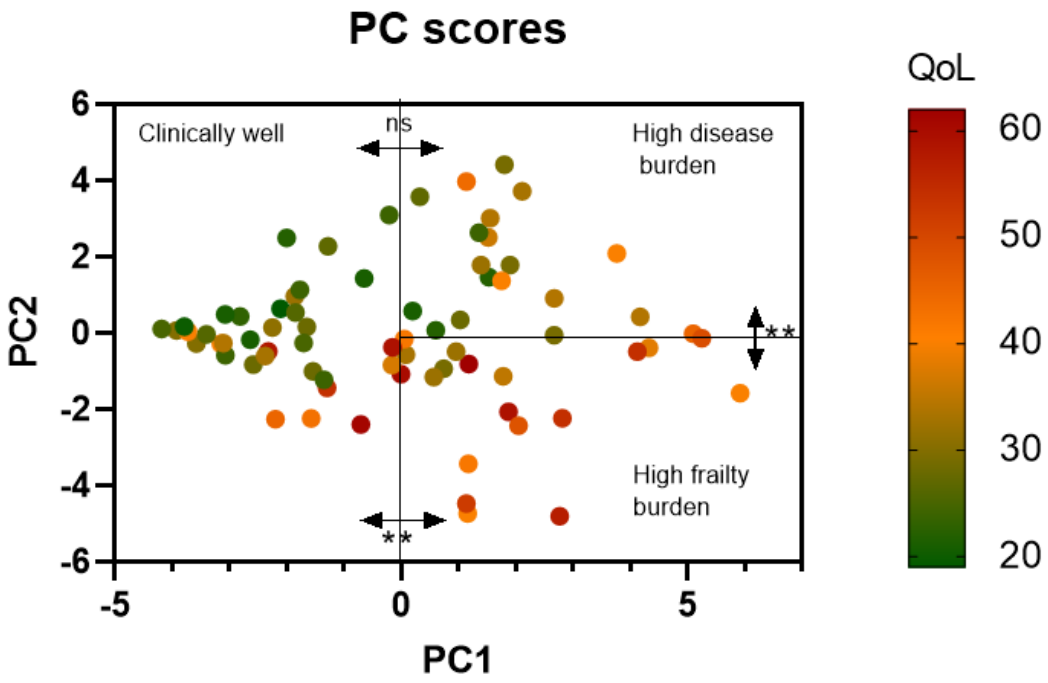


Figure 31: Heatmap of the Quality of life MY-20 Scores of the cohort, separated by novel PCA groupings

This heatmap shows that there is a significant difference in the self-reported quality of life scores of the cohort between the novel PCA groupings. Mann Whitney testing showed there was inferior quality of life reported in the frailty group when compared to the clinically well group (p= 0.003) and the high disease burden group (p=0.003). No difference was found in QoL between the clinically well and high disease burden groups.

Patients were also assessed for difference in quality of life based of line of therapy. This is a surrogate marker for the impact of myeloma therapy and the effects of chronic disease. Linear transformation was used to compare the cohorts for a minimally important difference in change of quality of life between patients treated with different lines of therapy (Table 21 and

Table ***22***). The minimally important clinical difference is a measure of clinically relevant difference in quality of life, rather than statistical significance. This will be scrutinised in more detail in the conclusions section.

There was an improvement in 'future perspective' quality of life between patients starting second line therapy compared to newly diagnosed patients. This was however reversed when comparing second and third-line patients, with third line patients have a worse future perspective. There was also a worse reported body image, along with worse 'Pain symptoms' and 'Systemic symptoms' between patients starting forth line therapy, compared to those commencing third line therapy.

Table 21: Mean transformed MY-20 scale scores for patients treated with different lines of therapy

Line of therapy	Pain symptom Score	Systemic symptom score		Mean body image score	Mean future perspective score
First	34.77	18.11		72.84	53.91
Second	25.93	25.93		79.63	64.81
Third	32.48	22.51		79.49	52.14
Fourth	44.44	34.26		62.50	51.39

This table demonstrates that there is no obvious step wise deterioration or increased burden in symptoms with each cycle, but rather a more complicated pattern of reported symptomatology

Table 22: Differences in mean transformed MY-20 scale scores treated with different lines of therapy, with corresponding minimum important difference

	Second – first line	Third – second line	Fourth – third line	Minimum important difference (MID)
Future perspective score difference	10.91	-12.68	-0.75	9
Body image score difference	6.79	-0.14	-16.99	13
Pain symptom score difference	-8.85	6.55	11.97	10
Systemic symptom score difference	7.82	-3.42	11.75	10

This table shows that a minimally important clinical difference in burden of symptoms was detectable in some of the symptom scales, mostly deteriorating between third and fourth line therapy. Green shading denotes improvement, pink shading denotes inferior QoL.

For future perspective and body image, a negative score denotes worse quality of life
For 'Pain symptoms' and 'Systemic symptoms', a negative score denotes improved quality of life. Shaded boxes denote changes that have reached the minimum important clinical threshold.

7.4.5: Longitudinal analysis showed QoL to be variable at diagnosis compared to the six month time point

There were 22 patients who achieved at least a partial remission (PR) and in whom MY-20 data was available at the six-month time point. Table 23 summarises the data. In this patient subgroup there is a non-statistically significant reduction in pain symptoms (p=0.19), with an associated non-significant increase in systemic symptoms (p=0.10). There was however a significant improvement in symptoms relating to concerns over 'future perspective' (p= 0.02) (Figure 32)

Table 23: Baseline and six-month data for patients achieving at least a partial remission

	Baseline Mean (SD)	6-month mean (SD)	P value
Pain symptoms	12.2 (4.4)	10.9 (3.3)	0.19
Systemic symptoms	14.0 (4.5)	15.8 (4.3)	0.10
Body image	1.6 (1.0)	2 (0.9)	0.15
Future perspective	7.4 (2.8)	6.3 (1.5)	0.02
Overall Score	33.2 (11.0)	35.0 (7.2)	0.67

This table demonstrates patients achieving at least a partial remission failed to have a significant improvement in reported symptoms in three of the four scales but did have an improved future perspective

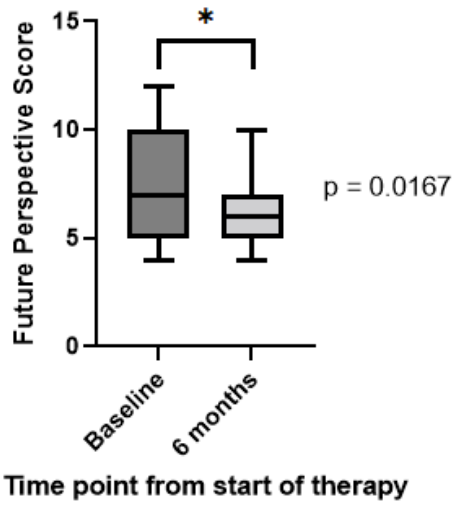


Figure 32: Baseline and six-month data for symptoms associated to concerns over future perspective in patients achieving at least a partial remission

This figure shows a significant improvement in future perspective score for those patients achieving at least a partial remission after 6 months of therapy

In subgroup analysis (n=11), there was a statistically significant reduction in 'Pain symptoms' in patients who were newly diagnosed and receiving first line therapy (p=0.0469). This subgroup was drawn from all first line patients who completed the MY20 survey at baseline. There was no difference in the other scales. The significant reduction in pain symptoms was not reproduced in relapsed patients receiving second line therapy and beyond.

7.4.6: There was an association with survival and overall MY-20 Questionnaire Score

The MY-20 questionnaire was also used to look for differences in overall survival depending on QoL score. Transformed Quality of life scores from three of the four scales were compared for a statistical difference in overall survival. This was not performed in the 'body image scale' as it contained only one question.

The cut-offs used to separate patients into groups were based on frequency of distribution plots for each individual scale. All three scales had a bimodal distribution of values so cut-offs were based on the frequency of distribution histograms, an example of which is demonstrated in Figure 33. The cut-off scores for each of the scales and cumulative MY-20

score are shown in Table 24. This allowed separation into two clinical groups: 'Good' and 'Poor' quality of life

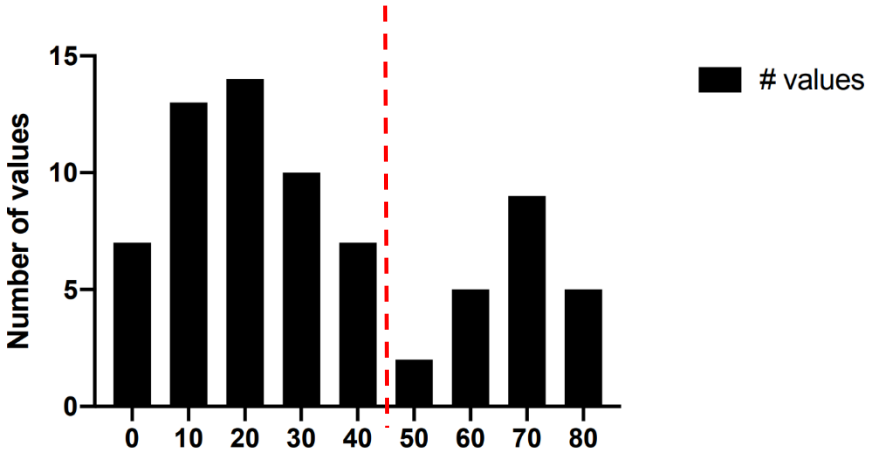


Figure 33: Histogram of distribution of scores for the Pain scale questions. Dotted line indicated score 'cut-off' to delineate the 'Good' and 'Poor' quality-of-life groups

This figure indicated the bimodal distribution of scores for the pain symptom complex, explaining the cut off values used in analysis.

Table 24: Cut-off values for different QoL Scales and overall MY-20 Score

Scale	Good QoL Range	Poor QoL Range
Symptom Complex 1	≤ 40	> 40
Symptom Complex 2	≤ 25	> 25
Future perspective*	≤ 40	> 40
Total MY-20 Score (raw untransformed score)	≤ 40	> 40

** To allow comparison the transformed future perspective score was subtracted from 100. This allows all scales to have 'good' QoL as low scores and 'poor' QoL as high scores*

There was a significant difference in overall survival when using the overall MY-20 score. Those scoring ≤ 40 points had a longer overall survival (p= 0.04). However, despite this, there was no statistically significant difference in quality of life between the 'good' and 'poor' quality of life groups in each of the scales (not individual scales or scores). There was

however a non-significant trend to improved overall survival in the 'good' quality of life group in all three scales interrogated (Figure 34).

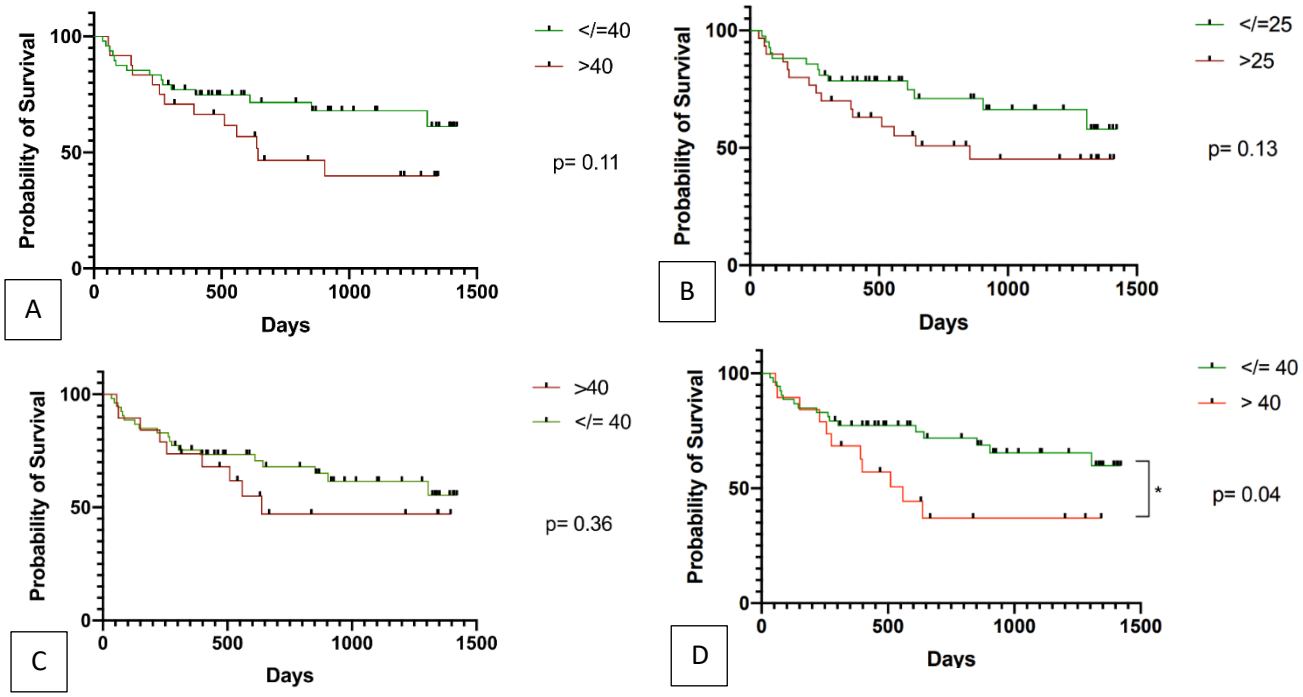


Figure 34: Kaplan Meir charts for overall survival in different MY-20 QoL groups (A) Pain symptoms, (B) Systemic symptoms, (C) Future perspective, (D) Overall MY-20 Score

This figure demonstrates that only the aggregate MY20 score shows a statistically significant difference in overall survival. That is to say that those patients with a score of less than or equal to 40 had a significantly longer overall survival than those with a score of more than 40 on the MY20 questionnaire.

7.4.7: Principal Component Analysis for MY-20 questions revealed differences between the Eigenvectors produced by PCA and the scales proposed in the original MY-20 study

PCA was used to interrogate the MY-20 questionnaire results. As previously discussed, PCA uses the covariance of different components in the dataset to group those variables that share similar characteristics (i.e. in this case questions that address similar or related symptoms). Since PCA is an unsupervised analysis. clusters of questions will be identified using the PCA algorithm purely based on common co-variances, and taking no acount for the previously defined 'scales'.

These 'scales' were initially produced using multi-trait scaling analysis (discussed in more detail in the conclusion). The purpose of using PCA on the MY-20 data was therefore to assess if the pre-existing scales showed clustering and similar co-varience, suggesting that they are indeed measuring related aspects of quality of life.

PCA separates data using different PCs termed PC1, PC2, PC3 etc. based on how much of the variability in the data set they can explain. Each PC will separate data in a different way, with each of the MY-20 questions (Eigenvectors) contributing a different 'loading'. Figure 35 shows the relative contributions of each of the MY-20 questions to PC1 and PC2. The further the arrow extends along the x-axis, the more the question contributes to the variance of PC1. The further the arrow extends along the y-axis (in either a positive or negative direction), the more the question contributes to the variance in PC2.

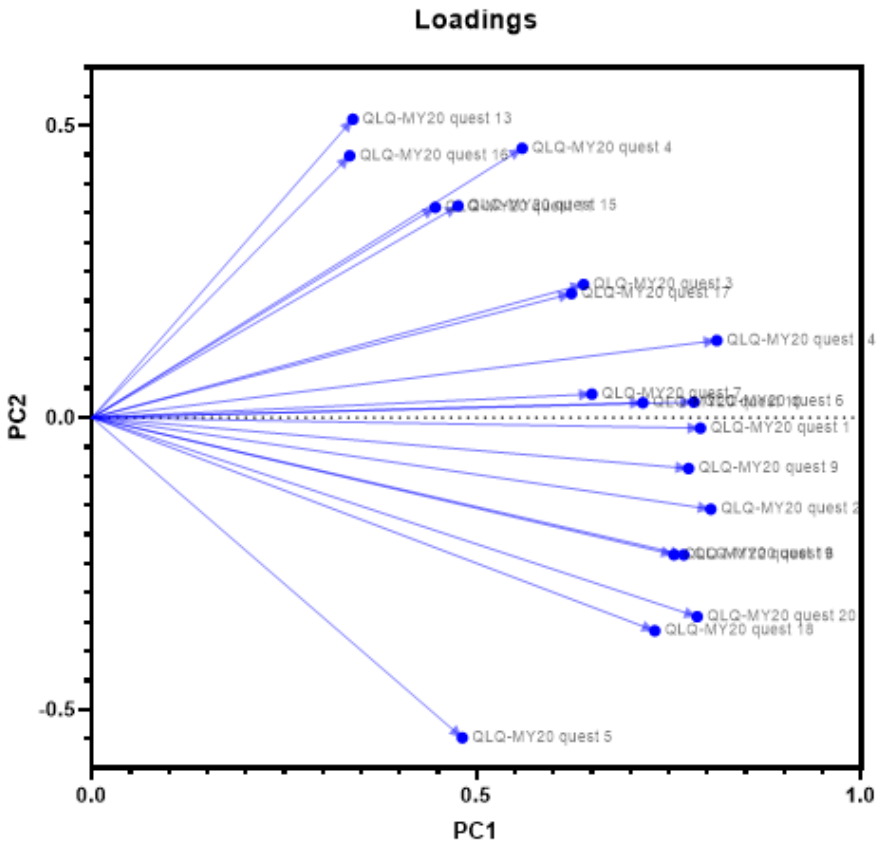


Figure 35: Eigenvector loadings for MY-20 QoL questionnaire for PC1 and PC2

On review of the data PC2 produced the most meaningful separation of data. The questions relating to 'chest pain' and 'tingling and numbness in hands and feet' were responsible for the most variation in the cohort (hence the largest loading values). The relative contribution of each of the questions to variance (i.e. the loading values) is shown in Figure 36.

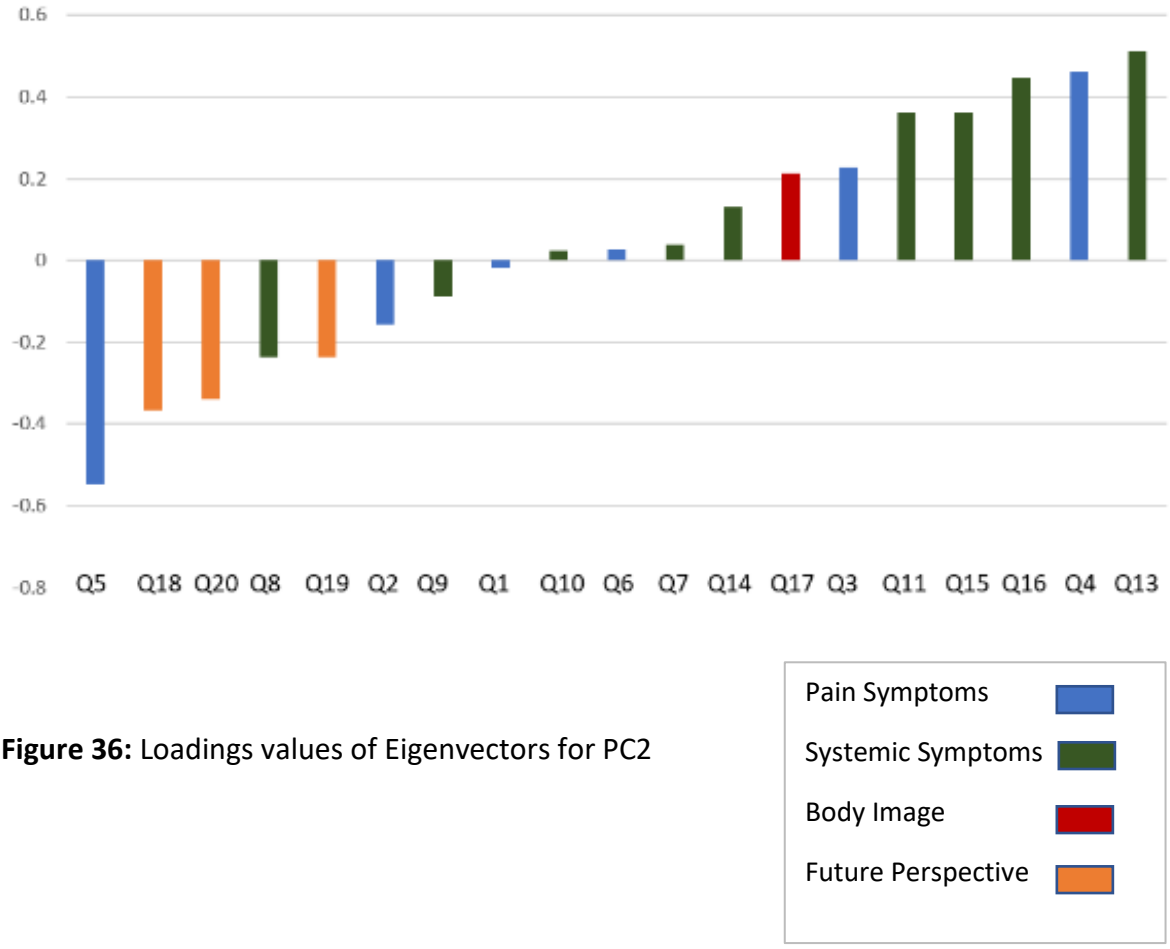


Figure 36: Loadings values of Eigenvectors for PC2

This figure shows the future perspectives questions clustering together, as do the systemic symptoms questions. This suggests these questions share a similar co-variance in this cohort

Figure 36 shows clearly that questions in the 'Future Perspective Scale (Orange) clearly cluster together, suggesting they share similar covariances. Questions in the 'Systemic symptoms (Green)' also show clustering, with the notable exception of Question 8. This question relates to the symptoms of thirst. Questions in 'Pain symptoms (blue)' are however dispersed with no clear clustering. This suggests that in our cohort, these questions may not be asking about a group of closely associated symptoms. This is contrary to the data from the Cocks *et al* validation study, in which multi-scale analysis demonstrated these symptoms were closely associated[193]. Finally in our cohort, Question 5 (relating to chest pain) is responsible for the most variation. This will be analysed further in the discussion.

7.5: Discussion

Selection of an appropriate quality-of-life questionnaire was found to be essential to achieve acceptable completion rates: The practicality and feasibility of employing the questionnaires is of prime importance. This specifically examined by the completion rate of each of the two questionnaires, both at the baseline and six-month time point. If the overall completion rate is poor or is skewed in such a way that a specific patient group is underrepresented, the utility of the tool is much diminished.

In our patient cohort 79% of patients enrolled in the survey completed the MY-20 survey at baseline, with only 34% of the QLQ C-30 questionnaires completed. Cocks *et al* undertook a field study to validate the MY-20 questionnaire, where they recruited 477 patients enrolled in clinical trials in Europe and USA[193]. They achieved a combined completion rate of both questionnaires of 72%. So, although our cohort of patients completed more of the MY-20 questionnaire than the historical cohort, our completion rate of the C-30 questionnaire was significantly lower.

The reasons for this difference are likely multifactorial. Firstly, our cohort are a 'real world' cohort, meaning they provide a genuine sample of patients being treated at a myeloma care centre. While some were enrolled in prospective trials, others were receiving stand of care therapy. It could therefore be hypothesised that trial patients are more motivated or engaged with the process of supplementary testing. Our patient cohort was also asked to fill out these questionnaires while attending clinic, for example between phlebotomy and doctor review. This setting may have added excess pressure on patients, resulting in lower completion rates. The previous trial cohort were asked to return their questionnaires by a specified date, rather than complete them on a given day.

It is also worth noting the discrepancy between the response rates to the different questionnaires. 45% more patients in our cohort completed the shorter MY-20 questionnaire. The obvious difference between the two is the length of the questionnaire, with the C-30 having ten more questions to answer. The order that the questionnaires were given to the patients may also have been an important factor. Cock *et al* showed the mean time to completion of both questionnaires was 12 minutes, with 83% completing it in less

than 15 minutes[193]**.** In fact, the MY-20 questionnaire piloted in this study contained four more questions than the current iteration.

A limitation of our study is that no data was collected regarding 'time to complete survey' and patient opinion of survey. This significantly limits our ability to explain the lower completion rate. Patient factors for the low completion rate may including stress, difficulty reading/ writing and time pressures. It may also be the case that some patients were not asked to complete the questionnaires at the specified time points. Lastly, there is a significant difference in demographics of our cohort compared to this validation cohort. 94% of the historical cohort were newly diagnosed, compared to just 40% of our cohort.[193] It may therefore be that symptoms related to relapsed disease or side effects of previous therapy (e.g. peripheral neuropathy) may account for our lower questionnaire completion rate.

In a German 'real world study' of the feasibility of performing the C-30 and MY-20 questionnaires, it was reported that 49% of patients required assistance to complete the surveys.[194] Unfortunately, we did not collect this data for our cohort, but this level of assistance would require a significant amount of additional time from the clinical team. We were however able to show that there was no significant difference between the baseline demographics (in terms of age, line of therapy and ECOG performance status) of those patients who completed the MY-20 questionnaire versus those who did not. This is an important finding as it would suggest that the ability or willingness to fill in the survey does not select out or exclude a particular demographic of patient.

At the six-month time point, 50% of the patients in a position to complete the questionnaires completed both surveys, compared to 76% in the Cocks cohort.[193] Our reduced return rate is again likely multi-causal, with similar reasons to the lower completion rates at baseline. It is worth noting that this data collection project had a specially funded research nurse, part of who's role was questionnaire collection. It therefore seems logical to suggest that return rates may be even lower if the responsibility fell on other members of the clinical team, with other pre-existing priorities and duties.

Unfortunately, an insufficient number of QLQ C-30 questionnaires were filled out in our cohort to provide enough data for meaningful analysis. This is however by itself an

interesting observation and may suggest completing both questionnaires in the real-world setting is not feasible.

In summary, questionnaire completion rates in our cohort were lower than historical validation 'field study' cohorts. We did however achieve comparable return rates of the shorter MY-20 questionnaire. The questionnaire was also well filled in, with <1% of data points missing. It may therefore be sensible to prioritise the shorter, myeloma specific MY-20 questionnaire in the real-world clinical setting, where time is a scarce and precious commodity.

A diagnosis of new or relapsed myeloma causes a wide-ranging impact on quality of life: The mean MY-20 scores recorded by our patient group are similar to those reported in the Cocks *et al* field study Table 25).[193] The only major difference is that our ECOG 0 population had a higher average MY-20 score, indicative of a inferior QoL. This is likely multi-causal but may in part be due to inter-clinician variability in use of the EOCG status. It may also be due to underlying demographic differences. The mean ECOG PS 4 score should also be treated with caution, as our cohort had just one patient in this category. Despite this, it does appear that our cohort and the Cocks study cohort are comparable in terms of baseline mean MY-20 scores.

Table 25: Table comparing mean overall MY-20 questionnaire scores by EOCG Performance Score

EOCG Grade	Mean total MY-20 Score (Study cohort)	Mean total MY-20 Score (Cocks *et al* study)
0	30	19
1	36	39
2	40	42
3	46	44
4	52	64

This table compares the mean total MY20 score by ECOG status in our cohort compared to a historic cohort. Our cohort had a lower MY20 score in ECOG 0 group and a lower mean score in the frailest ECOG 4 group

Patients in our data set also had an inferior QoL (higher MY-20 scores) if their Instrumental Activities of Daily Living score (IADL) was low. A low IADL score denotes poorer function, and those patients with a score of 1-5 had a mean MY-20 score of 39 (n=19), compared to an MY-20 score of 34 in those with a higher IADL score of 6-8 (n=53). This was not reported on in the Cocks *et al* study[193] but supports the hypothesis that measures of baseline function are associated with quality of life.

Myeloma patients in our cohort were most troubled by the bone aches and pains (mean score 2.6) and 'thinking about (their) illness' (mean score 2.6). In the Engelhardt German study, the most significant symptoms were reported to be pain and fatigue.[194] This difference may be due to the timings of the surveys. In our cohort, we performed the questionnaire prior to commencing a line of therapy. The German study asked participants to complete the quality-of-life questionnaire on a particular set date, rather than at a specific treatment timepoint. This means that their cohort will be a mixture of treated and partially treated patients.

The baseline scores for each of the scales can also be compared to the German cohort. In order to do this the raw data was transformed in a linear fashion (Appendix 8). At baseline

our cohort had a mean transformed 'Pain symptom' score of 34, compared to 36 in the German cohort.[194] Interestingly our cohort had a lower 'Systemic symptom' score of 19, compared to 26 in the German cohort[194]. This difference is likely accounted for by the fact that our questionnaires were conducted prior to the next cycle of therapy. This is supported by the fact that the mean score for treatment side effect rose to 27 by the six-month time point.

In summary, our cohort had similar baseline characteristics in terms of quality of life as other historical cohorts. While this is not by itself an important clinical finding, it is reassuring to note that our population reports similar issues with quality of life. Aches and pains were the prominent physical symptom identified in our cohort, with 'thinking about (their) illness' being the main psychological concern. This is useful information for the treating clinician to be aware prior to starting a new line of therapy. Pharmacological and holistic interventions can be focused on these specific issues.

ECOG performance status is associated with changes in some aspects of quality of life: As previously mentioned, Cocks *et al* performed a 'field study' to validate the MY-20 questionnaire in the myeloma cohort.[193] The 'construct validity' of the different 'scales' they used was assessed using multi-trait scaling analysis. In practice, this methodology allowed each of the 20 questions to be grouped with the other questions to which they most strongly correlated. This resulted in the production of four 'scales' or domains: disease symptoms (Pain symptoms), side effects of treatment (Systemic symptoms), body image and future perspective (Introduction and Appendix 7). We therefore sought to interrogate the quality-of-life data in our cohort to validate the findings of the original study.

The Cocks paper found a statistically significant difference between Pain symptoms, Systemic symptoms and body image at baseline between the patients with an ECOG performance status of 0, 1 or 2 compared to those with an ECOG performance status of 3 or 4.[193] In our cohort we reproduced these findings, with a statistically significant difference between the two groups for Pain symptoms (p= 0.04), Systemic symptoms (p=0.01) and body image (p=0.02). We also found a non-significant trend to increased concern over

future perspective in the poor performance status group, mirroring the results of the historical cohort.

A limitation of our cohort is the relatively small numbers of individuals in the ECOG 3-4 cohort (n=7). While it is reasonable to analyse our cohort in the same way as previous studies, these small numbers may lead to misleading results. However, we re-analysed our cohort, but comparing ECOG PS group 0 (n=29) vs PS 1 (n=26) Vs PS 2, 3 or 4 (n=17). Reassuringly we again found a statistically significant worsening in MY-20 score on the pain and systemic symptoms scales, adding further evidence that worsening performance status does indeed lead to a deterioration in quality of life.

These results add further evidence that a poor performance status is associated with more significant psychological and physical symptoms. While this may seem like a logical and intuitive finding, it is nonetheless useful to demonstrate. This finding would suggest that limited resources, such as psychological support (and referral to clinical psychology) should be focussed on the population most at need, in this case the poor performance group. The findings may also highlight to clinicians that patients with a poor performance status may require extra holistic assessment of their clinical symptoms and needs. It is also encouraging that a subjective clinician assessment of performance status (using ECOG PS) has a strong association with quality of life.

It is however important to note that only 8% of our cohort had an ECOG performance status of 3 or 4. This suggests they are underrepresented in the cohort, a problem also encountered in the initial assessment.[193]

Beta-2-microglobulin level is associated with changes in some aspects of quality of life: The original validation study was unable to find any statistically significant association with Beta-2-microglobulin (B2M) level and any of the scales.[193] However, in this cohort there was a significant association between having a lower B2M level (<4g/l) and lower Systemic symptom score at baseline (p= 0.002). B2M is a protein that forms part of the HLA molecule and is increased in a variety of malignancies, including multiple myeloma. Higher levels are suggestive of increased disease activity and are of prognostic importance.[195]

The Systemic symptom scale was originally labelled as 'The Treatment Side Effect Scale'. This title appears to be somewhat of a misnomer, as patients reported these symptoms prior to commencing therapy. While some of the questions are specific and common treatment side effects (neuropathy, alopecia), some also relate to non-specific symptoms that may be due to the disease itself (indigestion, feeling ill). It is therefore intuitive that patients with higher disease activity (B2M > 4g/l) will be more symptomatic. The reason for the difference between the cohorts is not clear and further analysis may be required to confirm our association. The difference may in part be due to the different baseline demographics, with many more of our cohort being assessed at relapse, rather than at diagnosis. We hypothesise that patients with higher disease activity and reduced functional reserve (secondary to previous lines of therapy) may be more symptomatic, with worse quality of life scores.

International Myeloma Working Group (IMWG) Classification is associated with changes in some aspects of quality of life: The original 'field validation' study did not seek to interrogate their data for differences in quality of life between the different IMWG classification groups. Indeed, the MY-20 scale pre-dates this patient classification method. This classification was chosen for analysis as the IMWG frailty classification is currently being used as part of the FiTNEss study, a UK nationwide trial for transplant ineligible patients.[196] Interestingly, in our patient cohort we observed a significant difference between the quality of life reported on the 'Systemic symptom Scale' at baseline between the fit, intermediate and frail groups (p= 0.003). It is useful to be aware that the frail cohort are reporting a worse quality of life in one of the scales. If this result were to be corroborated, it may help to inform patient and clinician decision making regarding treatment decisions and resource allocation.

We did not however have sufficient data to assess if there was a significant improvement in quality of life in each of these IMWG frailty groups after six months of therapy. This is perhaps the most important component to explore in future studies. If the quality of life of

the frail group does not improve with treatment, then it is imperative patients are counselled about this prior to committing to a further line of therapy.

Multiple Myeloma International Staging System (ISS) is associated with changes in some aspects of quality of life: There was a statistically significant association between ISS and quality of life on the 'Systemic symptom scale (p=0.001). This was also noted on multiple regression analysis in the Englehardt cohort.[194] The constituents of the ISS are B2M and albumin, which allows classification of patients into prognostic risk groups 1,2 and 3 with Group 3 being the worst prognostic group. It is therefore logical that this poor risk cohort would be burdened with worse symptoms on the Symptom Complex 2 Scale. It is however interesting that out cohort do not have worse symptoms on the Pain symptom scale in the poorer risk ISS groups.

This finding raises a question regarding how the symptom scales were produced and how each of questions was grouped in the initial validation study[193]. Multi-trait scaling analysis was used to group the questions together which are most strongly correlated, and their convergent validity was subsequently assessed. Each of the scales were found to have a correlation coefficient of greater than 0.4, suggesting they are indeed correctly grouped. This will be further discussed in the subsequent PCA section.

There are some clinically important differences in the quality of life at patients commencing different numbers of lines of therapy: We also sought to interrogate our cohort for differences in quality of life between patients who had received different numbers of lines of therapy. This was done using linear transformation of the aggregate scores of each of the scales (Appendix 8). This method has been undertaken in previous studies. [193,194] The 'minimally important difference measurement' has previously been validated for the MY-20 questionnaire to discern significant differences between patient's receiving different numbers of lines of therapy. [197] The minimally important difference is a measure of clinically relevant difference in quality of life, rather than statistical significance. The argument is that these two measures are not necessarily measuring the same thing. Sully *et al* used a triangulation method of anchor and distribution-based analysis, along with

qualitative data to propose their 'minimal important difference' for each of the quality-of-life scales.[197]

One interesting finding was that there was an improvement in the future perspective of patients when starting second line therapy, compared to newly diagnosed patients. This may reflect a familiarity with the diagnosis and less acute distress. Indeed, relapsing patients will have some warning or impending need for a new line of therapy, such as a rising paraprotein of SFLC ratio. This is an important finding as it highlights the importance of psychological support and the utility of a specialist nurse at the time of diagnosis.

Somewhat surprisingly there was no worsening in the Pain symptom scale, the Systemic symptom scale or body image between lines of therapy until the fourth cycle. At the time of starting this cycle, patients had worse quality of life on these three scales. This is extremely clinically useful information and could be used to council patients prior to commencing new lines of therapy. They will be able to give a more informed consent to therapy, knowing that there is likely to be a deterioration in these symptoms by the start of cycle four.

Our findings are however not mirrored by the German MY-20 study, where the most significant deterioration of symptoms was in patients between fourth line and best supportive care[194]. A limitation of our study is the failure to include patients on best supportive care. However, as previously mentioned, a strength of our study is that the questionnaires were performed prior to commencing a treatment cycle, thus giving a step wise picture of change in quality of life. The German cohort questionnaires were completed as a 'snapshot' rather than at a specific timepoint in the patient journey. Thus, their data may be more affected by concurrent treatment side effects of those on active treatment.

In summary, associations between quality of life and a selection of clinically significant parameters reflecting disease activity, frailty and prognosis were found in our cohort. These findings largely mirrored previous studies. Novel findings in our cohort include an association between high B2M at baseline and worse quality of life on the 'Systemic symptom scale', as well as sequential worsening quality of life on the same scale as the level of frailty increased in the IMWG classification. We also found a clinically important deterioration in quality of life for patients commencing fourth line therapy for myeloma therapy. These findings should

be tested in other cohorts but are of potential clinical importance in the medical and holistic management of myeloma patients.

Patients who achieve at least partial remission have an improvement in the psychological aspects of quality of life: For patients who achieved at least a partial remission and in whom we had quality of life data at the baseline and six-month time point, there was no statistically significant change in symptoms reported in the pain or systemic symptom scales (mainly physical symptoms). There was however and improvement in symptoms related to 'future perspectives' at this timepoint.

These findings are different to the Cocks field study, where scores for pain symptoms and body image improved, reflecting an improved quality of life at the follow-up time point. [193] In this cohort Systemic symptom scores (attributes to disease side effects) significantly worsened. The small size of our cohort likely explains why we were unable to find these associations. There were only 22 patients included in this analysis for our cohort. Furthermore, our patient cohort was also more heavily pre-treated, potentially meaning improvement in symptoms may not be the same as in earlier lines of therapy. Indeed, in sub-analysis of our patients receiving first line therapy, there was a statistically significant improvement in the Pain symptoms at the six-month time point.

Aggregate MY-20 Scores are associated with a difference in overall survival in our cohort: Previous work by Wisloff *et al* has reported that the score from several scales in the EORTC CQC-30 Questionnaire can be used to predict survival in their cohort of patients.[198] This is however a historical cohort from the 1990s, so may not be directly comparable.

The aggregate scores for each scale for the patients in our cohort were first analysed for a difference in overall survival. The 'cut-off values' to separate patients in 'Good' and 'Poor' quality of life were created using frequency of distribution histograms. The results were bimodal allowing appropriate scale-specific cut off values to be chosen. There was no statistically significant difference between the 'Good QoL' group and the 'Poor QoL' group for each of the individual scales, but each showed a trend to a shorter overall survival if

quality of life was classified as poor. This is perhaps somewhat surprising given the association between QoL scores and predictive scores such as Multiple Myeloma ISS. It is possible that sample size has limited the significance of the results.

However, there was a statistically significant difference with overall survival and quality of life reported on the total MY-20 score (p=0.04). Using a cut off of 40, patients with a MY-20 score of 40 had a reduced overall survival compared to patients with a better self-reported quality of life. This finding therefore replicates the results Wisloff achieved using the EORTC CQC-30 Questionnaire. [198] It is also of clinical relevance that patients with self-reported good quality of life can expect to have longer overall survivals and will impact on any Quality Adjusted Life Years calculations for future drug approvals.

Principal Component Analysis for MY-20 Questionnaire Answers reveals different clustering of questions when compared to the scales identified in the oritional study: The Cocks *et al* study[193] used multi-trait scaling analysis to decide which scale a question was placed. The logic here was to include questions in the scale to which they had the best correlation. 'Convergent validity' was defined as having a correlation of more than 0.4 between the question and the scale. These scales were then compared to the other scales for 'discriminant validity'. This is to ensure that the scales are measuring different components of quality of life. Scales were therefore based on close 'convergent validity' but a wide 'divergent validity' (greater than two times the standard error of a different scale).

PCA is a non-supervised statistical method for assessing variance and association between sets of data. It uses an algorithm to separate data, based on the factors responsible for most of the variance in the data set, allowing patients with the most in common to be grouped together. It also shows the strength of the impact of each of the variables on the analysis. We used the individual scores of each of the MY-20 questions for each patient and used PCA to demonstrate the questions responsible for most variation in quality of life.

Figure 35 shows the Eigenvector loading values for each of the question on PC2. PC2 was the component selected as it provided the most meaningful separation of data. The

questions at the extremes of the x-axis are the questions that are responsible for most of the variation in the analysis. In our cohort question 13 (tingling in hands and feet) and question 5 (chest pain) are responsible to most of the variation in quality of life.

We then went on to compare the pattern of separation of data by PCA to the original 'scales' in the Cocks *et al* study[193]. There were some interesting differences and similarities between the questions responsible for most covariance in PCA for our cohort and the historical cohort.

Analysis clearly showed a clustering of the 'Future Perspective Scale (orange)' towards the left of the x-axis. This would support the finding that these questions have a similar co-variance, thus as measuring a similar aspect of quality of life, in this case the psychological domain.

There was also a clustering or 'Systemic symptom questions to the right of the x-axis. This again suggests a covariance between these questions (with the exception of question 13, relating to thirst). There is no obvious logic why this question should have different variance to the others in the same scale, but it does appear to be different. The questions in the 'Pain symptoms' (blue) do not however cluster and are dispersed across PC2. This suggests this scale may be measuring a loosely related cluster of symptoms rather than a discrete group of co-linked issues. This may not be of significant clinical importance, but it does mean that these symptoms are less likely to addressed by a specific clinical intervention and may require a more detailed review to address.

Finally, the symptom of chest pain (question 5) was the strongest contributor to variance of PC2 and was separated from the vast majority of questions in Symptom Complex 1 and 2. This is of interest as NT-proBNP has been shown to be an independent prognostic marker in myeloma and has been integrated into novel prognostic scoring algorithms.[199] In this PCA, it was more closely associated with the questions in the 'future perspective scale', relating largely to psychological symptomatology. The term 'chest pain' is not further defined in the questionnaire, so could be interpreted as having a musculoskeletal chest pain or another such variation. When probed there was no statistical difference in the NT-proBNP in those patients reporting chest pain and those that did not. This would have been an important finding as NT-proBNP has been validated as a marker of cardiac amyloid, as well as having

prognostic utility in myeloma.[200] Furthermore, there was no significant survival difference between those reporting chest pain and those not. Patient reported chest pain on a QoL assessment can therefore not be used as a surrogate marker for cardiac amyloid or raised NT-proBNP.

In summary, our PCA of this cohort suggests that the questions related to psychological aspects of the disease do cluster and demonstrate similar covariances. The other scales are less well defined in our cohort when using PCA. This does not challenge the utility of the MY20 tool but does suggest that it is measuring a more disparate collection of symptoms, rather than a focused number of interlinked variables in the aforementioned scales. The utility may therefore lie in picking out particular troublesome symptoms for the individual (that may not have been raised during the consultation with the clinician) rather than defining a specific pattern of symptoms that can be managed with a single intervention.

7.5: Conclusions

First and foremost, the data from the MY-20 questionnaire shows the significant impact the diagnosis and subsequent treatment of myeloma causes, with patients reporting significant issues relating to their quality of life. It therefore seems both relevant and important to collect such data, especially given the fact treatment is non curative. The questionnaire may also provide the treating clinician with a more detailed breakdown of symptoms than was attained in consultation, hence a second opportunity to identify troublesome symptoms.

Analysis of quality of life in this cohort has shown that it is associated with clinical variables, including ECOG performance status, B2M level (surrogate for disease activity) IMWG Frailty classification and ISS score. Demonstrating this association emphasises the conclusion that in order to improve quality of life, all these parameters need to be addressed when treating the patient.

A limitation of this study is sample size, especially at the six month follow up point. Due to the sample size, we were not able to demonstrate significant changes in the quality of life for the individual as treatment progressed. This should be a key criterion for future research.

In summary, the MY-20 questionnaire is feasible and acceptable to patients. It provides insight into a range of variables that affect quality of life and is associated with several important clinical parameters. We feel it should be used as standard of care in the setting of newly diagnosed or relapsed myeloma to help clinicians optimise therapy and make appropriate treatment choices. The MY-20 questionnaire does not however replace clinical and biochemical parameters in the assessment of multiple myeloma but serves as a useful clinical adjunct.

8.1: Summary Conclusion

There are a variety of scoring, risk classification and predictive models to help clinicians with the management of patients with myeloma. Each has its strengths and weaknesses and can help identify patients with good or poor prognosis. The goal of this research was not to devise yet another complicated scoring system, but to help identify health and frailty in the setting of myeloma.

This research has demonstrated that unsupervised analysis in the form of PCA can successfully classify patients into novel clinical groups, reflecting different underlying disease processes. Standard of care tests and pre-existing scores can be used in this way to produce three biology different and clinically meaningful novel cohorts of patients: clinically well, high disease burden and high frailty burden. These novel grouping have significant potential clinical utility to guide management decisions, survival, and prognostication.

The question of whether PCA groupings are superior to the IMWG frailty score is an important one. The IMWG Frailty classification has been adopted as the standard of care for assessment of frailty in the UK. It is relatively simple to use, can be calculated in clinic and has been integrated into the Myeloma XIV trial.[104] When using this system in our cohort, it did not show a difference in overall survival. However, our PCA model does not just consider frailty, it differentiates it from burden of disease. This is often difficult to do at diagnosis and may result to a patient with a high disease burden being confused with a frail individual.

Another advantage of our novel cohorts is the intriguing finding of different patterns of mortality in the high disease burden and high frailty burden groups. Both the high disease and high frailty burden have inferior survival rates to the clinically well group. The early 200 day mortality of the high disease and frailty groups overlaps. After this point (number at risk in each group = 16), the curves diverge, with a plateau in survival of the high disease burden group, while the high frailty burden group demonstrates ongoing mortality. This ability to identify a group with early significant mortality, but with improved survival odds in the longer term (after 200 days) is not a feature of established classification systems such as R-ISS or IMWG Frailty classification. This is therefore a potentially significant clinical finding of our novel cohorts that may be of potential clinical utility.

While it is important to note that the novel PCA groupings provide information regarding the specific biological processes underlying the disease in the three cohorts, it also provides valuable information regarding patients' quality of life. We found a statistically inferior quality of life in the high frailty burden group when compared to the clinically well or high disease burden group. This therefore shows that the PCA grouping provide a multidimensional insight into the disease and its impact on a human, as well as scientific level.

Despite the notable advantages of the novel groups identified by PCA in this settling, this method does pose a significant issue. PCA is a predicative model that does not allow the retrospective inclusion of patients into the data set. Therefore, in practice, if one were to see a new patient in clinic, it is impossible to put this patient into a calculation and assess into which group they will be classified. This means PCA analysis alone fails the 'clinical feasibility' test.

It was therefore important that if the novel PCA groups were to be of clinical use in future research, a method of identifying these cohorts had to be sought to allow prospective classification. We have been able to demonstrate that by using RF analysis (a form of supervised analysis) of protein biomarkers identified my SWATH-MS, along some confirmatory statistics, it is possible to accurately identify and classify patients into our novel PCA groupings. This therefore gives rise to the possibility of prospectively identifying patients into these three distinct clinical groups in future research.

8.2: Future Work

A strength of this research is the amount of detailed demographic and baseline data available for the cohort, enabling a big data approach to statistical analysis. However, the cohort is limited to 91 patient entries, with 49 serum samples undergoing proteomic analysis. It is therefore essential that data be subjected to a future confirmatory study using larger data set, with a consistent study end point, for example five years. This would also allow for assessment of survival and quality of life in each of the three groups in patients undergoing different treatment regimens.

While data was collected prospectively, the analysis and novel PCA groupings were identified retrospectively. The next logical step would be to use the biomarkers identified and to prospectively classify myeloma patients into the three groups and assess for previously identified findings of difference in survival and quality of life. It would also be useful to collect long term follow up data, to allow analysis of progression free survival and therapy related toxicities in each group. This was not possible in our cohort due to limited data.

The identification of a small set of biomarkers that can accurately classify patients into the novel groups is significant. In a prospective future study, these protein biomarkers could be evaluated by an ELISA technique or a related multiplex approach. In this way, the identification process is scalable and more financially feasible than mass spectrometry on serum samples.

Finally, the serum biomarkers we have identified could be used in combination with other baseline demographics such as ECOG performance status or gait speed and be combined into a clinically feasible and meaningful prognostic system. If a biomarker was to be used in this way, it would also be of benefit to further explore the biological plausibility of any candidate biomarker to establish why it should be of clinical importance in the setting of myeloma or frailty. However, as noted in the introduction, novel classification or prognostic scores should be approached with caution. Any future score involving biomarkers must be financially viable and provide more informative data than those that already exist. This must be the benchmark for any future work in this area.

9.0: Appendices

Appendix 1: International Myeloma Working Group Frailty Score (based on Palumbo *et al*; Blood 2015)

International Myeloma Working Group Frailty Score		
Fit	**Intermediate-Fit**	**Frail**
All of: Age ≤ 75 years ADL > 4 IADL > 5 CCI ≤ 1	Age 76-80 years	Age > 80 (regardless of ADL, IADL, CCI)
	Or ADL ≤ 4	Or Age 76-80 and either ADL ≤ 4, IADL ≤ 5, CCI ≥ 2
	Or IADL ≤ 5	Or Age ≤ 75years and at least two of: ADL ≤ 4, IADL ≤ 5, CCI ≥ 2
	Or CCI ≥ 2	

Appendix 2: Eastern Oncology Cooperative Group (ECOG) performance Status Score (based on Oken *et al* [1]

ECOG Score	**Description**
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity, but ambulatory and able to carry out work of a light and sedentary nature.
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours.
4	Completely disabled. Cannot carry out any self-care. Totally confined to bed or chair
5	Dead

Appendix 3

The following tables give a detailed breakdown of the characteristics of interest in the myeloma cohort studies in this research. These data are summarised in the results section and compared to published cohorts.

Age demographics of patient cohort

Age range	Number of patients	Percentage
< 50	4	4 %
50-59	18	20 %
60-69	27	30 %
70-79	29	32 %
> 80	13	14%

Classification of subtypes of multiple myeloma in patient cohort

Subtype of multiple myeloma	Number of patients	Percentage (%)
IgA kappa	7	8
IgA lamda	5	6
IgD lamda	2	2
IgG and IGA lamda	2	2
IgG kappa	35	38
IgG lamda	17	19
IgM	1	1
kappa light chain	12	13
Lamda light chain	9	10
Non secretory	1	1

Number of patients in cohort undergoing different lines of therapy

Line of therapy	Number of patients	Percentage (%)	
1	35	38 %	
2	20	22 %	
3	18	20 %	
4	9	10 %	
5	8	9 %	
6	1	1 %	

ECOG performance Status Scores of cohort

ECOG Performance Score	Number	Percentage (%)
0	33	36%
1	35	39%
2	14	15%
3	7	8%
4	1	1 %

ISS of patients in our cohort

ISS	Number	Percentage (%)
1	19	21 %
2	33	36 %
3	39	43 %

IMWG Frailty Assessment Group in our cohort

IMWG Cohort	Number	Percentage (%)
1	34	38%
2	23	26%
3	33	36%

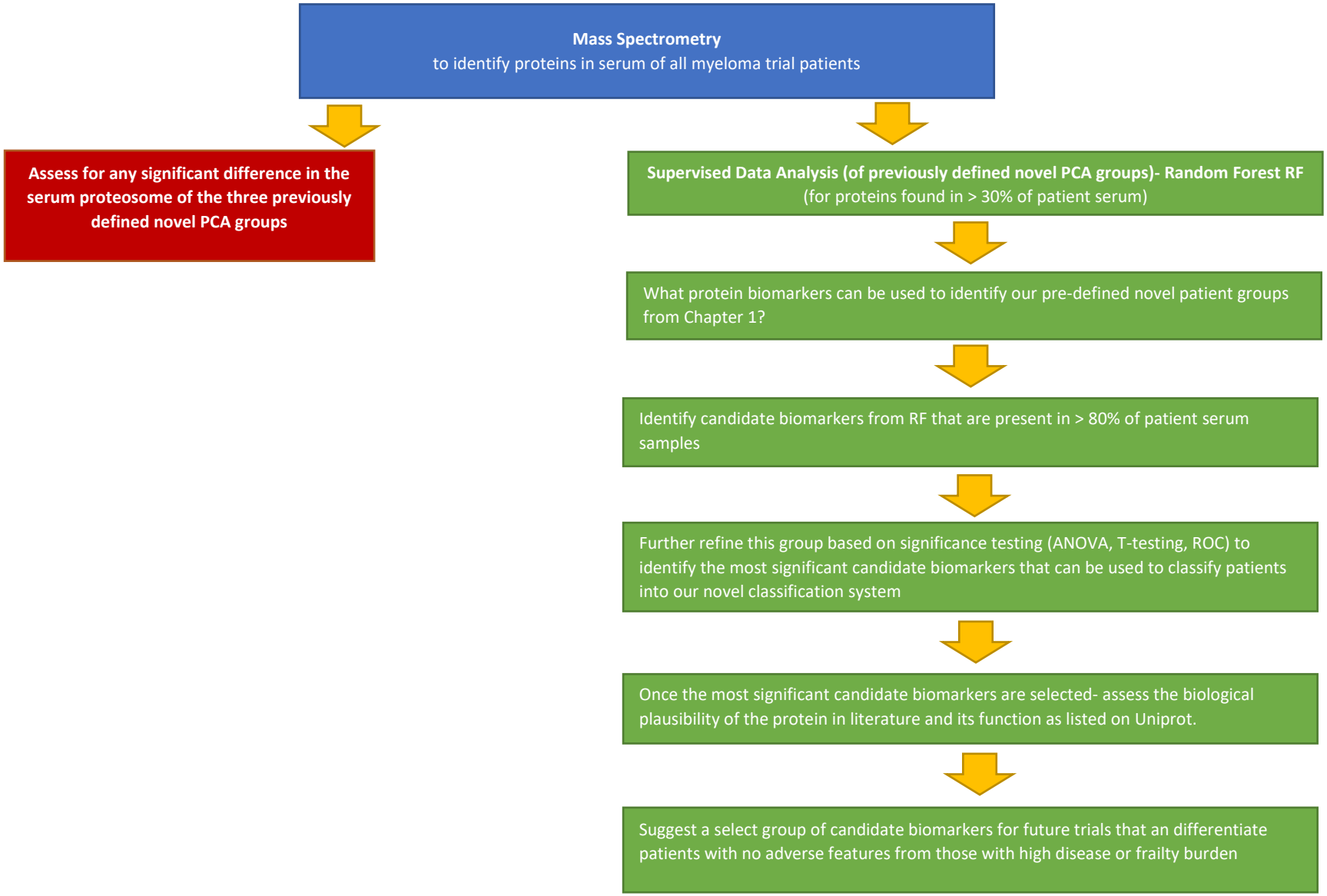
Summary of patient discrete variables at baseline

	Mean	Median	Range
Line of therapy	2	2	1- 6
Number of comorbidities	4	4	0- 13
HCT-CI Score	2	2	0- 9
Number of medications	6	6	0- 20
ISS Stage	2	2	1- 3
ECOG PS	1	1	0- 4
CCI	2	1	0- 6
ADL	6	6	1-6
IADL	7	8	1-8

Summary of patient continuous variables at baseline

	Mean	Median	Range	25% Q	75% Q	SD	SEM
Age	67	67	38-87	60	76	11.1	1.15
Albumin	32	34	14-43	29	37	6.13	0.64
B2M	6.2	5	1.2- 22.2	3.00	7.70	4.69	0.49
LDH	246	210	108- 994	180	265.0	148.4	15.56
Hb	105	104	77- 139	90	118	16.96	1.78
Neutrophils	3.78	3.16	0.07- 16.05	2.24	4.69	2.43	0.26
Platelets	195	202	7- 452	134	250	94.8	9.94
Creatinine	136	94	18- 765	69	147	120.7	12.65
eGFR	60	67	6- 97	34	87	27.64	2.90
Corrected calcium	2.48	2.45	2.17- 3.27	2.35	2.56	0.19	0.02
CRP	18.81	4	1- 225	1	17	37.04	3.88
NT-proBNP	1060	241	7- 25177	104	797	2846	298.4
Paraprotein	17.22	12	0- 86.48	0	28.3	20.24	2.12
Pathological SFLC	1377	582.5	11- 11356	246.9	1413	2054	215.3
D-dimer	2347	920	27- 80000	539.0	1796	8369	877.3
Gait Speed	0.87	0.94	0- 2.13	0.25	1.25	0.59	0.06
Grip strength	25.03	23.30	0- 51.60	18.60	30.6	10.11	1.06

Appendix 4: Flow chart of analytical processes in Chapter 2



Appendix 5

This appendix lists the ten proteins that are best able to discriminate between the novel PCA groups. These proteins can be considered candidate biomarkers to differentiate patients between each of the three cohorts. These proteins were identified using log difference in abundance and a T-test for difference in quantitative levels (significance of $p < 0.05$)

This table showing the ten proteins with the most different quantitative levels between the myeloma with no adverse features group (Group 1) and the myeloma with high disease burden group (Group 2).

logFC	p.Value	Accession	Protein names	Gene
1.58	0.01	P46782	40S ribosomal protein S5 (Small ribosomal subunit protein uS7)	RPS5
1.31	0.03	Q02413	Desmoglein-1 (Cadherin family member 4)	DSG1
-1.30	0.00	P61769	Beta-2-microglobulin	B2M
-1.29	0.00	Q15848	Adiponectin (30 kDa adipocyte complement-related protein)	ADIPOQ
1.22	0.02	Q02880	DNA topoisomerase 2-beta	TOP2B
-1.22	0.00	P00746	Complement factor D	CFD
-1.22	0.00	P61626	Lysozyme C	LYZ
1.21	0.02	Q9P2E9	Ribosome-binding protein 1 (180 kDa ribosome receptor homolog)	RRBP1
-1.20	0.03	P04040	Catalase	CAT
1.16	0.03	Q9Y613	FH1/FH2 domain-containing protein 1 (Formin homolog overexpressed in spleen 1) (FHOS)	FHOD1

Negative log2 fold change means DOWN in Group 1 vs Group 2, positive log2 fold change means UP in Group 1 "Well" vs Group 2 “High"

This table shows the ten proteins with the most different qualitative levels between the myeloma with high disease burden group (Group 2) and myeloma with high frailty burden group (Group 3).

logFC	P.Value	Accession	Protein names	Gene
6.57	0.01	P48147	Prolyl endopeptidase (PE)	PREP
-3.52	0.04	P35527	Keratin, type I cytoskeletal 9 (Cytokeratin-9)	KRT9
-2.79	0.04	P07910	Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1/C2)	HNRNPC
-2.77	0.01	Q15631	Translin	TSN
-2.70	0.01	P48047	ATP synthase subunit O, mitochondrial (ATP synthase peripheral stalk subunit OSCP)	ATP5PO
-2.57	0.01	Q6PI48	Aspartate--tRNA ligase, mitochondrial	DARS2
-2.50	0.05	P08758	Annexin A5 (Anchorin CII)	ANXA5
-2.45	0.03	P02533	Keratin, type I cytoskeletal 14 (Cytokeratin-14) (CK-14)	KRT14
-2.32	0.04	P16083	Ribosyldihydronicotinamide dehydrogenase [quinone]	NQO2
2.26	0.03	Q9Y490	Talin-1	TLN1

Negative log2 fold change means DOWN in Group 2 vs Group 3, positive log2 fold change means UP in Group 2 vs Group 3

This table shows the ten proteins with the most different qualitative levels between the myeloma with no adverse features group (Group 1) and myeloma with high frailty burden group (Group 3).

logFC	P.Value	Accession	Protein names	Gene
-2.10	0.01	Q9P2T1	GMP reductase 2 (GMPR 2)	GMPR2
-1.76	0.02	P07942	Laminin subunit beta-1 (Laminin B1 chain)	LAMB1
1.53	0.04	P60900	Proteasome subunit alpha type-6 (27 kDa prosomal protein) (PROS-27)	PSMA6
-1.34	0.00	P16070	CD44 antigen (CDw44) (Epican) (Extracellular matrix receptor III)	CD44
-1.28	0.00	P37837	Transaldolase	TALDO1
-0.97	0.00	P61626	Lysozyme C	LYZ
-0.97	0.01	P61769	Beta-2-microglobulin	B2M
-0.96	0.04	P80723	Brain acid soluble protein 1 (22 kDa neuronal tissue-enriched acidic protein)	BASP1
-0.95	0.00	P13929	Beta-enolase	ENO3
0.93	0.01	P07451	Carbonic anhydrase 3	CA3

Negative log2 fold change means DOWN in Group 1 vs Group 3, positive log2 fold change means UP in Group 1 vs Group 3.

Appendix 6

This appendix lists the important proteins, identified by RF analysis to distinguished between the groups in pair wise analysis (Group 1 against Group 2, Group 2 against Group 3, Group 1 against Group 3

(A) Important proteins to distinguish between myeloma with no adverse features and myeloma with high disease burden. (B) Important proteins to distinguish between myeloma with high disease burden and myeloma with high frailty burden. (C) Important proteins to distinguish between myeloma with no adverse features and myeloma with high frailty burden

A

Protein names	Gene
Complement factor D	CFD
Cystatin-C (Cystatin-3) (Gamma-trace) (Neuroendocrine basic polypeptide) (Post-gamma-globulin)	CST3
Protein AMBP	AMBP
Apolipoprotein D (Apo-D) (ApoD)	APOD
Gelsolin (AGEL) (Actin-depolymerizing factor) (ADF) (Brevin)	GSN
CD44 antigen (CDw44) (Epican) (Extracellular matrix receptor III)	CD44
Pigment epithelium-derived factor (PEDF) (Cell proliferation-inducing gene 35 protein) (EPC-1) (Serpin F1)	SERPINF1
6-phosphogluconate dehydrogenase, decarboxylating	PGD
Lysozyme C	LYZ
Beta-2-microglobulin	B2M
Tropomyosin alpha-4 chain (TM30p1) (Tropomyosin-4)	TPM4
EGF-containing fibulin-like extracellular matrix protein 1 (Extracellular protein S1-5) (Fibrillin-like protein)	EFEMP1
Adiponectin (30 kDa adipocyte complement-related protein)	ADIPOQ
Lymphatic vessel endothelial hyaluronic acid receptor 1 (LYVE-1) (Cell surface retention sequence-binding protein 1)	LYVE1
Insulin-like growth factor-binding protein 6 (IBP-6) (IGF-binding protein 6) (IGFBP-6)	IGFBP6
Endothelial protein C receptor (Activated protein C receptor) (APC receptor) (Endothelial cell protein C receptor) (CD antigen CD201)	PROCR

Adipocyte plasma membrane-associated protein (Protein BSCv)	APMAP
Leucine-rich alpha-2-glycoprotein (LRG)	LRG1
Succinate--CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	SUCLG1
Inter-alpha-trypsin inhibitor heavy chain H3 (ITI heavy chain H3)	ITIH3
Histone H1.0 (Histone H1') (Histone H1(0)) [Cleaved into: Histone H1.0, N-terminally processed]	H1-0
Complement component C7	C7
Propionyl-CoA carboxylase alpha chain, mitochondrial (PCCase subunit alpha)	PCCA
Hepatocyte growth factor-like protein (Macrophage stimulatory protein)	MST1
Importin subunit beta-1 (Importin-90) (Karyopherin subunit beta-1)	KPNB1
Ectonucleoside triphosphate diphosphohydrolase 5 (NTPDase 5)	ENTPD5
Toll-interacting protein	TOLLIP
Dynamin-1-like protein	DNM1L
Apolipoprotein E (Apo-E)	APOE
Angiogenin	ANG
Brain acid soluble protein 1 (22 kDa neuronal tissue-enriched acidic protein)	BASP1
Ficolin-2 (37 kDa elastin-binding protein)	FCN2
Ferritin light chain (Ferritin L subunit)	FTL
Apolipoprotein A-IV (Apo-AIV) (ApoA-IV) (Apolipoprotein A4)	APOA4
Ribonuclease pancreatic	RNASE1

B

Protein names	**Gene**
Prothrombin	F2
Apolipoprotein E (Apo-E)	APOE
Apolipoprotein D (Apo-D) (ApoD)	APOD
6-phosphogluconate dehydrogenase, decarboxylating	PGD
Aspartate--tRNA ligase, mitochondrial	DARS2
Ribosome-binding protein 1 (180 kDa ribosome receptor homolog) (RRp) (ES/130-related protein) (Ribosome receptor protein)	RRBP1

Fumarate hydratase, mitochondrial (Fumarase) (HsFH)	FH
Interleukin enhancer-binding factor 3 (Double-stranded RNA-binding protein 76)	ILF3
Gelsolin (AGEL) (Actin-depolymerizing factor) (ADF) (Brevin)	GSN
Kinesin-1 heavy chain (Conventional kinesin heavy chain) (Ubiquitous kinesin heavy chain) (UKHC)	KIF5B
Nuclear pore complex protein Nup50 (50 kDa nucleoporin) (Nuclear pore-associated protein 60 kDa-like) (Nucleoporin Nup50)	NUP50
Lymphatic vessel endothelial hyaluronic acid receptor 1 (LYVE-1) (Cell surface retention sequence-binding protein 1)	LYVE1
Pigment epithelium-derived factor (PEDF) (Cell proliferation-inducing gene 35 protein) (EPC-1) (Serpin F1)	SERPINF1
Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1/C2)	HNRNPC
Neutrophil gelatinase-associated lipocalin (NGAL) (25 kDa alpha-2-microglobulin-related subunit of MMP-9) (Lipocalin-2)	LCN2
60S ribosomal protein L10a (CSA-19) (Large ribosomal subunit protein uL1)	RPL10A
Ubiquitin-like modifier-activating enzyme 1	UBA1
Glyoxalase domain-containing protein 4	GLOD4
Calpain small subunit 1 (CSS1) (Calcium-activated neutral proteinase small subunit)	CAPNS1
Inter-alpha-trypsin inhibitor heavy chain H1 (ITI heavy chain H1)	ITIH1
Protein AMBP	AMBP
Propionyl-CoA carboxylase alpha chain, mitochondrial (PCCase subunit alpha)	PCCA
Cellular nucleic acid-binding protein (CNBP) (Zinc finger protein 9)	CNBP
Attractin (DPPT-L) (Mahogany homolog)	ATRN
Cell cycle and apoptosis regulator protein 2 (Cell division cycle and apoptosis regulator protein 2)	CCAR2
Proteasome subunit beta type-9	PSMB9
Complement factor H (H factor 1)	CFH
Syntaxin-binding protein 2 (Protein unc-18 homolog 2) (Unc18-2) (Protein unc-18 homolog B) (Unc-18B)	STXBP2
Far upstream element-binding protein 2 (FUSE-binding protein 2) (KH type-splicing regulatory protein) (KSRP) (p75)	KHSRP

Proteasome subunit alpha type-6 (27 kDa prosomal protein) (PROS-27) (p27K)	PSMA6
Epididymis-specific alpha-mannosidase	MAN2B2
Centromere protein F (CENP-F) (AH antigen) (Kinetochore protein CENPF) (Mitosin)	CENPF
von Willebrand factor (vWF) [Cleaved into: von Willebrand antigen 2 (von Willebrand antigen II)]	VWF
CD44 antigen (CDw44) (Epican) (Extracellular matrix receptor III)	CD44
Cadherin-5 (7B4 antigen) (Vascular endothelial cadherin) (VE-cadherin) (CD antigen CD144)	CDH5

C

Protein names	**Gene**
Ectonucleoside triphosphate diphosphohydrolase 5 (NTPDase 5)	ENTPD5
CD44 antigen (CDw44) (Epican) (Extracellular matrix receptor III)	CD44
Transaldolase	TALDO1
Lysozyme C	LYZ
Brain acid soluble protein 1 (22 kDa neuronal tissue-enriched acidic protein) (Neuronal axonal membrane protein NAP-22)	BASP1
EGF-containing fibulin-like extracellular matrix protein 1 (Extracellular protein S1-5) (Fibrillin-like protein) (Fibulin-3) (FIBL-3)	EFEMP1
Rab GDP dissociation inhibitor alpha (Rab GDI alpha)	GDI1
Pigment epithelium-derived factor (PEDF) (Cell proliferation-inducing gene 35 protein) (EPC-1) (Serpin F1)	SERPINF1
Four and a half LIM domains protein 1 (FHL-1) (Skeletal muscle LIM-protein 1) (SLIM) (SLIM-1)	FHL1
Importin subunit beta-1 (Importin-90) (Karyopherin subunit beta-1) (Nuclear factor p97)	KPNB1
Endothelial protein C receptor (Activated protein C receptor) (APC receptor)	PROCR
Beta-2-microglobulin	B2M
Apolipoprotein F (Apo-F) (Lipid transfer inhibitor protein) (LTIP)	APOF
Complement factor D	CFD

Kinesin-1 heavy chain (Conventional kinesin heavy chain) (Ubiquitous kinesin heavy chain) (UKHC)	KIF5B
Nuclear pore complex protein Nup50 (50 kDa nucleoporin) (Nuclear pore-associated protein 60 kDa-like) (Nucleoporin Nup50)	NUP50
Cystatin-C (Cystatin-3) (Gamma-trace) (Neuroendocrine basic polypeptide) (Post-gamma-globulin)	CST3
Carbonic anhydrase 1	CA1
Tropomyosin alpha-4 chain (TM30p1) (Tropomyosin-4)	TPM4
Apolipoprotein C-I (Apo-CI) (ApoC-I) (Apolipoprotein C1)	APOC1
Galactokinase	GALK1
Transitional endoplasmic reticulum ATPase (TER ATPase)	VCP
Decorin (Bone proteoglycan II) (PG-S2) (PG40)	DCN
Beta-enolase	ENO3
Complement factor I	CFI
Vitronectin (VN) (S-protein) (Serum-spreading factor) (V75)	VTN
5'-nucleotidase domain-containing protein 1	NT5DC1
Thyroxine-binding globulin (Serpin A7) (T4-binding globulin)	SERPINA7
Metalloproteinase inhibitor 1 (Erythroid-potentiating activity) (EPA) (Fibroblast collagenase inhibitor)	TIMP1
Twinfilin-2 (A6-related protein) (hA6RP) (Protein tyrosine kinase 9-like) (Twinfilin-1-like protein)	TWF2
60S ribosomal protein L3 (HIV-1 TAR RNA-binding protein B) (TARBP-B) (Large ribosomal subunit protein uL3)	RPL3
Inter-alpha-trypsin inhibitor heavy chain H3 (ITI heavy chain H3) (ITI-HC3) (Inter-alpha-inhibitor heavy chain 3)	ITIH3
Chloride intracellular channel protein 1 (Chloride channel ABP) (Nuclear chloride ion channel 27)	CLIC1
Cellular nucleic acid-binding protein (CNBP) (Zinc finger protein 9)	CNBP

Appendix 7: EORTC MY-20 Questionnaire Questions and Scales

Scale	Question	
Disease symptoms (Pain symptoms)	Have you had bone aches/ pains?	1
	Have you had pain in your back?	2
	Have you had pain in your hip?	3
	Have you had pain in your arms or shoulder?	4
	Have you had pain in your chest?	5
	If you had pain, did it increase with activity?	6
Side effects of treatment (Systemic symptoms)	Did you feel drowsy?	7
	Did you feel thirsty?	8
	Did you feel ill?	9
	Have you had a dry mouth?	10
	Have you lost any hair?	11
	Did you have tingling in your hands or feet?	13
	Did you feel restless or agitated?	14
	Have you had indigestion or heartburn?	15
	Have you had burning or sore eyes?	16
Body Image	Have you felt physically less attractive as a result of your disease or treatment?	17
Future perspective	Have you been thinking about your illness?	18
	Have you been worried about dying?	19
	Have you worried about your health in the future?	20

Appendix 8

Formula used for linear transformation of 'Scale means'

Based on calculations from:

https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf (accessed 21/10/21)

Functional Scales (Body image and future perspective)

$$\text{Score} = \left(1 - \frac{(RS-1)}{\text{Range}}\right) \times 100$$

Symptom scales (Disease symptoms and side effect scales)

$$\text{Score} = ((RS-1)/\text{range}) \times 100$$

Where

Raw: score (RS) = $(I_1 + I_2 + I_n)/n$

n= number of questions in scale

range = 3 (based on scale from 1-4)

10.0: References

[1] Handforth C, Clegg A, Young C, Simpkins S, Seymour MT, Selby PJ, *et al*. The prevalence and outcomes of frailty in older cancer patients: a systematic review. Ann Oncol. 2015;26(6):1091-101.

[2] Nielson CM, Bylsma LC, Fryzek JP, Saad HA, Crawford J. Relative Dose Intensity of Chemotherapy and Survival in Patients with Advanced Stage Solid Tumor Cancer: A Systematic Review and Meta-Analysis. The Oncologist. 2021;26(9):e1609-e18.

[3] Diamond E, Lahoud OB, Landau H. Managing multiple myeloma in elderly patients. Leuk Lymphoma. 2018;59(6):1300-11.

[4] Seesaghur A, Petruski-Ivleva N, Banks VL, Wang JR, Abbasi A, Neasham D, *et al*. Clinical features and diagnosis of multiple myeloma: a population-based cohort study in primary care. BMJ Open. 2021;11(10):e052759.

[5] Xue Q-L. The frailty syndrome: definition and natural history. Clinics in geriatric medicine. 2011;27(1):1-15.

[6] Martin FC, O'Halloran AM. Tools for Assessing Frailty in Older People: General Concepts. Adv Exp Med Biol. 2020;1216:9-19.

[7] Rodriguez-Mañas L, Fried LP. Frailty in the clinical scenario. Lancet, The. 2015;385(9968):e7-e9

[8] Turner G, Clegg A. Best practice guidelines for the management of frailty: a British Geriatrics Society, Age UK and Royal College of General Practitioners report. Age and Ageing. 2014;43(6):744-7.

[9] Gill TM, Gahbauer EA, Allore HG, Han L. Transitions between frailty states among community-living older persons. Arch Intern Med. 2006;166(4):418-23.

[10] Diebel LWM, Rockwood K. Determination of Biological Age: Geriatric Assessment vs Biological Biomarkers. Curr Oncol Rep. 2021;23(9):104.

[11] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, *et al*. Frailty in Older Adults: Evidence for a Phenotype. The Journals of Gerontology: Series A. 2001;56(3):M146-M57.

[12] Cesari M, Gambassi G, van Kan GA, Vellas B. The frailty phenotype and the frailty index: different instruments for different purposes. Age Ageing. 2014;43(1):10-2.

[13] Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. Age Ageing. 2018;47(2):193-200

[14] NICE Guidelines for Management of Multiple Myeloma. Accessed online 10/9/22: https://www.nice.org.uk/guidance/TA763/chapter/1-Recommendations

[15] Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. TheScientificWorldJournal. 2001;1:323-36.

[16] Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. J Gerontol A Biol Sci Med Sci. 2007;62(7):738-43.

[17] Lee H, Lee E, Jang IY. Frailty and Comprehensive Geriatric Assessment. J Korean Med Sci. 2020;35(3):e16.

[18] Ethun CG, Bilen MA, Jani AB, Maithel SK, Ogan K, Master VA. Frailty and cancer: Implications for oncology surgery, medical oncology, and radiation oncology. CA Cancer J Clin. 2017;67(5):362-77.

[19] Mohile SG, Dale W, Somerfield MR, Schonberg MA, Boyd CM, Burhenn PS, *et al*. Practical Assessment and Management of Vulnerabilities in Older Patients Receiving Chemotherapy: ASCO Guideline for Geriatric Oncology. J Clin Oncol. 2018;36(22):2326-47.

[20] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017;67(1):7-30.

[21] Mohile SG, Xian Y, Dale W, Fisher SG, Rodin M, Morrow GR, *et al*. Association of a cancer diagnosis with vulnerability and frailty in older Medicare beneficiaries. J Natl Cancer Inst. 2009;101(17):1206-15.

[22] Cancer Research UK website. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/incidence#heading-One Accessed 5/11/22

[23] Loecker C, Schmaderer M, Zimmerman L. Frailty in Young and Middle-Aged Adults: An Integrative Review. J Frailty Aging. 2021;10(4):327-33

[24] Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95.

[25] García-Giménez JL, Mena-Molla S, Tarazona-Santabalbina FJ, Viña J, Gomez-Cabrera MC, Pallardó FV. Implementing Precision Medicine in Human Frailty through Epigenetic Biomarkers. International journal of environmental research and public health [Internet]. 2021 2021/02//; 18(4). Available from: http://europepmc.org/abstract/MED/33672064

[26] Ferrucci L, Cavazzini C, Corsi A, Bartali B, Russo CR, Lauretani F, *et al*. Biomarkers of frailty in older persons. J Endocrinol Invest. 2002;25(10 Suppl):10-5.

[27] Perazza LR, Brown-Borg HM, Thompson LV. Physiological Systems in Promoting Frailty. Compr Physiol. 2022;12(3):3575-620.

[28] Cardoso AL, Fernandes A, Aguilar-Pimentel JA, de Angelis MH, Guedes JR, Brito MA, *et al*. Towards frailty biomarkers: Candidates from genes and pathways regulated in aging and age-related diseases. Ageing Research Reviews. 2018;47:214-77

[29] Heider M, Nickel K, Högner M, Bassermann F. Multiple Myeloma: Molecular Pathogenesis and Disease Evolution. Oncology Research and Treatment. 2021;44(12):672-81.

[30]Barwick BG, Gupta VA, Vertino PM, Boise LH. Cell of Origin and Genetic Alterations in the Pathogenesis of Multiple Myeloma. Frontiers in Immunology. 2019;10.

[31] von Schwedler U, Jck H-M, Wabl M. Circular DNA is a product of the immunoglobulin class switch rearrangement. Nature. 1990;345(6274):452-6.

[32] Bergsagel PL, Kuehl WM. Molecular Pathogenesis and a Consequent Classification of Multiple Myeloma. Journal of Clinical Oncology. 2005;23(26):6333-8.

[33] Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, *et al*. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol. 2014;15(12):e538-48.

[34] van de Donk NWCJ, Pawlyn C, Yong KL. Multiple myeloma. The Lancet. 2021;397(10272):410-27.

[35] Cancer Research UK (2020); Myeloma Statistics. Accessed online on 14/6/23: http://cancerresearchuk.org

[36] Rajan AM, Rajkumar SV. Interpretation of cytogenetic results in multiple myeloma for clinical practice. Blood Cancer Journal. 2015;5(10):e365-e.

[37] Abdallah N, Rajkumar SV, Greipp P, Kapoor P, Gertz MA, Dispenzieri A, et al. Cytogenetic abnormalities in multiple myeloma: association with disease characteristics and treatment response. Blood Cancer Journal. 2020;10(8):82.

[38] Cowan AJ, Allen C, Barac A, Basaleem H, Bensenor I, Curado MP, *et al*. Global Burden of Multiple Myeloma: A Systematic Analysis for the Global Burden of Disease Study 2016. JAMA Oncol. 2018;4(9):1221-7.

[39] Palumbo A, Bringhen S, Ludwig H, Dimopoulos MA, Bladé J, Mateos MV, *et al*. Personalized therapy in multiple myeloma according to patient age and vulnerability: a report of the European Myeloma Network (EMN). Blood. 2011;118(17):4519-29.

[40] Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, Bladé J, *et al*. International staging system for multiple myeloma. J Clin Oncol. 2005;23(15):3412-20.

[41] Cassuto JP, Krebs BP, Viot G, Dujardin P, Masseyeff R. beta 2 Microglobulin, a tumour marker of lymphoproliferative disorders. Lancet. 2. England1978. p. 108-9.

[42] Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, *et al*. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of Clinical Oncology. 2015;33(26):2863-9.

[43] Claps G, Faouzi S, Quidville V, Chehade F, Shen S, Vagner S, et al. The multiple roles of LDH in cancer. Nature Reviews Clinical Oncology. 2022;19(12):749-62.

[44] Pimentel M, Espinal O, Godinez F, Jimenez F, Martinez D, Mendoza N, *et al*. Consensus Statement: Importance of Timely Access to Multiple Myeloma Diagnosis and Treatment in Central America and the Caribbean. J Hematol. 2022;11(1):1-7.

[45] Ludwig H, Durie BGM, Bolejack V, Turesson I, Kyle RA, Blade J, *et al*. Myeloma in patients younger than age 50 years presents with more favorable features and shows better survival: an analysis of 10 549 patients from the International Myeloma Working Group. Blood. 2008;111(8):4039-47.

[46] Nilsson T, Höglund M, Lenhoff S, Rylander L, Turesson I, Westin J, *et al*. A pooled analysis of karyotypic patterns, breakpoints and imbalances in 783 cytogenetically abnormal multiple myelomas reveals frequently involved chromosome segments as well

as significant age- and sex-related differences. British Journal of Haematology. 2003;120(6):960-9.

[47] Avet-Loiseau H, Hulin C, Campion L, Rodon P, Marit G, Attal M, *et al*. Chromosomal abnormalities are major prognostic factors in elderly patients with multiple myeloma: the intergroupe francophone du myélome experience. J Clin Oncol. 2013;31(22):2806-9.

[48] Network Site Specific Group-haematology; University of Oxford Algorithm: Accessed online 14/6/23: https://nssg.oxford-haematology.org.uk/myeloma/op-policy-and-guidelines/MM-29-myeloma-algorithm.pdf

[49] Teoh PJ, Chng WJ. CAR T-cell therapy in multiple myeloma: more room for improvement. Blood Cancer Journal. 2021;11(4):84.

[50] Moreau P, Garfall AL, van de Donk NWCJ, Nahi H, San-Miguel JF, Oriol A, et al. Teclistamab in Relapsed or Refractory Multiple Myeloma. New England Journal of Medicine. 2022;387(6):495-505.
[51] Goldsmith SR, Vij R. Evolving Paradigms of Therapy for Multiple Myeloma: State of the Art and Future Directions. JCO Oncology Practice. 2021;17(7):415-8.

[52] Pozzi S, Marcheselli L, Bari A, Liardo EV, Marcheselli R, Luminari S, *et al*. Survival of multiple myeloma patients in the era of novel therapies confirms the improvement in patients younger than 75 years: a population-based analysis. British Journal of Haematology. 2013;163(1):40-6.

[53] Kint N, Delforge M. Concise review - Treatment of multiple myeloma in the very elderly: How do novel agents fit in? Journal of Geriatric Oncology. 2016;7(5):383-9.

[54] Habicht DW, Witham MD, McMurdo MET. The under-representation of older people in clinical trials: Barriers and potential solutions. The Journal of Nutrition Health and Aging. 2008;12(3):194-6.

[55] Clinicaltrials.gov Website: Accessed on 11/09/2022: https://clinicaltrials.gov/ct2/show/record/NCT03720041

[56] Kleber M, Ihorst G, Groß B, Koch B, Reinhardt H, Wäsch R, *et al*. Validation of the Freiburg Comorbidity Index in 466 Multiple Myeloma Patients and Combination With the International Staging System Are Highly Predictive for Outcome. Clinical Lymphoma Myeloma and Leukemia. 2013;13(5):541-51.

[57] Kim SM, Kim, M.J., Jung, C.W. Comparison of the freiburg and charlson comorbidity indices in predicting overall survival in elderly patients with newly diagnosed multiple myeloma. BioMed Research International: Hindawi Publishing Corporation; 2014.

[58] Sara B, Maria Victoria M, Sonja Z, Alessandra L, Antonietta Pia F, Albert O, *et al*. Age and organ damage correlate with poor survival in myeloma patients: meta-analysis of 1435 individual patient data from 4 randomized trials. Haematologica. 2013;98(6):980-7

[59] Engelhardt M, Domm AS, Dold SM, Ihorst G, Reinhardt H, Zober A, *et al*. A concise revised Myeloma Comorbidity Index as a valid prognostic instrument in a large cohort of 801 multiple myeloma patients. Haematologica. 2017;102(5):910-21

[60] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. JAMA. 1963;185(12):914-9.

[61] Lawton MP, Brody EM. ASSESSMENT OF OLDER PEOPLE: SELF-MAINTAINING AND INSTRUMENTAL ACTIVITIES OF DAILY LIVING. Nursing Research. 1970;19(3).

[62] Rubenstein LV, Calkins DR, Greenfield S, Jette AM, Meenan RF, Nevins MA, *et al*. Health status assessment for elderly patients. Report of the Society of General Internal Medicine Task Force on Health Assessment. J Am Geriatr Soc. 1989;37(6):562-9.

[63] Mor V, Laliberte L, Morris JN, Wiemann M. The Karnofsky performance status scale: An examination of its reliability and validity in a research setting. Cancer. 1984;53(9):2002-7.

[64] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, *et al*. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982;5(6):649-55.

[65] Takamatsu H, Honda S, Miyamoto T, Yokoyama K, Hagiwara S, Ito T, *et al*. Changing trends in prognostic factors for patients with multiple myeloma after autologous stem cell transplantation during the immunomodulator drug/proteasome inhibitor era. Cancer Science. 2015;106(2):179-85.

[66] Wildes TM, Tuchman SA, Klepin HD, Mikhael J, Trinkaus K, Stockerl-Goldstein K, *et al*. Geriatric Assessment in Older Adults with Multiple Myeloma. J Am Geriatr Soc. 2019;67(5):987-91.

[67] Hamaker ME, Jonker JM, de Rooij SE, Vos AG, Smorenburg CH, van Munster BC. Frailty screening methods for predicting outcome of a comprehensive geriatric assessment in elderly patients with cancer: a systematic review. Lancet Oncol. 2012;13(10):e437-44

[68] Cook G, Larocca A, Facon T, Zweegman S, Engelhardt M. Defining the vulnerable patient with myeloma—a frailty position paper of the European Myeloma Network. Leukemia. 2020;34(9):2285-94.

[69] Offidani M, Corvatta L, Polloni C, Centurioni R, Visani G, Brunori M, *et al*. Assessment of Vulnerability Measures and Their Effect on Survival in a Real-Life Population of Multiple Myeloma Patients Registered at Marche Region Multiple Myeloma Registry. Clinical Lymphoma, Myeloma & Leukemia. 2012;12(6):423-32.

[70] Palumbo A, Bringhen S, Mateos M-V, Larocca A, Facon T, Kumar SK, *et al*. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. Blood. 2015;125(13):2068-74.

[71] Cook G, Royle K-L, Pawlyn C, Hockaday A, Shah V, Kaiser MF, *et al*. A clinical prediction model for outcome and therapy delivery in transplant-ineligible patients with myeloma (UK Myeloma Research Alliance Risk Profile): a development and validation study. The Lancet Haematology. 2019;6(3):e154-e66.

[72] Redder L, Klausen TW, Vangsted AJ, Gregersen H, Andersen NF, Pedersen RS, *et al*. Validation of a New Clinical Prediction Model for Outcome in Newly Diagnosed Multiple Myeloma Patients Not Eligible for Autologous Stem-Cell Transplantation; A Population-Based Study from the Danish National Multiple Myeloma Registry. Blood. 2019;134(Supplement_1):1849-.

[73] Razjouyan J, Naik AD, Horstman MJ, Kunik ME, Amirmazaheri M, Zhou H, *et al*. Wearable Sensors and the Assessment of Frailty among Vulnerable Older Adults: An Observational Cohort Study. Sensors (Basel). 2018;18(5)

[74] Favaretto M, De Clercq E, Schneble CO, Elger BS. What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS One. 2020;15(2):e0228987.

[75] Wang W, Krishnan E. Big Data and Clinicians: A Review on the State of the Science. JMIR Med Inform. 2014;2(1):e1.

[76] Semenick D. TESTS AND MEASUREMENTS: The T-test. Strength & Conditioning Journal. 1990;12(1):36-7.

[77] Menyhart O, Weltz B, Győrffy B. MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction. PLoS One. 2021;16(6):e0245824.

[78] Principal Component Analysis for Special Types of Data. Principal Component Analysis. New York, NY: Springer New York; 2002. p. 338-72.

[79] Ringnér M. What is principal component analysis? Nature Biotechnology. 2008;26(3):303-4.

[80] Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She Q-B, Maurer M, *et al*. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. Proceedings of the National Academy of Sciences. 2007;104(18):7564-9.

[81] Rachel Wolff: Monkey Learn; 5 Types of Classification in Machine Learning (Internet) Aug 26th 2020: Available at: https://monkeylearn.com/blog/classification-algorithms/ Accessed online 18/3/23

[82] Stayam Kumar: Towards Data Science; 3 Techniquest to avoid overfitting of decision trees (Internet); 31st May 2021: Website: https://towardsdatascience.com/3-techniques-to-avoid-overfitting-of-decision-trees-1e7d3d985a09#:~:text=Is%20your%20Decision%20Tree%20Overfitting,fails%20to%20capture%20important%20patterns. Accessed online: 18/3/23

[83] Boulesteix A-L, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. WIREs Data Mining and Knowledge Discovery. 2012;2(6):493-507

[84] Nimrod G, Szilágyi A, Leslie C, Ben-Tal N. Identification of DNA-binding Proteins Using Structural, Electrostatic and Evolutionary Features. Journal of Molecular Biology. 2009;387(4):1040-53.

[85] Goldstein BA, Polley Ec Fau - Briggs FBS, Briggs FB. Random forests for genetic association studies. (1544-6115 (Electronic)).

[86] Weiwei H. Classification of sport actions using principal component analysis and random forest based on three-dimensional data. Displays. 2022;72:102135.

[87] Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, Bladé J, *et al*. International staging system for multiple myeloma. J Clin Oncol. 2005;23(15):3412-20

[88] Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, *et al*. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of Clinical Oncology. 2015;33(26):2863-9.

[89] Palumbo A, Bringhen S, Mateos M-V, Larocca A, Facon T, Kumar SK, *et al*. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. Blood. 2015;125(13):2068-74.

[90] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, *et al*. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982;5(6):649-55.

[91] Sorror ML, Maris MB, Storb R, Baron F, Sandmaier BM, Maloney DG, *et al*. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. Blood. 2005;106(8):2912-9.

[92] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83.

[93] Moreau P, Hulin C, Perrot A, Arnulf B, Belhadj K, Benboubker L, *et al*. Maintenance with daratumumab or observation following treatment with bortezomib, thalidomide, and dexamethasone with or without daratumumab and autologous stem-cell transplant in patients with newly diagnosed multiple myeloma (CASSIOPEIA): an open-label, randomised, phase 3 trial. The Lancet Oncology. 2021;22(10):1378-90.

[94] Banna GL, Cantale O, Haydock MM, Battisti NML, Bambury K, Musolino N, *et al*. International Survey on Frailty Assessment in Patients with Cancer. The Oncologist. 2022

[95] Rocci A BJ. The Role of Frailty in Patients with Multiple Myeloma. Manchester Royal Infirmary Foundation Trust: Myeloma UK; 2017.

[96] National institute of cancer Guidelines. Accessed online: 9/11/21 https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm#ctc_50

[97] Sive J, Cuthill K, Hunter H, Kazmi M, Pratt G, Smith D, *et al*. Guidelines on the diagnosis, investigation and initial treatment of myeloma: a British Society for Haematology/UK Myeloma Forum Guideline. British Journal of Haematology. 2021;193(2):245-68.

[98] Dray S, Josse J. Principal component analysis with missing values: a comparative survey of methods. Plant Ecology. 2015;216(5):657-67.

[99] Palumbo A, Bringhen S, Mateos M-V, Larocca A, Facon T, Kumar SK, *et al*. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. Blood. 2015;125(13):2068-74

[100] Engelhardt M, Sandra Maria D, Gabriele I, Alexander Z, Mandy M, Heike R, *et al*. Geriatric assessment in multiple myeloma patients: validation of the International Myeloma Working Group (IMWG) score and comparison with other common comorbidity scores. Haematologica. 2016;101(9):1110-9.

[101] Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, Bladé J, *et al*. International staging system for multiple myeloma. J Clin Oncol. 2005;23(15):3412-20.

[102] Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, *et al*. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of Clinical Oncology. 2015;33(26):2863-9.

[103] Avet-Loiseau H, Hulin C, Campion L, Rodon P, Marit G, Attal M, *et al*. Chromosomal abnormalities are major prognostic factors in elderly patients with multiple myeloma: the

intergroupe francophone du myélome experience. J Clin Oncol. 2013;31(22):2806-9.

[104] Clinicaltrials.gov Website: Accessed on 18/11/21: https://clinicaltrials.gov/ct2/show/record/NCT03720041

[105] Afram G, Gran C, Borg Bruchfeld J, Wagner AK, Hussain A, Alici E, *et al*. Impact of performance status on overall survival in patients with relapsed and/or refractory multiple myeloma: Real-life outcomes of daratumumab treatment. Eur J Haematol. 2020;105(2):196-202.

[106] Blagden SP, Charman SC, Sharples LD, Magee LRA, Gilligan D. Performance status score: do patients and their oncologists agree? British Journal of Cancer. 2003;89(6):1022-7.

[107] Avilés A, Zepeda G, Guzmán R, Talavera A, García EL, Díaz-Maqueo JC. [Prognostic importance of beta-2-microglobulin in multiple myeloma]. Rev Invest Clin. 1992;44(2):215-20.

[108] Engelhardt M, Domm AS, Dold SM, Ihorst G, Reinhardt H, Zober A, *et al*. A concise revised Myeloma Comorbidity Index as a valid prognostic instrument in a large cohort of 801 multiple myeloma patients. Haematologica. 2017;102(5):910-21.

[109] Kumar S, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, Colby C, *et al*. Revised Prognostic Staging System for Light Chain Amyloidosis Incorporating Cardiac Biomarkers and Serum Free Light Chain Measurements. Journal of Clinical Oncology. 2012;30(9):989-95.

[110] Richards AM, Crozier IG, Yandle TG, Espiner EA, Ikram H, Nicholls MG. Brain natriuretic factor: regional plasma concentrations and correlations with haemodynamic state in cardiac disease. British heart journal. 1993;69(5):414-7.

[111] Abe Y, Kobayashi T, Narita K, Kobayashi H, Kitadate A, Takeuchi M, *et al*. N-Terminal Pro-Brain Natriuretic Peptide (NT-proBNP) Is Associated with Both Left Ventricular Diastolic Dysfunction and Myeloma-Related Renal Insufficiency and Robustly Predicts Mortality in Patients with Symptomatic Multiple Myeloma. Blood. 2018;132(Supplement 1):4478-.

[112] Milani P, Vincent Rajkumar S, Merlini G, Kumar S, Gertz MA, Palladini G, *et al*. N-terminal fragment of the type-B natriuretic peptide (NT-proBNP) contributes to a simple new frailty score in patients with newly diagnosed multiple myeloma. American journal of hematology. 2016;91(11):1129-34.

[113] Dambha-Miller H, Everitt H, Little P. Clinical scores in primary care. British Journal of General Practice. 2020;70(693):163

[114] Chanukuppa V, Taware R, Taunk K, Chatterjee T, Sharma S, Somasundaram V, *et al*. Proteomic Alterations in Multiple Myeloma: A Comprehensive Study Using Bone Marrow

Interstitial Fluid and Serum Samples. Frontiers in Oncology. 2021;10(3164)

[115] Cardoso AL, Fernandes A, Aguilar-Pimentel JA, de Angelis MH, Guedes JR, Brito MA, *et al*. Towards frailty biomarkers: Candidates from genes and pathways regulated in aging and age-related diseases. Ageing Research Reviews. 2018;47:214-77.

[116] de Magalhães JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics (Oxford, England). 2009;25(7):875-81.

[117] Falandry C, Gilson E, Rudolph KL. Are aging biomarkers clinically relevant in oncogeriatrics? Crit Rev Oncol Hematol. 2013;85(3):257-65

[118] Saedi AA, Feehan J, Phu S, Duque G. Current and emerging biomarkers of frailty in the elderly. Clinical interventions in aging. 2019;14:389-98.

[119] Stastna M, Van Eyk JE. Secreted proteins as a fundamental source for biomarker discovery. Proteomics. 2012;12(4-5):722-35.

[120] Brando C. Protein Biomarkers; The correlation between Proteins, Health Conditions and Exogenous Products. Pharmaceutical Outsourcing Newsletter. Posted online16/5/22. Available online: https://www.pharmoutsourcing.com/Featured-Articles/586534-Protein-Biomarkers-The-Correlation-Between-Proteins-Health-Conditions-and-Exogenous-Products/#disqus_thread (Accessed 13/11/22)

[121] Sidoli S, Lin S, Xiong L, Bhanu NV, Karch KR, Johansen E, *et al*. Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications. Mol Cell Proteomics. 2015;14(9):2420-8

[122] Grissa D, Pétéra M, Brandolini M, Napoli A, Comte B, Pujos-Guillot E. Feature Selection Methods for Early Predictive Biomarker Discovery Using Untargeted Metabolomic Data. Frontiers in Molecular Biosciences. 2016;3.

[123] UniProt Consortium; Find Your Protein; Wellcome Trust Genomics Campus, Cambridge; Internet. Accessed online 6/4/22: https://www.uniprot.org/

[124] National Library of Medicine; Gene Search; Bethesda, USA; Internet Accessed online 6/4/22: https://www.ncbi.nlm.nih.gov/gene/

[125] String Consortium; Protein Protein Interaction Networks; European Molecular Biology Laboratory; Internet; Accessed online 6/4/22: https://string-db.org/

[126] Miao J, Zhu W. Precision–recall curve (PRC) classification trees. Evolutionary Intelligence. 2022;15(3):1545-69.

[127] Telikicherla D, Marimuthu A, Kashyap MK, Ramachandra YL, Mohan S, Roa JC, *et al*. Overexpression of ribosome binding protein 1 (RRBP1) in breast cancer. Clinical proteomics. 2012;9(1):7-.

[128] Tsai HY, Yang YF, Wu AT, Yang CJ, Liu YP, Jan YH, *et al*. Endoplasmic reticulum ribosome-binding protein 1 (RRBP1) overexpression is frequently found in lung cancer patients and alleviates intracellular stress-induced apoptosis through the enhancement of GRP78. Oncogene. 2013;32(41):4921-31.

[129] Lv S, Shi Z, Wang X, Zheng P, Li H, Han Q, Li Z. Ribosome Binding Protein 1 Correlates with Prognosis and Cell Proliferation in Bladder Cancer. *Onco Targets Ther*.

2020;13:6699-6707

[130] Pan Y, Cao F, Guo A, Chang W, Chen X, Ma W, *et al*. Endoplasmic reticulum ribosome-binding protein 1, RRBP1, promotes progression of colorectal cancer and predicts an unfavourable prognosis. Br J Cancer. 2015;113(5):763-72.

[131] Pan X, Wu H, Chen G, Li W. Prognostic Value of Enolase Gene Family in Colon Cancer. Med Sci Monit. 2020;26:e922980.

[132] Carreras J, Hamoudi R, Nakamura N. Artificial Intelligence Analysis of Gene Expression Data Predicted the Prognosis of Patients with Diffuse Large B-Cell Lymphoma. Tokai J Exp Clin Med. 2020;45(1):37-48.

[133] Zhang Y, Yang C, Zhang M, Liu H, Gong C, Zhang J, *et al*. Interleukin enhancer-binding factor 3 and HOXC8 co-activate cadherin 11 transcription to promote breast cancer cells proliferation and migration. Oncotarget. 2017;8(64):107477-91.

[134] Moamer A, Hachim IY, Binothman N, Wang N, Lebrun JJ, Ali S. A role for kinesin-1 subunits KIF5B/KLC1 in regulating epithelial mesenchymal plasticity in breast tumorigenesis. EBioMedicine. 2019;45:92-107.

[135] Cardoso CM, Groth-Pedersen L, Høyer-Hansen M, Kirkegaard T, Corcelle E, Andersen JS, *et al*. Depletion of kinesin 5B affects lysosomal distribution and stability and induces peri-nuclear accumulation of autophagosomes in cancer cells. PLoS One. 2009;4(2):e4424.

[136] Shan Y, Yang G, Huang H, Zhou Y, Hu X, Lu Q, *et al*. Ubiquitin-Like Modifier Activating Enzyme 1 as a Novel Diagnostic and Prognostic Indicator That Correlates With Ferroptosis and the Malignant Phenotypes of Liver Cancer Cells. Front Oncol. 2020;10:592413.

[137] Ohwada C, Nakaseko C, Koizumi M, Takeuchi M, Ozawa S, Naito M, *et al*. CD44 and hyaluronan engagement promotes dexamethasone resistance in human myeloma cells. Eur J Haematol. 2008;80(3):245-50

[138] Bjorklund CC, Baladandayuthapani V, Lin HY, Jones RJ, Kuiatse I, Wang H, *et al*. Evidence of a role for CD44 and cell adhesion in mediating resistance to lenalidomide in multiple myeloma: therapeutic implications. Leukemia. 2014;28(2):373-83.

[139] Dahl IM, Rasmussen T, Kauric G, Husebekk A. Differential expression of CD56 and CD44 in the evolution of extramedullary myeloma. Br J Haematol. 2002;116(2):273-7.

[140] Vincent T, Mechti N. IL-6 regulates CD44 cell surface expression on human myeloma cells. Leukemia. 2004;18(5):967-75

[141] Hassn Mesrati M, Syafruddin SE, Mohtar MA, Syahir A. CD44: A Multifunctional Mediator of Cancer Progression. Biomolecules. 2021;11(12).

[142] Perl A, Hanczko R, Telarico T, Oaks Z, Landas S. Oxidative stress, inflammation and carcinogenesis are controlled through the pentose phosphate pathway by transaldolase. Trends Mol Med. 2011;17(7):395-403.

[143] Kočevar N, Odreman F, Vindigni A, Grazio SF, Komel R. Proteomic analysis of gastric cancer and immunoblot validation of potential biomarkers. World J Gastroenterol. 2012;18(11):1216-28.

[144] Nunomiya K, Shibata Y, Abe S, Inoue S, Igarashi A, Yamauchi K, *et al*. Relationship between Serum Level of Lymphatic Vessel Endothelial Hyaluronan Receptor-1 and Prognosis in Patients with Lung Cancer. J Cancer. 2014;5(3):242-7.

[145] Karinen S, Juurikka K, Hujanen R, Wahbi W, Hadler-Olsen E, Svineng G, *et al*. Tumour cells express functional lymphatic endothelium-specific hyaluronan receptor in vitro and in vivo: Lymphatic mimicry promotes oral oncogenesis? Oncogenesis. 2021;10(3):23.

[146] Banerji S, Ni J, Wang SX, Clasper S, Su J, Tammi R, *et al*. LYVE-1, a new homologue of the CD44 glycoprotein, is a lymph-specific receptor for hyaluronan. J Cell Biol. 1999;144(4):789-801.

[147] Sathyan S, Ayers E, Gao T, Milman S, Barzilai N, Verghese J. Plasma proteomic profile of frailty. Aging Cell. 2020;19(9):e13193.

[148] Rassart E, Desmarais F, Najyb O, Bergeron KF, Mounier C. Apolipoprotein D. Gene. 2020;756:144874.

[149] Sanchez D, Ganfornina MD. The Lipocalin Apolipoprotein D Functional Portrait: A Systematic Review. Front Physiol. 2021;12:738991

[150] Sasaki Y, Negishi H, Koyama R, Anbo N, Ohori K, Idogawa M, *et al*. p53 family members regulate the expression of the apolipoprotein D gene. J Biol Chem. 2009;284(2):872-83.

[151] Sekikawa S, Onda T, Miura N, Nomura T, Takano N, Shibahara T, *et al*. Under expression of α-1-microglobulin/bikunin precursor predicts a poor prognosis in oral squamous cell carcinoma. Int J Oncol. 2018;53(6):2605-14.

[152] Chhetri JK, Ma L, Zheng Z, Liu FY, Zhao J, Gu ZQ, *et al*. Apolipoprotein E Polymorphism and Frailty: Apolipoprotein ε4 Allele Is Associated with Fatigue but Not Frailty Syndrome in a Community-Dwelling Older Population Cohort. J Nutr Health Aging. 2021;25(4):410-5.

[153] Rockwood K, Nassar B, Mitnitski A. Apolipoprotein E-polymorphism, frailty and mortality in older adults. J Cell Mol Med. 2008;12(6b):2754-61.

[154] Xu X, Wan J, Yuan L, Ba J, Feng P, Long W, *et al*. Serum levels of apolipoprotein E correlates with disease progression and poor prognosis in breast cancer. Tumor Biology. 2016;37(12):15959-66.

[155] Ren L, Yi J, Li W, Zheng X, Liu J, Wang J, *et al*. Apolipoproteins and cancer. Cancer Medicine. 2019;8(16):7032-43.

[156] Strandberg TE, Levinson SL, DiNubile MJ, Jyväkorpi S, Kivimäki M. Association of plasma gelsolin with frailty phenotype and mortality among octogenarian community-dwelling men: a cohort study. Aging Clin Exp Res. 2022.

[157] Chiu CT, Wang PW, Asare-Werehene M, Tsang BK, Shieh DB. Circulating Plasma Gelsolin: A Predictor of Favorable Clinical Outcomes in Head and Neck Cancer and Sensitive Biomarker for Early Disease Diagnosis Combined with Soluble Fas Ligand. Cancers (Basel). 2020;12(6).

[158] Chen Z, Li K, Yin X, Li H, Li Y, Zhang Q, *et al*. Lower Expression of Gelsolin in Colon Cancer and Its Diagnostic Value in Colon Cancer Patients. J Cancer. 2019;10(5):1288-

96.

[159] Pan S, Chen R, Brand RE, Hawley S, Tamura Y, Gafken PR, *et al*. Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. J Proteome Res. 2012;11(3):1937-48.

[160] Jiang X, Bai XY, Li B, Li Y, Xia K, Wang M, *et al*. Plasma Inter-Alpha-Trypsin Inhibitor Heavy Chains H3 and H4 Serve as Novel Diagnostic Biomarkers in Human Colorectal Cancer. Dis Markers. 2019;2019:5069614.

[161] Hamm A, Veeck J, Bektas N, Wild PJ, Hartmann A, Heindrichs U, *et al*. Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: A systematic expression analysis. BMC Cancer. 2008;8(1):25.

[162] Seki R, Yamagishi S, Matsui T, Yoshida T, Torimura T, Ueno T, *et al*. Pigment epithelium-derived factor (PEDF) inhibits survival and proliferation of VEGF-exposed multiple myeloma cells through its anti-oxidative properties. Biochem Biophys Res Commun. 2013;431(4):693-7.

[163] Cheng G, Song C. Association of pigment epithelium derived factor expression with cancer progression and prognosis: a meta-analysis study. Discov Oncol. 2021;12(1):61.

[164] Wojtukiewicz MZ, Hempel D, Sierko E, Tucker SC, Honn KV. Endothelial Protein C Receptor (EPCR), Protease Activated Receptor-1 (PAR-1) and Their Interplay in Cancer Growth and Metastatic Dissemination. Cancers. 2019;11(1).

[165] Hu J, Duan B, Jiang W, Fu S, Gao H, Lu L. Epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1) suppressed the growth of hepatocellular carcinoma cells by promoting Semaphorin 3B(SEMA3B). Cancer medicine. 2019;8(6):3152-66.

[166] Zhu X-J, Liu J, Xu X-Y, Zhang C-D, Dai D-Q. Novel tumor-suppressor gene epidermal growth factor-containing fibulin-like extracellular matrix protein 1 is epigenetically silenced and associated with invasion and metastasis in human gastric cancer. Mol Med Rep. 2014;9(6):2283-92.

[167] Rahmati Nezhad P, Riihilä P, Knuutila JS, Viiklepp K, Peltonen S, Kallajoki M, *et al*. Complement Factor D Is a Novel Biomarker and Putative Therapeutic Target in Cutaneous Squamous Cell Carcinoma. Cancers (Basel). 2022;14(2).

[168] Annweiler C, Bataille R, Ferrière N, Douillet D, Fantino B, Beauchet O. Plasma Beta-2 Microglobulin as a Marker of Frailty in Older Adults: A Pilot Study. The Journals of Gerontology: Series A. 2011;66A(10):1077-9.

[169] Kim M, Suzuki T, Kojima N, Yoshida H, Yoshida Y, Hirano H, *et al*. Association Between Serum β(2) -Microglobulin Levels and Prevalent and Incident Physical Frailty in Community-Dwelling Older Women. J Am Geriatr Soc. 2017;65(4):e83-e8.

[170]Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, Bladé J, *et al*. International staging system for multiple myeloma. J Clin Oncol. 2005;23(15):3412-20.

[171] Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, *et al*. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of Clinical Oncology. 2015;33(26):2863-9.

[172] Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, *et al*. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of

SWATH-mass spectrometry. Nature Communications. 2017;8(1):291.

[173] van der Gugten JG. Tandem mass spectrometry in the clinical laboratory: A tutorial overview. Clinical Mass Spectrometry. 2020;15:36-43.

[174] Ketha H, Kaur S, Grebe SK, Singh RJ. Clinical applications of LC-MS sex steroid assays: evolution of methodologies in the 21st century. Current Opinion in Endocrinology, Diabetes and Obesity. 2014;21(3).

[175] Barnidge DR, Tschumper RC, Theis JD, Snyder MR, Jelinek DF, Katzmann JA, *et al*. Monitoring M-proteins in patients with multiple myeloma using heavy-chain variable region clonotypic peptides and LC-MS/MS. J Proteome Res. 2014;13(4):1905-10.

[176] Hmmier A, Dowling P. DIGE Analysis Software and Protein Identification Approaches. In: Ohlendieck K, editor. Difference Gel Electrophoresis: Methods and Protocols. New York, NY: Springer New York; 2018. p. 41-50.

[177] Alhajj M, Farhana A. Enzyme Linked Immunosorbent Assay. [Updated 2022 Feb 2]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Accessed 25/5/22. Available from: https://www.ncbi.nlm.nih.gov/books/NBK555922/

[178] Bjelosevic S, Pascovici D, Ping H, Karlaftis V, Zaw T, Song X, *et al*. Quantitative Age-specific Variability of Plasma Proteins in Healthy Neonates, Children and Adults*. Molecular & Cellular Proteomics. 2017;16(5):924-35.

[179] Ignjatovic V, Lai C, Summerhayes R, Mathesius U, Tawfilis S, Perugini MA, *et al*. (2011) Age-Related Differences in Plasma Proteins: How Plasma Proteins Change from Neonates to Adults. PLoS ONE 6(2): e17213.

[180] Santos-Lozano A, Valenzuela PL, Llavero F, Lista S, Carrera-Bastos P, Hampel H, *et al*. Successful aging: insights from proteome analyses of healthy centenarians. Aging. 2020;12(4):3502-15.

[181] Ferraris VA. Commentary: Should we rely on receiver operating characteristic curves? From submarines to medical tests, the answer is a definite maybe! The Journal of Thoracic and Cardiovascular Surgery. 2019;157(6):2354-5.

[182] Prizment AE, Linabery AM, Lutsey PL, Selvin E, Nelson HH, Folsom AR, *et al*. Circulating Beta-2 Microglobulin and Risk of Cancer: The Atherosclerosis Risk in Communities Study (ARIC). Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2016;25(4):657-64.

[183] Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ. 2002;324(7351):1417

[184] Röllig C, Knop S, Bornhäuser M. Multiple myeloma. The Lancet. 2015;385(9983):2197-208.

[185] Stewart AK, Dimopoulos MA, Masszi T, Špička I, Oriol A, Hájek R, *et al*. Health-Related Quality-of-Life Results From the Open-Label, Randomized, Phase III ASPIRE Trial Evaluating Carfilzomib, Lenalidomide, and Dexamethasone Versus Lenalidomide and Dexamethasone in Patients With Relapsed Multiple Myeloma. Journal of Clinical Oncology. 2016;34(32):3921-30

[186] Maes H, Delforge M. Optimizing quality of life in multiple myeloma patients: current options, challenges and recommendations. Expert Review of Hematology.

2015;8(3):355-66.

[187] Stevens A, Abrams K, Brazier J *et al*. Criteria for assessing patient based outcome measures for use in clinical
trials. In: Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, eds. *The advanced handbook of methods in evidence based healthcare.* London: Sage, 2001:181–194

[188] Fallowfield L. Quality of life: a new perspective for cancer patients. Nature Reviews Cancer. 2002;2(11):873-

[189] Saunders CM, Fallowfield LJ. Survey of the attitudes to and use of quality of life measures by breast cancer specialists in the UK. The Breast. 1996;5(6):425-6.

[190] EORTC QLQ-C30 Questionnaire. Accessed online 11/10/21: https://www.eortc.org/app/uploads/sites/2/2018/08/Specimen-QLQ-C30-English.pdf

[191] EORTC MY-20 Questionnaire. Accessed online 11/10/21 https://www.eortc.org/app/uploads/sites/2/2018/08/Specimen-MY20-English.pdf

[192] Wisløff F, Eika S, Hippe E, Hjorth M, Holmberg E, Kaasa S, *et al*. Measurement of health-related quality of life in multiple myeloma. British Journal of Haematology. 1996;92(3):604-13.

[193] Cocks K, Cohen D, Wisløff F, Sezer O, Lee S, Hippe E, *et al*. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. European Journal of Cancer. 2007;43(11):1670-8.

[194] Engelhardt M, Ihorst G, Singh M, Rieth A, Saba G, Pellan M, *et al*. Real-World Evaluation of Health-Related Quality of Life in Patients With Multiple Myeloma From Germany. Clinical Lymphoma Myeloma and Leukemia. 2021;21(2):e160-e75.

[195] Avilés A, Zepeda G, Guzmán R, Talavera A, García EL, Díaz-Maqueo JC. [Prognostic importance of beta-2-microglobulin in multiple myeloma]. Rev Invest Clin. 1992;44(2):215-20.

[196] Cairns D, Pawlyn C, Royle K-L, Best P, Bird J, Bowcock S, *et al*. Frailty-Adjusted Therapy in Transplant Non-Eligible Patients with Newly Diagnosed Multiple Myeloma (FiTNEss): A UK Myeloma Research Alliance Study, Myeloma XIV. Blood. 2019;134(Supplement_1):3153-.

[197] Sully K, Trigg A, Bonner N, Moreno-Koehler A, Trennery C, Shah N, *et al*. Estimation of minimally important differences and responder definitions for EORTC QLQ-MY20 scores in multiple myeloma patients. Eur J Haematol. 2019;103(5):500-9

[198] WislØff F, Hjorth M. Health-related quality of life assessed before and during chemotherapy predicts for survival in multiple myeloma. British Journal of Haematology. 1997;97(1):29-37.

[199] Kumar S, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, Colby C, *et al*. Revised Prognostic Staging System for Light Chain Amyloidosis Incorporating Cardiac Biomarkers and Serum Free Light Chain Measurements. Journal of Clinical Oncology. 2012;30(9):989-95

[200] Lehrke S, Steen H, Kristen AV, Merten C, Lossnitzer D, Dengler TJ, *et al*. Serum levels of NT-proBNP as surrogate for cardiac amyloid burden: new evidence from gadolinium-

enhanced cardiac magnetic resonance imaging in patients with amyloidosis. Amyloid. 2009;16(4):187-95