Development of clinical prediction rules for reducing delays in the diagnosis of multiple myeloma

CONSTANTINOS KOSHIARIS

Green Templeton College

A thesis submitted to the Division of Medical Sciences at the

University of Oxford for the Degree of

Doctor of Philosophy

Michaelmas Term 2018

Nuffield Department of Primary Care Health Sciences

University of Oxford



This thesis is dedicated to my parents and sister.

For their endless love, support and encouragement

Acknowledgments

I would like to thank my DPhil supervisors, Associate Professor Ann Van den Bruel and Dr Jason Oke for all the help, advice and support they have offered me. This thesis would not have been possible without their guidance.

Dr Brian D Nicholson for introducing me to the topic, and providing knowledge and expertise on the subject.

Professor William Hamilton and Dr Elizabeth Shephard for sharing their databases and for their hospitality when visiting the University of Exeter

Associate Professor Clare Bankhead and Dr Sarah Lay-Flurrie for their input and help during the development of the CPRD study protocol and performing the data cut.

My fellow DPhil candidates and colleagues in the Gibson building for always supporting me throughout the DPhil. In particular I would like to thank Ali Albasri, Charlotte Albury, Jack O'Sullivan, Dr Joseph Lee, Dr Oliver Van Hecke, Dr Ben Fletcher, Esan Oluwaseun, Kerstin Frie and Markus Schichtel.

Finally I would like to thank my whole family for their endless love, support and encouragement.

This thesis presents independent research funded by funded by the Primary Care Research Trust, the University of Oxford, and National Institute for Health Research (NIHR) Oxford CLAHRC. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health.

Contents

Acknowledgments 3
Abstract 8
Chapter 1: Multiple myeloma and diagnostic delays: Background and literature review 10
1.1 Introduction 10
1.2 Epidemiology of multiple myeloma and risk factors 13
1.3 Complications, prognosis and clinical features of myeloma 14
1.4 Current NICE guidelines for diagnosis of multiple myeloma in the UK 14
1.4.1 Primary care investigations 15
1.4.2 Secondary care investigations 15
1.5 Treatment for multiple myeloma 16
1.6 Diagnostic trajectories for patients with multiple myeloma 18
1.6.1 Importance of reducing delays in cancer diagnosis 18
1.6.2 Delays observed in multiple myeloma and the association with cancer outcomes 18
1.6.3 Cancer referral pathways and suspicion of cancer in primary the primary care setting 20
1.7 Use of clinical prediction rules to aid clinical decision making 21
1.8 Specific aims and objectives of the thesis 22
1.9 Studies and structure of the thesis 23
1.10 Dissemination of results 24
References 25
Chapter 2: Quantifying diagnostic intervals in myeloma: a systematic review and meta-analysis 34
2.1 Introduction 34
2.2 Methods 35
2.2.1 Outcomes, Search strategy and inclusion/exclusion criteria 36
2.2.2 Data extraction 36
2.2.3 Risk of bias tool 37
2.2.4 Analysis and synthesis of data 37
2.3 Results 38
2.3.1 Study characteristics 39
2.3.2 Definitions of intervals and symptoms 42
2.3.3 Risk of bias assessment 44
2.3.4 Estimation of the length of intervals 45
2.4 Discussion 49
2.4.1 Summary of findings 49
2.4.2 Strengths and weaknesses 49

2.4.3 Sources of Heterogeneity 52
2.4.4 Comparison of findings with existing literature 53
2.4.5 Implications for future research and clinical practice 54
2.5 Conclusion and implications for next chapter 55
References 56
Chapter 3: Identifying opportunities for early diagnosis of myeloma: a case-control study based on Electronic health care records (CPRD) 61
3.1 Chapter Introduction 61
3.2 Research questions 62
3.3 Methods 62
3.3.1 Study design 62
3.3.2 Variable definitions 63
3.3.3 Statistical analysis 64
3.4 Results 67
3.4.1 Timings of symptoms and investigations 68
3.4.2 Rule in and rule out tests 73
3.4.3 Quantification of diagnostic pathway 77
3.5 Discussion 79
3.5.1 Summary 79
3.5.2 Strengths and limitations 80
3.5.3 Comparison with existing literature and findings from the previous chapter 81
3.5.4 Implications for practice 83
3.6 Conclusion and implications for next chapter 85
References 86
Chapter 4: Derivation of clinical prediction rules for early diagnosis of myeloma: a retrospective cohort study based on CPRD data 90
4.1 Chapter Introduction 90
4.2 Objectives and aims 91
4.3 Methods 91
4.3.1 Study design 91
4.3.2 Sample size considerations 92
4.3.3 Variable definitions 92
4.3.4 Statistical analysis 94
4.4 Results 102
4.4.1 Baseline characteristics of derivation and validation datasets 102
4.4.2 Univariate analysis 104

4.4.3 Multiple imputation 107
4.4.4 Full blood count model (model 1) 108
4.4.5 Difference in blood test parameters models (model 2) 111
4.4.6 All tests model (model 3) 113
4.5 Discussion 117
4.5.1 Summary of results 117
4.5.2 Strengths and limitations 117
4.5.3 Comparison with existing literature and finding from previous chapter 120
4.5.4 Developed clinical prediction rules 121
4.5.5 Steps for the next chapter 124
References 125
Chapter 5: Statistical validation, clinical utility and implementation of diagnostic prediction rules for early diagnosis of myeloma 128
5.1 Chapter Introduction 128
5.2 Aims 129
5.3 Methods 129
5.3.1 Discrimination and calibration 130
5.3.2 Clinical utility 132
5.3.3 Net benefit and Decision curve analysis 133
5.3.4 Combining estimates in multiple imputation 137
5.4 Results 138
5.4.1 Test performance (individual and combinations) 138
5.4.2 Model validation - Discrimination and Calibration 141
5.4.3 Comparison of different strategies for diagnosis of myeloma 143
5.4.4 Comparison of different strategies using net benefit approaches 144
5.5 Discussion 147
5.5.1 Summary 147
5.5.2 Comparison with existing literature and findings from previous chapters 147
5.5.3 Selection of final model based on clinical, statistical and implementation criteria 148
5.5.4 Threshold determination 149
5.5.5 Strength and limitations 149
5.5.6 Simulation of different scenarios based on the selected clinical prediction rule and presentation of risk prediction tool for myeloma 151
5.5.7 Implications for future research 153
5.5.8 Implications for future practice 154
References 155

6. Discussion 160
6.1 Summary of findings 160
6.2 Strengths and limitations of the thesis 164
6.3 Implications 170
6.3.1 Clinical Implications 170
6.3.2 Future research 172
6.3.3 Concluding remarks 174
References 175
APPENDIX 1 – Publication from thesis 179
APPENDIX 2 - Publication from thesis 190
APPENDIX 3 – GPUpdate (myeloma) 199
APPENDIX 4 – Search strategy for systematic review 203
APPENDIX 5 –Risk of bias summary 204
APPENDIX 6 – Risk of bias summary 205
APPENDIX 7 – Sensitivity analysis using random effects meta-analysis 206
APPENDIX 8 –Read codes 208
APPENDIX 9 - The rule-out value of individual and combinations of investigations for myeloma (FBC instead of haemoglobin). 240
APPENDIX 10 - Pre/Post - test probabilities 241
APPENDIX 11 – ISAC protocol approval 242
APPENDIX 12 - Missing data patterns for blood tests 243
APPENDIX 13 - Kaplan Meir plots / proportional hazards assumptions examination for binary predictors 244
APPENDIX 14 - Imputation diagnostics 252
APPENDIX 15 – Probability plots 258
APPENDIX 16– TRIPOD statement 259

Abstract

Background

Multiple myeloma is a type of blood cancer which starts at the bone marrow. Fifty percent of patients have more than three consultations in primary care before being referred to secondary care. This implies that myeloma patients can experience delays in the diagnosis which could contribute to the relatively poor prognosis. The aim of the thesis was to develop clinical prediction rules that could potentially reduce the time to diagnosis of myeloma patients.

Methods

The thesis comprises three main studies. The first is a systematic review which maps and quantifies the diagnostic pathway of myeloma patients. In the second study, the features of myeloma (symptoms and blood test abnormalities) were examined to identify which present the earliest. The most optimal combinations of blood tests for diagnosing myeloma were also examined. In the third study, a series of clinical prediction rules were developed and validated using the Clinical Research Datalink data that could be used to expedite the diagnosis of myeloma.

Results

Myeloma patients experience substantial delays with 50% experiencing a diagnostic interval greater than three months and 25% more than eight months. Early symptoms of myeloma include back pain, rib pain, chest pain and infections while fractures, weight loss and nausea manifest later in disease progression. For blood tests haemoglobin and inflammatory markers like Erythrocyte sedimentation rate (ESR) and plasma viscosity (PV) can become abnormal up to two years before diagnosis while values in calcium and creatinine manifest later. C-reactive protein is not a useful inflammatory marker for the diagnosis of myeloma. A combination of normal haemoglobin, either ESR or PV and calcium can rule out most myeloma cases. A clinical prediction rule containing demographics (age, gender, BMI), risk factors (MGUS)

symptoms (nosebleeds, back pain, chest pain, rib pain) and the parameters of the full blood count (white cell count, haemoglobin, platelets and mean corpuscular volume) showed good discrimination (AUC: 0.84, 95% CI: 0.82-0.87, R^2: 0.57, 95% CI: 0.53-0.62 and D-statistic: 2.4, 95% CI: 2.2-2.6) and good calibration.

Conclusions

Myeloma patients experience substantial delays in their diagnosis. General practitioners should examine patients that present with myeloma related symptoms using a combination of a full blood count, an inflammatory marker (ESR or PV) and calcium. The clinical prediction rule developed can be useful when myeloma is not suspected and not all tests are ordered. The rule could be applied in the laboratory or on the electronic systems that general practitioners use, but further work is required before implementation.

Chapter 1: Multiple myeloma and diagnostic delays: Background and literature review

1.1 Introduction

Multiple myeloma is a haematological malignancy that starts in the bone marrow (1,2). The bone marrow is a spongy material responsible for the production of stem cells which in turn develop into the three main types of blood cells: the red blood cells (responsible for carrying oxygen in the body), platelets (responsible for clotting the blood and controlling bleeding) and the white blood cells (responsible for fighting infections).

Myeloma affects a particular type of white cells known as plasma cells. Plasma cells produce immunoglobulins (also known as antibodies) which are responsible for destroying viruses, bacteria and foreign cells found in the body. In the presence of an infection the bone marrow produces more plasma cells and immunoglobulins in order to fight it. In multiple myeloma, rather than healthy plasma cells a large number of abnormal plasma cells are produced that secrete an antibody called paraprotein (M-protein) which also reproduces and accumulates. The accumulation of the abnormal cells and M-proteins overload the bone marrow and disrupt the production of healthy red and white blood cells, and platelets. Figures 1 and 2 illustrate the uncontrollable myeloma cell growth and the difference between normal and abnormal plasma cells. Due to the fact that myeloma affects the plasma cells it can develop in any part of the body where plasma cells can be found such as the spine, ribcage or pelvis. This is the main reason why myeloma is known as multiple.

Figure 1: Myeloma cell growth

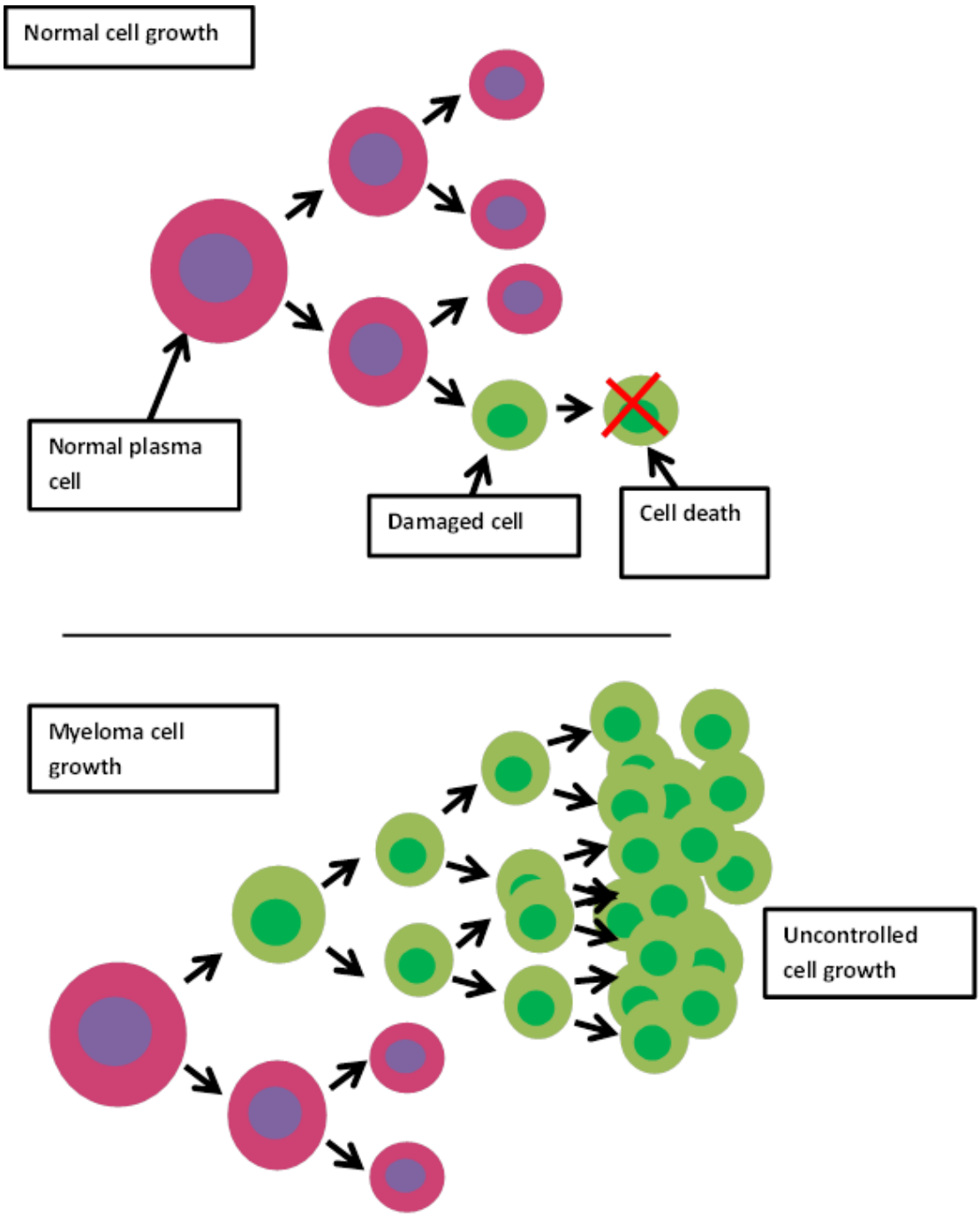
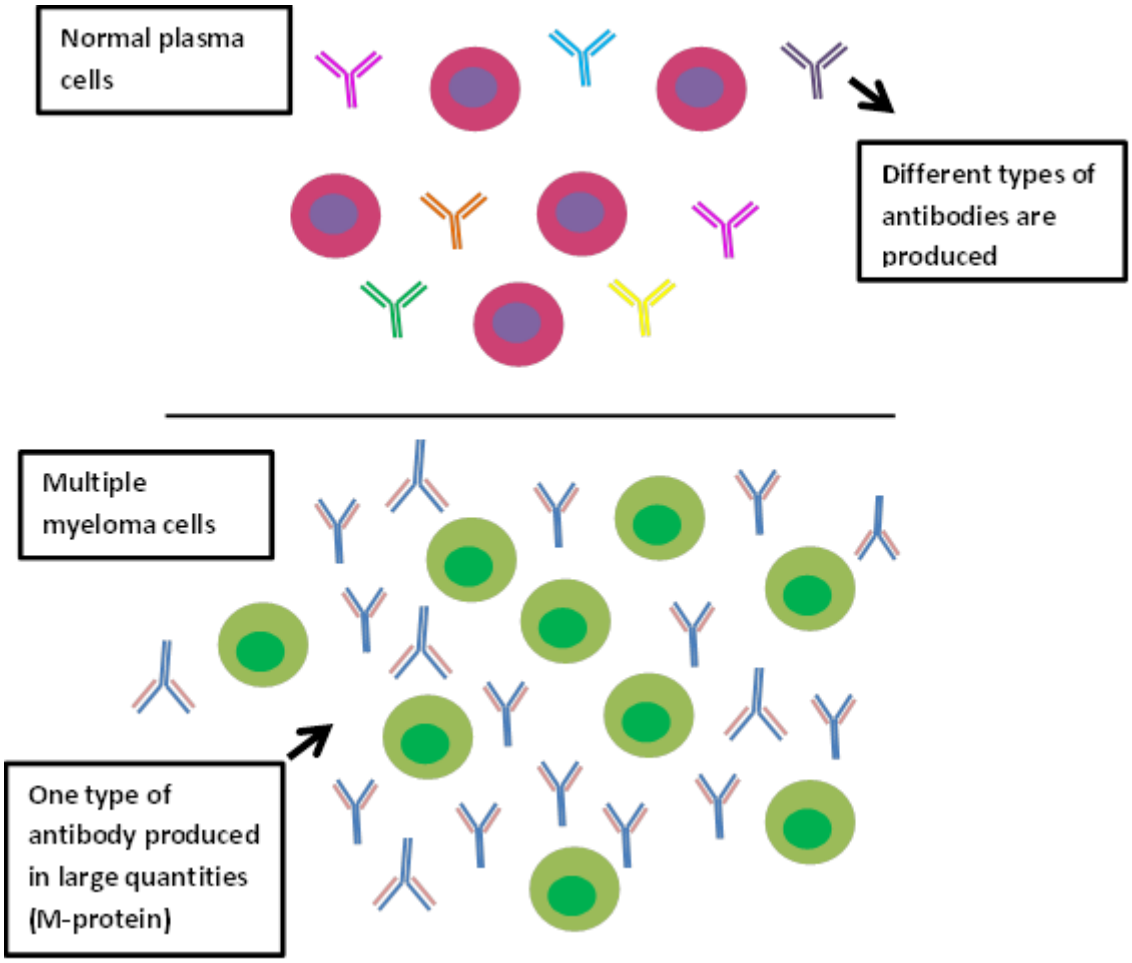


Figure 2: Myeloma plasma cells



Multiple myeloma has been shown to be preceded by a premalignant plasma-cell disorder called monoclonal gammopathy of undetermined significance (MGUS) which is asymptomatic and is considered a non-cancerous condition (3). MGUS is detected in patients incidentally when tests are done for other reasons. MGUS patients are at higher risk of developing myeloma than the general population. It has been shown that the risk is increased by approximately 1% per year (4). More specifically the risk of progression since MGUS diagnosis is 17% at 10 years, 34% at 20 years and 39% at 25 years (5). Smouldering myeloma is a term used to describe an intermediate stage between MGUS and multiple myeloma. At this stage the disease is still asymptomatic but the risk profile for this type of myeloma is different compared to MGUS as in this case the annual risk of progression to multiple myeloma is 10% (6).

Currently only symptomatic myeloma patients are treated. MGUS patients are monitored annually by having a clinical examination and blood tests that will be described in a later section (7). Criteria have been developed in order to determine whether a particular patient has symptomatic myeloma and thus requires treatment, or one of the asymptomatic forms of the disease which require observation (8).

1.2 Epidemiology of multiple myeloma and risk factors

Myeloma is the second most common haematological cancer after leukaemia accounting for 2% of all cancer cases. Nevertheless it is a very rare disease with a yearly incidence of approximately 9 per 100 000 population (9), mainly affecting the elderly population - the median age at diagnosis is 70 years (10). The incidence rate increased by 69% since 1975 in Great Britain and is expected to increase by 11% until 2035 based on cancer incidence projections (11).

Although survival in multiple myeloma has improved in the last decade due to the introduction of new treatment methods like stem cell transplantation, myeloma is still behind other cancers (12,13). The one year survival is 77%, five year survival 47% and ten year survival 33% (14). Compared with the other major haematological malignancies myeloma has worse 5 year survival rates than leukaemia (52%) and lymphoma at all points (5 year- 85%) (15,16)

Myeloma is more common in elderly, males and the black population (17). A person with a first degree relative with myeloma has approximately twice the risk of the general population (18). Other factors that have been shown to increase the risk of the disease include monoclonal gammopathies of undetermined significance (MGUS) that were described in the previous section, people that had a solid organ transplant, HIV and high BMI (19–22). Currently, only 14% of myeloma cases are considered preventable (23).

1.3 Complications, prognosis and clinical features of myeloma

Multiple myeloma often leads to infections, anaemia, bone lesions, hypercalcemia, renal failure and neurological problems(24,25). These complications are mainly caused by the destruction of red and white blood cells and make the patient more prone to serious infections.

Renal impairment is one of the most frequent complications and is present in more than 20% of myeloma patients (25,26). The abnormal proteins produced by the myeloma cells travel to the kidneys where they block the tubules and alter their filtering property reducing renal function and leading eventually to renal failure. Renal failure is considered a medical emergency and at least 10% of myeloma patients require dialysis (27,28). Renal failure and infection contribute to more than 50% of myeloma related deaths (29,30). Hypercalcaemia, caused by bone destruction have been identified as independent prognostic factor for survival (31).

Other serious complications of myeloma which require emergency treatment and also contribute to mortality include spinal cord compression and pathological fractures (32,33). Spinal cord compression is a neurological emergency and it can lead to permanent loss of function if not treated immediately.

All these complications give rise to vague symptoms. For example, anaemic patients tend to experience fatigue, exhaustion, dizziness and shortness of breath while high calcium and reduced renal function can lead to nausea, fatigue (similar to anaemia), chest pain, confusion etc. Other very common symptoms include musculoskeletal symptoms, mainly back pain and rib pain. Bleeding can also occur due to the reduction in platelets which does not allow blood to clot properly. Many patients experience weight loss which is also a very common symptom in other cancers.

1.4 Current NICE guidelines for diagnosis of multiple myeloma in the UK

Currently the National Institute for Health and Care Excellence (NICE) guidelines for multiple myeloma diagnosis recommend the following (34):

1.4.1 Primary care investigations

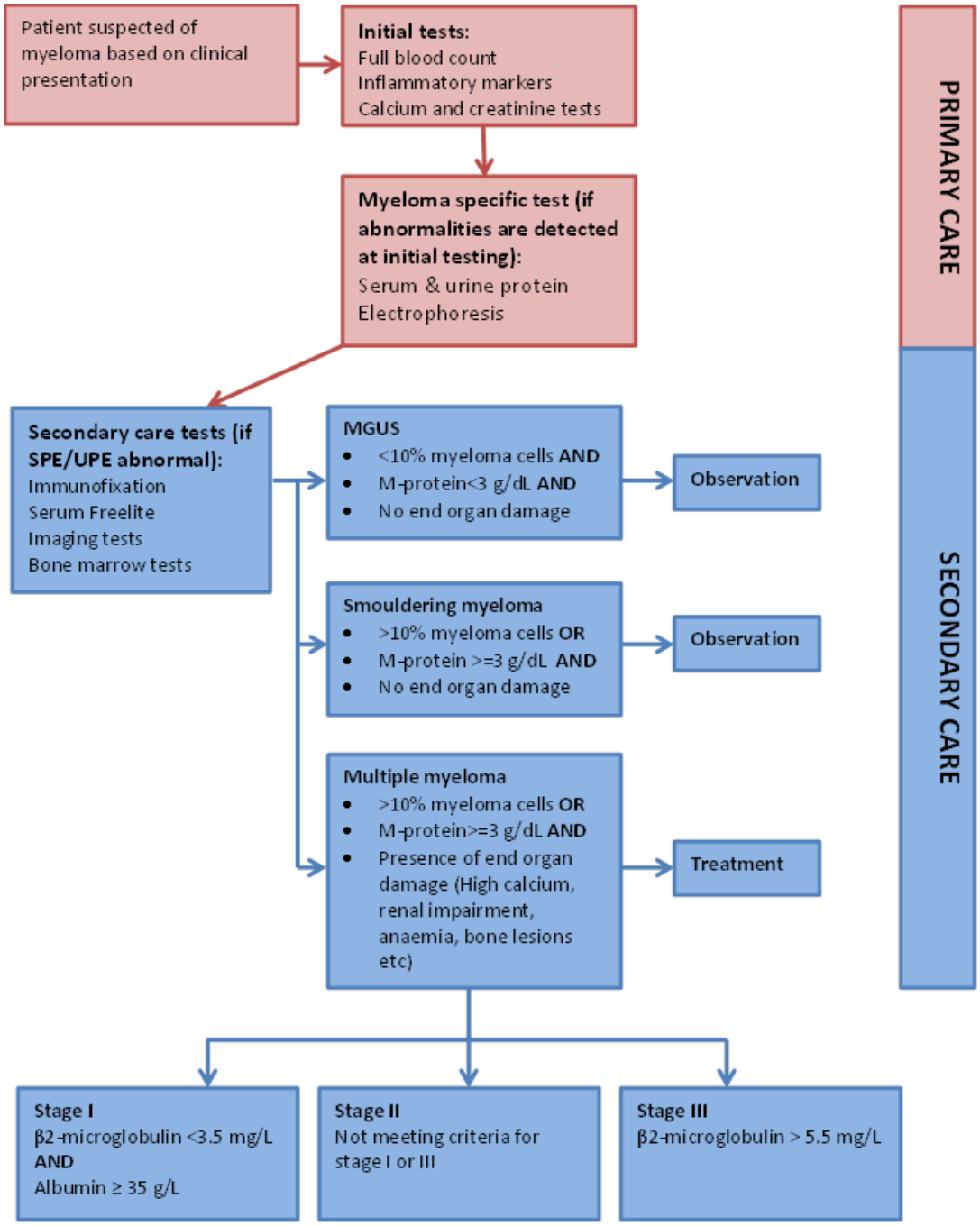
Myeloma should be suspected in patients who present in primary care with non-explained bone pain especially with lower back or chest pain, fatigue, symptoms of high calcium (bone pain, abdominal pain, depression, confusion, muscle weakness, constipation, thirst, polyuria), Weight loss, hyperviscosity symptoms (headache, visual disturbance, cognitive impairment, mucosal bleeding, breathlessness), symptoms of spinal cord compression (sensory loss, paraesthesia, limb weakness, walking difficulty, sphincter disturbance, spinal deformity), or fever.

Patients who are suspected of myeloma should be examined by a range of tests that include a full blood count, calcium and an inflammatory marker (35). Abnormal results in these tests should be followed by tests for myeloma cell proteins in the urine and blood (protein electrophoresis). The presence of these proteins makes a diagnosis of myeloma more likely. In 10-15% of myeloma cases, the myeloma protein is only detectable in urine, which means both urine and serum needs to be tested (36). All tests are available in primary care, and an abnormal result on the electrophoresis tests should result in an urgent referral to a haematologist department.

1.4.2 Secondary care investigations

Once referred to secondary care a patient undergoes further investigations in order to confirm a diagnosis of myeloma. A patient is likely to have additional biochemical tests such as serum and urine Immunofixation, imaging and possibly a bone marrow biopsy. A normal bone marrow has fewer than 5% per cent normal plasma cells while in myeloma this is expected to be more than 10% clonal plasma cells. Presence of more than 60% clonal plasma cells is usually considered a myeloma defining event. In addition to the tests described above, patients who are diagnosed with myeloma also get tested with β2-microglobulin and albumin which are prognostic tests to establish the stage of the disease with the International Staging System (ISS) (37). Figure 3 illustrates the full diagnostic pathway for multiple myeloma including criteria for determining the type and stage of the disease.

Figure 3: Diagnostic pathway for myeloma in the UK



1.5 Treatment for multiple myeloma

Currently myeloma is a treatable but not a curable disease (38). Myeloma treatments aim to reduce and control the burden of the disease, minimize symptoms, prolong life and improve its quality. Myeloma treatments revolve around high dose chemotherapy which is harmful and toxic and affects the production of new blood cells in the bone marrow. This can result in

serious and life-threatening complications due to severe cytopenia (very low white cells, red cells, platelets etc.). The introduction of a new treatment called high dose therapy and autologous stem cell transplantation was able to significantly improve survival, with patients in the high dose group having a 5 year survival rate of 52% compared to 12% for patients in the conventional-dose group (12,13). The procedure involves the collection of the patient's stem cells, followed by high dose chemotherapy in order to kill the myeloma cells. The collected stem cells are then transplanted back to the patient helping with the production of new healthy blood cells.

Usually autologous stem cell transplantation with high dose chemotherapy are recommended only for patients younger than 65 years with no severe comorbidities, although some studies indicated that certain patients over the age 65 could also benefit from this treatment (39–41). Patients with multiple comorbidities like diabetes or dysfunction in major organs like the kidneys, heart, lung etc. are usually not eligible for autologous stem cell transplantation (42). Patients' fitness can be evaluated with scoring systems such as the Charlson comorbidity index for example (43). Treatment with reduced-dose approaches is recommended for patients who are not eligible for high-dose treatments (42).

Renal failure is one of the main reasons for which patients are excluded from autologous stem cell transplantation because high dose therapy leads to high toxicity in this patient group. One study demonstrated that lower dosage treatment can be applied to patients with renal failure (44). Amongst the patients who are dialysis dependent due to severe renal failure, only 13% become dialysis independent following autologous stem cell transplantation (45). Diagnosing myeloma before significant impact on renal function through earlier detection could improve patient prognosis by preserving treatment options.

1.6 Diagnostic trajectories for patients with multiple myeloma

1.6.1 Importance of reducing delays in cancer diagnosis

In 2016 NHS England invested more than £200 million for improving cancer services and reduce delays in the diagnosis of cancer (46). A systematic review that examined the association between longer time to diagnosis and cancer outcomes showed that several cancers could benefit from a shorter time to diagnosis in terms of survival, earlier stage at diagnosis and better quality of life but the benefits vary depending on the type of cancer (47). For example bladder cancer patients who experienced a delay of more than nine months had twice the risk of dying compared to patients diagnosed within three months (48). Similarly in pancreatic cancer (which is one of the worst cancers in terms of survival) increased time to diagnosis has been shown to be associated with increased mortality (an increase in risk of approximately 2% per week of delay) (49). In addition it has been shown that cancer patients that have three or more consultations are more likely to report worse care experience compared to patients with fewer consultations suggesting that long waiting times also affect patient-reported care experience (50).

1.6.2 Delays observed in multiple myeloma and the association with cancer outcomes

Multiple myeloma has been classified as one of the most difficult cancers to diagnose which can be attributed to the non-specific nature of the symptoms (51). Half of myeloma patients have more than three consultations in primary care before they get referred to secondary care (52). In comparison, fewer than 7% of breast cancer patients have more than three primary care consultations before being referred. Similar results were also observed in a survey of myeloma patients conducted by Myeloma Research UK where 47% of patients reported that they had to visit their GP more than four times before being referred to hospital for cancer (53). Myeloma patients have the highest percentage of emergency presentations of all cancers with approximately 30% being diagnosed through that route (54). The same study has also demonstrated that myeloma patients who present through emergency services have the worst survival compared to any other route (one-year survival via emergency 72.6% vs 88.6% via

other routes). 77% of myeloma patient diagnosed through ES had consulted with their GP and over half of them more than 3 times which leads to longer time to diagnosis (55,56).

In a study of 103 myeloma patients, only 30% of patients were diagnosed within three months after their first symptom, and in 40% more than six months passed until diagnosis (25). The median diagnostic interval for myeloma patients, which is the time from when a patient presents to a healthcare professional with a myeloma related symptom until diagnosis is approximately five months (57). In addition, 37% of myeloma patients reported that their health worsened while they were waiting to see a specialist compared to 20% for the rest of the cancers (53).

One of the most challenging aspects of research conducted on early cancer diagnosis is to establish an association between time to diagnosis and cancer outcomes especially survival. A systematic review has demonstrated that most studies that examine the association between diagnostic intervals and cancer outcomes (such as survival, stage, treatment response etc.) are of low quality and tend to have a lot of heterogeneity (47). This heterogeneity is usually attributed to the different definitions of time intervals and outcomes that are used in the different studies. In addition the waiting time paradox is also observed in a lot of studies, where the relationship between survival and waiting times is opposite from what is expected i.e. people with short delays have worse survival (58). This could be explained by the fact that patients that are seriously ill and have symptoms that are very indicative of cancer are promptly investigated which results in very short diagnostic intervals. All of the above demonstrate that investigating the association between diagnostic delays and cancer survival outcomes is quite complex and is not straightforward to quantify or establish.

Nevertheless, despite the difficulty in establishing the relationship between waiting times and cancer survival myeloma is also one of the cancers in which delays have been associated with worse outcomes. A study has shown that all the patients who had a delay in their diagnosis of more than six months had more than two complications which included infections, neurological problems, renal disease, bone problems and anaemia while 40% of patients diagnosed within

three months had no complications (24). This is important because a patient's treatment is determined by their overall health, meaning that delays may indirectly result in less effective treatment. For example, patients with renal failure are excluded from autologous stem cell transplantation and high dose chemotherapy. In addition, delays can affect the disease-free survival of myeloma patients because delays are associated with a higher cancer stage which is in turn associated with worse survival (59). Patients with comorbidities are more likely to experience diagnostic delays compared to patients with no comorbidities with a study reporting an odds ratio of 1.6 (95% CI: 1.3-2.0) for this association (60). Considering the median age at diagnosis is 70 years, comorbidities are quite likely (61).

In summary, delays in myeloma diagnosis can be explained by the non-specific nature of the symptoms, the rarity of the condition (annual incidence of 9 per 100 000 people) and the advanced age at which myeloma tends to occur. Overall there is evidence suggesting that reducing diagnostic delays in myeloma is important. Efforts should focus on early detection of symptomatic patients ideally before patients start experiencing multiple complications like irreversible renal failure or deterioration of their overall health status leading to presentation to the emergency services.

1.6.3 Cancer referral pathways and suspicion of cancer in primary the primary care setting

Current NICE guidelines recommend that patients who present with cancer related symptoms with a positive predictive value of 3% or above should be referred to secondary care through a suspected cancer referral pathway which aims for the patient to be seen within two weeks (62,63). Some examples of symptoms or combination of symptoms that achieve that threshold include haematuria which has a PPV of 3.9 % (95% CI: 3.5-4.6) for bladder cancer, a breast lump which has a PPV of 4.8 % (95% CI: 3.6-5.4) for breast cancer, and rectal bleeding combined with diarrhoea that has a PPV of 3.4 % (95% CI: 2.1-6.0) for colorectal cancer (64–66).

For myeloma, most of the symptoms experienced by patients (see section 1.5.1) are also very common in many other conditions meaning a symptom based approach is not as useful as in other cancers. For example, the most common symptom of myeloma is back pain– percentages vary from 30% to 60%; back pain has a PPV of 0.1% which means that only 1 in 1000 patients with back pain will have myeloma (67). Only when symptoms are combined with blood tests will positive predictive values go over the threshold of 3%, for example the combination of back pain and high calcium (26).

1.7 Use of clinical prediction rules to aid clinical decision making

One of the most common tools that are used in medicine to aid with the decision making process are clinical prediction rules (68). Prediction rules are mathematical equations which combine multiple factors in order to estimate the probability of disease being present or manifesting in some time in the future.

Properly developed clinical prediction rules may have certain advantages over human clinical decision making alone. One of the main advantages is that for the same set of factors a clinical prediction rule will give the same result while human judgement has been shown to be inconsistent (68). From a diagnostic point of view they can be very useful in situations where the disease is very rare and the clinicians have limited personal experience.

Currently there is a lack of clinical prediction rules for myeloma. To date, only one clinical prediction rule has been developed that could potentially be used for myeloma (69). A major limitation of this prediction rule was that the outcome used for the development of the model was abnormal serum/urine protein electrophoresis (SPE/UPE) rather than myeloma which limits its applicability. This is in stark contrast with the multiple prediction rules that have been developed for most cancers. For example a recent systematic review identified that for colorectal cancer alone, 52 prediction rules have been developed (70).

One of the main reasons for not creating prediction rules for myeloma is the rarity of the disease. When creating clinical prediction rules the optimal design a cohort study (71–73). But

to conduct a cohort study in low prevalence settings thousands or even millions of patients need to be recruited. The introduction of big databases like the CPRD allowed the development of prediction rules for rare conditions. For example a clinical prediction rule developed for pancreatic cancer using the Qresearch database had a sample of approximately 2.5 million patients of which 1415 patients had pancreatic cancer (74,75).

One possible explanation for the delays in diagnosing myeloma is that the disease is not suspected early enough and multiple consultations are required before myeloma is suspected as the cause of a patient's symptoms. A clinical prediction rule could prove to be useful, if it could be applied when myeloma is not necessarily suspected but some common tests are ordered like the full blood count or inflammatory markers which are very common in primary care and they are ordered when patients present with non-specific symptoms and for a variety of conditions (76).

1.8 Specific aims and objectives of the thesis

In the thesis I will answer the following questions:

1. How long are the delays in the diagnosis of myeloma and where exactly are they occurring?

This is an important question as it is still not very clear how long the time to diagnosis of myeloma is and how long the patients spend in each healthcare setting (primary or secondary). Answering this question will allow me to identify where in the clinical pathway a clinical prediction rule might be most effective.

2. Which symptoms and blood test abnormalities manifest early and which late during the myeloma disease progression

Currently there is limited evidence on the timing of symptoms and blood test abnormalities. Identifying the biomarkers that change the earliest during the disease progression can help to

determine what should be included in the clinical prediction rule. Tests or biomarkers that change very close to the diagnosis are unlikely to be useful.

3. Which test combinations are the most useful for ruling in/out the disease?

Most of the blood tests that are used in primary care for myeloma and that precede the electrophoresis tests have been examined separately but never in combination and not within a primary care setting. Comparing the diagnostic value of blood tests together will enable me to identify the most optimal combinations.

The results of questions 1-3 will inform the final and main objective of the thesis which is

4. Development and validation of clinical predictions rule for the diagnosis of myeloma

1.9 Studies and structure of the thesis

To answer the questions specified stated above in section 1.9, the thesis presents four results chapters in the order in which they were completed. The methods used in all chapters of the thesis was quantitative and included wide range of study designs like a systematic review, a nested case-control study and a retrospective cohort study.

Chapter 2 presents the results of a systematic review which aimed to quantify the length of the diagnostic intervals observed in myeloma patients from onset of first symptom to confirmation of diagnosis.

Chapter 3 presents the findings of the nested case-control study which answered questions 2 and 3 (section 1.9); mapping the disease progression for symptoms and blood tests and identification of the most optimal test combinations for myeloma. This study, in collaboration with the University of Exeter, was a secondary analysis of the case-control study using CPRD data that was used to examine the positive predictive value of symptoms in combination with blood tests (26). Information from this chapter will inform the design of the study presented in chapters 4 and 5 like for example selection of the most useful tests in terms of early diagnosis, length of follow up etc.

Chapter 4 presents a series of clinical prediction rules for the diagnosis of myeloma developed from a retrospective cohort study.

Chapter 5 is focused on the validation and examination of clinical utility of the clinical prediction rules developed in chapter 4 and the determination of the rule that is likely to be the most useful for diagnosing myeloma

Chapters 4 and 5 used a different CPRD data cut compared to the one in chapter 3.

1.10 Dissemination of results

The results in chapters 2 (systematic review) and 3 (case-control study) have been published, and as such, sections of the methods, results and discussions are reproduced (77,78). This work has not been submitted for any other course or qualification at the University of Oxford or elsewhere. The articles are available in full following the final chapter in appendices 1 and 2. In addition GPUpdate updated their guidelines for myeloma by including the results of the aforementioned study (appendix 3)

References

1. Röllig C, Knop S, Bornhäuser M. Multiple myeloma. Lancet. 2015;385(9983):2197–208.

2. Palumbo A, Anderson K. Multiple Myeloma. N Engl J Med. 2011;364(11):1046–60.

3. Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, Hayes RB, et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. Blood. 2009;113(22):5412–7.

4. Rajkumar SV, Kyle RA, Therneau TM, Clark RJ, Bradwell AR, Melton LJ, et al. Presence of monoclonal free light chains in the serum predicts risk of progression in monoclonal gammopathy of undetermined significance. Br J Haematol. 2004;127(3):308–10.

5. Kyle RA, Rajkumar SV, RAJKUMAR SVE. Management of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). Oncology (Williston Park). 2011;25(7):578–86.

6. Rajkumar SV, Landgren O. Smoldering multiple myeloma. 2015;125(20):3069–76.

7. Bird J, Behrens J, Westin J, Turesson I, Drayson M, Beetham R, et al. UK Myeloma Forum (UKMF) and Nordic Myeloma Study Group (NMSG): guidelines for the investigation of newly detected M-proteins and the management of monoclonal gammopathy of undetermined significance (MGUS). Br J Haematol. 2009;147(1):22–42.

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

9. Cancer registration statistics, England - Office for National Statistics [Internet]. [cited

2019 Jan 18]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/final2016

10. Palumbo A, Bringhen S, Ludwig H, Dimopoulos MA, Bladé J, Mateos M V., 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.

11. Vélez R, Turesson I, Landgren O, Kristinsson SY, Cuzick J. Incidence of multiple myeloma in Great Britain, Sweden, and Malmö, Sweden: the impact of differences in case ascertainment on observed incidence trends. BMJ Open. 2016;6(1):e009584.

12. Palumbo A, Cavallo F, Gay F, Di Raimondo F, Ben Yehuda D, Petrucci MT, et al. Autologous Transplantation and Maintenance Therapy in Multiple Myeloma. N Engl J Med. 2014;371(10):895–905.

13. Attal M, Harousseau J-L, Stoppa A-M, Sotto J-J, Fuzibet J-G, Rossi J-F, et al. A Prospective, Randomized Trial of Autologous Bone Marrow Transplantation and Chemotherapy in Multiple Myeloma. N Engl J Med. 1996;335(2):91–7.

14. Myeloma survival statistics | Cancer Research UK [Internet]. [cited 2018 Jul 9]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/survival#heading-Zero

15. Leukaemia (all subtypes combined) survival statistics | Cancer Research UK [Internet]. [cited 2018 Aug 15]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia/survival#heading-Zero

16. Hodgkin lymphoma survival statistics | Cancer Research UK [Internet]. [cited 2018 Aug 15]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/hodgkin-lymphoma/survival#heading-Zero

17. Landgren O, Gridley G, Turesson I, Caporaso NE, Goldin LR, Baris D, et al. Risk of monoclonal gammopathy of undetermined significance (MGUS) and subsequent multiple myeloma among African American and white veterans in the United States. Blood. 2006;107(3):904–6.

18. Frank C, Fallah M, Ji J, Sundquist J, Hemminki K. The population impact of familial cancer, a major cause of cancer. Int J Cancer. 2014;134(8):1899–906.

19. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet. 2008;371(9612):569–78.

20. Shiels MS, Cole SR, Kirk GD, Poole C. A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals. J Acquir Immune Defic Syndr. 2009;52(5):611–22.

21. Engels EA, Clarke CA, Pfeiffer RM, Lynch CF, Weisenburger DD, Gibson TM, et al. Plasma cell neoplasms in US solid organ transplant recipients. Am J Transplant. 2013;13(6):1523–32.

22. McShane CM, Murray LJ, Landgren O, O'Rorke MA, Korde N, Kunzmann AT, et al. Prior autoimmune disease and risk of monoclonal gammopathy of undetermined significance and multiple myeloma: a systematic review. Cancer Epidemiol Biomarkers Prev. 2014;23(2):332–42.

23. Brown KF, Rumgay H, Dunlop C, Ryan M, Quartly F, Cox A, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br J Cancer. 2018;118(8):1130–41.

24. Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. Br J Haematol. 2003;121(5):749–57.

25. Kariyawasan CC, Hughes D a., Jayatillake MM, Mehta a. B. Multiple myeloma: causes and consequences of delay in diagnosis. Qjm. 2007;100(10):635–40.

26. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. Br J Gen Pr. 2015;65(631):e106-13.

27. Dimopoulos MA, Kastritis E, Rosinol L, Bladé J, Ludwig H. Pathogenesis and treatment of renal failure in multiple myeloma. Leukemia. 2008;22(8):1485–93.

28. Wirk B. Renal failure in multiple myeloma: a medical emergency. Bone Marrow Transplant. 2011;46(6):771–83.

29. Augustson BM, Begum G, Dunn JA, Barth NJ, Davies F, Morgan G, et al. Early mortality after diagnosis of multiple myeloma: analysis of patients entered onto the United kingdom Medical Research Council trials between 1980 and 2002--Medical Research Council Adult Leukaemia Working Party. J Clin Oncol. 2005;23(36):9219–26.

30. Abbott KC, Agodoa LY. Multiple myeloma and light chain-associated nephropathy at end-stage renal disease in the United States: patient characteristics and survival. Clin Nephrol. 2001;56(3):207–10.

31. Oyajobi BO. Multiple myeloma/hypercalcemia. Arthritis Res Ther. 2007;9 Suppl 1(Suppl 1):S4.

32. Ha K-Y, Kim Y-H, Kim H-W. Multiple Myeloma and Epidural Spinal Cord Compression : Case Presentation and a Spine Surgeon's Perspective. J Korean Neurosurg Soc. 2013;54(2):151.

33. Smith D, Yong K. Multiple myeloma. BMJ. 2013;346:f3863–f3863.

34. Multiple myeloma - NICE CKS [Internet]. [cited 2018 Jul 9]. Available from: https://cks.nice.org.uk/multiple-myeloma#!diagnosissub

35. Watson J, Round A, Hamilton W. Raised inflammatory markers. BMJ. 2012;344:e454–e454.

36. Zhang J, Sun W, Huang Z, Chen S, Zhong Y, Hu Y, et al. Light chain multiple myeloma, clinic features, responses to therapy and survival in a long-term study. World J Surg Oncol. 2014;12:234.

37. Greipp PR. International Staging System for Multiple Myeloma. J Clin Oncol. 2005;23(15):3412–20.

38. Mai EK, Haas E-M, Lücke S, Löpprich M, Kunz C, Pritsch M, et al. A systematic classification of death causes in multiple myeloma. Blood Cancer J. 2018;8(3):30.

39. Harousseau J-L, Moreau P. Autologous Hematopoietic Stem-Cell Transplantation for Multiple Myeloma. N Engl J Med. 2009;360(25):2645–54.

40. Badros A, Barlogie B, Siegel E, Morris C, Desikan R, Zangari M, et al. Autologous stem cell transplantation in elderly multiple myeloma patients over the age of 70 years. Br J Haematol. 2001;114(3):600–7.

41. Kumar SK, Dingli D, Lacy MQ, Dispenzieri A, Hayman SR, Buadi FK, et al. Autologous stem cell transplantation in patients of 70 years and older with multiple myeloma: Results from a matched pair analysis. Am J Hematol. 2008;83(8):614–7.

42. Palumbo A, Rajkumar SV, San Miguel JF, Larocca A, Niesvizky R, Morgan G, et al. International Myeloma Working Group Consensus Statement for the Management, Treatment, and Supportive Care of Patients With Myeloma Not Eligible for Standard Autologous Stem-Cell Transplantation. J Clin Oncol. 2014;32(6):587–600.

43. 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.

44. Tosi P, Zamagni E, Ronconi S, Benni M, Motta M, Rizzi S, et al. Safety of autologous hematopoietic stem cell transplantation in patients with multiple myeloma and chronic renal failure. Leukemia. 2000;14(7):1310–3.

45. El Fakih R, Fox P, Popat U, Nieto Y, Shah N, Parmar S, et al. Autologous Hematopoietic Stem Cell Transplantation in Dialysis-Dependent Myeloma Patients. Clin Lymphoma Myeloma Leuk. 2015;15(8):472–6.

46. NHS England » Diagnosing cancer earlier and faster [Internet]. [cited 2018 Jul 10]. Available from: https://www.england.nhs.uk/cancer/early-diagnosis/

47. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. Br J Cancer. 2015;112 Suppl:S92-107.

48. Hollenbeck BK, Dunn RL, Ye Z, Hollingsworth JM, Skolarus TA, Kim SP, et al. Delays in diagnosis and bladder cancer mortality. Cancer. 2010;116(22):5235–42.

49. Gobbi PG, Bergonzi M, Comelli M, Villano L, Pozzoli D, Vanoli A, et al. The prognostic role of time to diagnosis and presenting symptoms in patients with pancreatic cancer. Cancer Epidemiol. 2013;37(2):186–90.

50. Mendonca SC, Abel GA, Saunders CL, Wardle J, Lyratzopoulos G. Pre-referral general practitioner consultations and subsequent experience of cancer care: evidence from the English Cancer Patient Experience Survey. Eur J Cancer Care (Engl). 2016;25(3):478–90.

51. Lyratzopoulos G, Wardle J, Rubin G. Rethinking diagnostic delay in cancer: how difficult is the diagnosis? BMJ. 2014;349:g7400–g7400.

52. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel G a. Variation in number of general practitioner consultations before hospital referral for cancer: Findings from the

2010 National Cancer Patient Experience Survey in England. Lancet Oncol. 2012;13(4):353–65.

53. Galinsky J, Richard S, Morris M. Myeloma patients' self-reported experiences of care and treatment. Cancer Nurs Pract. 2016;15(7):22–6.

54. Howell D, Smith A, Appleton S, Bagguley T, Macleod U, Cook G, et al. Multiple myeloma: routes to diagnosis, clinical characteristics and survival - findings from a UK population-based study. Br J Haematol. 2017;177(1):67–71.

55. Abel GA, Mendonca SC, McPhail S, Zhou Y, Elliss-Brookes L, Lyratzopoulos G. Emergency diagnosis of cancer and previous general practice consultations: insights from linked patient survey data. Br J Gen Pract. 2017;67(659):e377–87.

56. Lyratzopoulos G, Abel G a, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. Br J Cancer. 2013;108(3):686-90.

57. Din NU, Ukoumunne OC, Rubin G, Hamilton W, Carter B, Stapley S, et al. Age and Gender Variations in Cancer Diagnostic Intervals in 15 Cancers: Analysis of Data from the UK Clinical Practice Research Datalink. PLoS One. 2015;10(5):e0127717.

58. Tørring ML, Frydenberg M, Hansen RP, Olesen F, Vedsted P. Evidence of increasing mortality with longer diagnostic intervals for five common cancers: A cohort study in primary care. Eur J Cancer. 2013;49(9):2187–98.

59. Ong F, Hermans J, Noordijk EM, Wijermans PW, Kluin-Nelemans JC. Presenting signs and symptoms in multiple myeloma: High percentages of stage III among patients without apparent myeloma-associated symptoms. Ann Hematol. 1995;70(3):149–52.

60. Friese CR, Abel GA, Magazu LS, Neville BA, Richardson LC, Earle CC. Diagnostic delay and complications for older adults with multiple myeloma. Leuk Lymphoma.

2009;50(3):392–400.

61. Piccirillo JF, Vlahiotis A, Barrett LB, Flood KL, Spitznagel EL, Steyerberg EW, et al. The changing prevalence of comorbidity across the age spectrum. Crit Rev Oncol Hematol. 2008;67(2):124–32.

62. Barraclough K. New NICE guidance on referral for cancer. BMJ. 2015;351:h3640.

63. Suspected cancer: recognition and referral | Guidance and guidelines | NICE.

64. Hamilton W, Round A, Sharp D, Peters TJ. Clinical features of colorectal cancer before diagnosis: a population-based case–control study. Br J Cancer. 2005;93(4):399–405.

65. Walker S, Hyde C, Hamilton W. Risk of breast cancer in symptomatic women in primary care: a case–control study using electronic records. Br J Gen Pract. 2014;64(629):e788–93.

66. Shephard EA, Stapley S, Neal RD, Rose P, Walter FM, Hamilton WT. Clinical features of bladder cancer in primary care. Br J Gen Pract. 2012;62(602):598–604.

67. Kyle RA, Gertz MA, Witzig TE, Lust JA, Lacy MQ, Dispenzieri A, et al. Review of 1027 Patients With Newly Diagnosed Multiple Myeloma. Mayo Clin Proc. 2003;78(1):21–33.

68. Adams ST, Leveson SH. Clinical prediction rules. BMJ. 2012;344:d8312.

69. Thakkinstian A, Tran H, Reeves G, Murch S, Attia J. A clinical decision rule to aid ordering of serum and urine protein electrophoresis for case-finding of paraproteins in hospitalized inpatients. J Gen Intern Med. 2008;23(10):1688–92.

70. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk Prediction Models for Colorectal Cancer: A Systematic Review. Cancer Prev Res (Phila). 2016;9(1):13–26.

71. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al.

Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683–90.

72. Lee YH, Bang H, Kim DJ. How to Establish Clinical Prediction Models. Endocrinol Metab (Seoul, Korea). 2016;31(1):38–44.

73. Moons KGM, Altman DG, Reitsma JB, Ioannidis JP a, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162(1):W1–73.

74. Home - QResearch [Internet]. Available from: https://www.qresearch.org/

75. Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: Derivation and validation of an algorithm. Br J Gen Pract. 2012;62(594):38–45.

76. Busby J, Schroeder K, Woltersdorf W, Sterne JAC, West N, West S. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices : using routine data to identify research priorities. 2013; 63(609):e 256–66.

77. Koshiaris C, Oke J, Abel L, Nicholson BD, Ramasamy K, Van den Bruel A. Quantifying intervals to diagnosis in myeloma: a systematic review and meta-analysis. BMJ Open. 2018;8(6):e019758.

78. Koshiaris C, Van den Bruel A, Oke JL, Nicholson BD, Shephard E, Braddick M, et al. Early detection of multiple myeloma in primary care using blood tests: a case–control study in primary care. Br J Gen Pract. 2018;68(674):e586–93.

Chapter 2: Quantifying diagnostic intervals in myeloma: a systematic review and meta-analysis

2.1 Introduction

Diagnostic delays are associated with poorer outcomes in Myeloma. Reducing the time to diagnosis could result in better outcomes for patients (1,2). Myeloma diagnosis is not straightforward and it can vary from patient to patients in terms of how and where it takes place i.e. diagnosis through referral or emergency presentation or detected by incidental testing etc.

The diagnostic pathway is a term used to describe the different stages of cancer diagnosis, from the first occurrence of symptoms through to the confirmation of the disease by specialists. A number of different studies have identified some important milestones along the diagnostic pathway which most patients are expected to go through (3,4). These include the onset of the first symptom(s) which can be attributed to myeloma, first presentation to primary care with that particular symptom, first presentation to secondary care and final diagnosis.

Using these milestones the following intervals could be defined. Figure 1 illustrates the patients pathway based on these definitions

Patient interval: Time from onset of first symptom to first presentation to primary care.

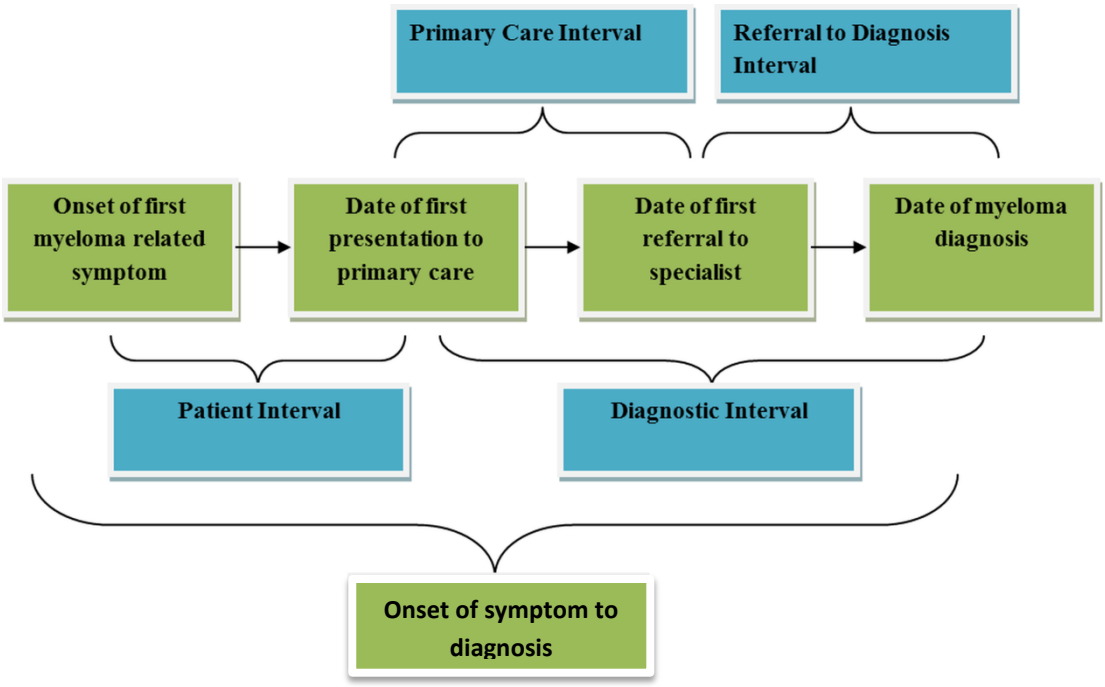
Primary care interval: Time from first presentation to primary care with a cancer related symptom to first referral to secondary care services.

Referral to diagnosis interval: Time from first referral to secondary care to confirmed myeloma diagnosis

Diagnostic interval: Time from fist presentation to primary care with a cancer related symptom to date of confirmed diagnosis.

Onset of symptom to diagnosis interval: Time from symptom onset to confirmation of diagnosis.

Figure 1: Patient pathway to diagnosis



Delays are possible at each of these intervals: patients may not attribute their symptoms to a potentially serious disease thus delaying their visit to the doctor, and clinicians in primary or secondary care may equally not suspect myeloma and spend time investigating other potential causes. Delays occurring after the patient present to the doctor are usually attributed to healthcare and are subject to improvement.

The aim of this chapter is to quantify the different intervals of the diagnostic pathway for multiple myeloma patients from symptom onset to confirmation of diagnosis.

2.2 Methods

To quantify the intervals in the diagnostic pathway, a systematic search of the literature was carried out and the evidence combined using meta-analysis. The aim was to quantify each of the five intervals in the myeloma diagnosis pathway.

2.2.1 Outcomes, Search strategy and inclusion/exclusion criteria

The outcomes used were the five intervals as defined in section 2.2.1 which includes: the patient, primary, referral to diagnosis, diagnostic and total intervals. Medline and EMBASE were searched up until January 2018 for articles estimating the length of any of the intervals described above. The search strategy was a modified version of the strategy used in another systematic review where they explored the association of time to diagnosis and different outcomes in all cancers (5). In addition citation searching of important references like the Aarhus statement was done and also search in the reference list of systematic reviews that were dealing with similar questions (3,5). Search strategy included the terms myeloma, time to diagnosis, time to referral, time to consultation and different variations of the above terms which could indicate a quantification of one of the pre-specified intervals. Additional terms included delay in diagnosis, delay in presentation etc. A copy of the search strategy can be seen in appendix 4. Articles that estimated the length of at least one of the intervals in days or months were included. Articles that reported numbers of consultations or referrals rather than the length of time were not included. Although number of consultations indicates opportunities for earlier diagnosis they do not allow us to directly estimate the length of a particular interval. Studies that focused on children or adolescents (< 18 years old) and on the asymptomatic form of the disease (monoclonal gammopathies or smouldering myeloma) were also excluded. Papers had to have an abstract in English to be selected for full text assessment but no full text articles were excluded based on language.

2.2.2 Data extraction

Data were extracted independently by two different reviewers in a pre-specified excel spreadsheet. Study characteristics included author, year of publication, country of data collection, type of study, myeloma related symptoms, and sample size. Information on the definitions of the different time points and intervals were also extracted. Statistics for the length of the different outcomes included the median, interquartile range, mean and standard deviation

(SD). Authors were contacted if the data were not reported in the main articles or if the data were not in the appropriate format (i.e. duration of intervals reported as a categorical variable rather than continuous).

2.2.3 Risk of bias tool

Risk of bias was assessed by two independent reviewers using the Aarhus statement checklist tool and disagreements were discussed with a third reviewer (3). The Aarhus statement is a tool designed to evaluate the risk of bias of studies that are quantifying onset of symptom to diagnosis in cancer. It consists of 20 items which evaluate the appropriateness of the different definitions the studies use to define the starting and ending points of each interval under investigation.

2.2.4 Analysis and synthesis of data

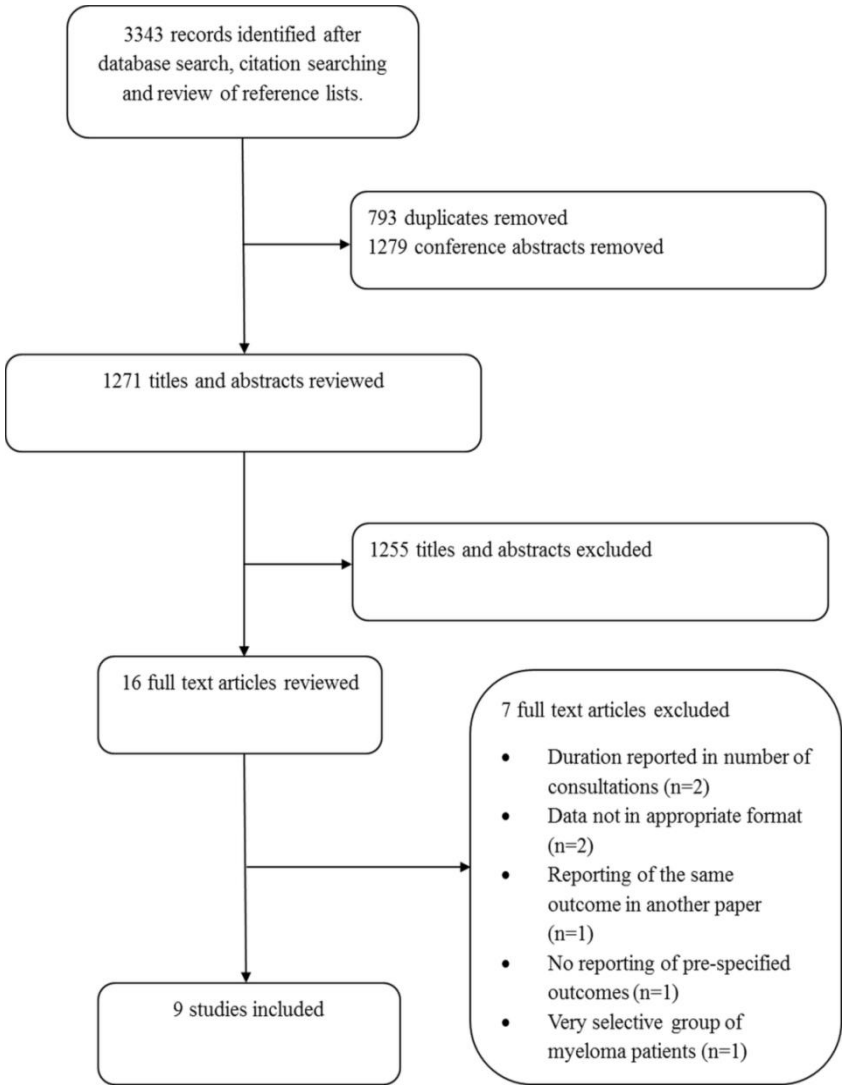
Because of the skewed nature of time intervals, especially the ones that are quantifying the duration of illness are unlikely to be normally distributed, the median and interquartile range (IQR) are better for summarising the data. This is evident by the fact that all studies were reporting the 25^{th}, 50^{th} and 75^{th} percentiles for the different intervals they were investigating. For intervals reported by more than one study, the pooled estimate was calculated by taking a weighted mean for each percentile. The weight was obtained by dividing the sample size in each study with the total number of patients. Sensitivity analysis was conducted by excluding the study with the highest risk of bias, to explore its effect on the overall estimate. A distribution was fitted through the three weighted estimates in order to better illustrate the shape of the different time intervals. The log-normal distribution was chosen as time intervals in cancer are usually skewed to the right (6). An additional sensitivity analysis was conducted by meta-analysing the means and standard deviations by using the established methods. Transformations of the median and IQR to mean and standard deviation were used when these were not reported (7). Additional sensitivity analysis included exclusion of the studies with the highest risk of bias.

2.3 Results

The literature search identified 3343 citations from the EMBASE/MEDLINE searches, citation and reference list searching. After removal of duplicates and conference abstracts, 1271 citations were left for screening. Based on titles and abstracts, 1255 abstracts were excluded. 16 full text articles were screened and 9 were included in the review. Figure 2 illustrates study selection flowchart. Some studies were excluded based on certain reasonable criteria that were developed during the process of the review in addition to the pre-specified criteria. The post-hoc criteria included papers that were using the same databases and outcomes, investigation of intervals that were not pre-specified and were not based on theoretical frameworks (i.e. Aarhus statement) and papers that were reporting only selective myeloma patients (i.e. patients presenting with only a specific symptom or complication).

In total 7 full text articles were excluded and the main exclusion reasons include the following. Two studies reported the duration of the interval in number of consultations (8,9). One study did not report any of the pre-specified outcomes (10). Three papers were making use of the same database and in particular the National Audit of Cancer diagnosis in Primary Care 2009-2010 (11). One of these three papers was excluded as it was investigating the same outcome as one of the other two papers so inclusion of this paper would results in entering the same patients twice in the analysis(12). Two papers were not reporting the interval in the appropriate format i.e. categorical data rather than continuous (1,13). One paper was reporting the interval for a very specific group of patients, in particular patients that experience renal impairment making it a very selective population compared to the other studies (14).

Figure 2: Study flowchart



2.3.1 Study characteristics

The studies took place between 2009 and 2018 and the sample size varied from 107 to 3831 patients. Five studies reported intervals in various cancers(15–19) ; three reported only myeloma(20–22); and one for haematological malignancies(23). Six studies were conducted in the UK(15–19,23), one in the USA(20), one in Hungary(21) and one in Israel(22). Two UK studies used data from two separate Clinical Practice Research Datalink (CPRD) cohorts, a database of routinely collected electronic primary care records (15,16). Two other UK studies used data from the English National Audit of Cancer Diagnosis in Primary Care 2009-2010(11): both of these studies were reporting the primary care interval and one of those two the patient

interval. The primary care interval was extracted from the study (table 1) that had the bigger sample size (17,18). Another UK study used data from the English National Audit of Cancer Diagnosis in Primary care 2014 and the last UK study was a patient survey on patients diagnosed with haematological malignancies(19,23). The study conducted in Hungary analysed data collected from patients treated in a haematology centre and the study conducted in the USA analysed a retrospective database collected from the Surveillance, Epidemiology, and End Results Program (SEER) program which is responsible for collecting cancer incidence and survival data from cancer registries and covers approximately 28% of the US population (20,21). The study conducted in Israel used data from the Israeli health maintenance (HMO) organisation linked with the Israel National Cancer Registry (NCR) (22). Detailed information on the study characteristics can be seen in table 1.

Table 1: Study characteristics

Study design	Study period	Population characteristics (age, gender)	Myeloma patients (Total)	Outcome measure (Interval)
Friese et al 2009, (USA) (20)				
Retrospective analysis	1992-2002	mean age 76.3 46% males	3831	Diagnostic
Howell et al 2013, (UK) (23)				
Survey	2004-2011	Median age 69.9 66.9% males	341	Patient Diagnostic Onset of symptom to diagnosis
Lyratzopoulos et al 2013, (UK) (17)				
Audit data	2009-2010	Not reported	176	Primary Care
Varga et al 2014, (Hungary) (21)				
Retrospective analysis	Not reported	median age 60 50% males	193	Diagnostic
Neal et al 2014, (UK) (16)				
Retrospective analysis	2001-2002	mean age 72 53% males	221	Diagnostic
Din et al 2015, (UK) (15)				
Retrospective analysis	2007-2010	median age 72 56% males	500	Diagnostic
Lyratzopoulos et al 2015, (UK) (18)				
Audit data	2009-2010	Not reported	124	Patient
Goldscbmidt et al 2016, (Israel) (22)				
Retrospective analysis	2002-2011	median age 63 53% males	107*	Diagnostic
Swann et al 2017, (UK) (19)				
Audit data	2014	Not reported	202	Primary Diagnostic

*The total sample size for this study was 110 patients out of which 7 were diagnosed with plasmacytoma. The analysis was conducted on 107 patients that had complete data

2.3.2 Definitions of intervals and symptoms

Table 2 lists the symptoms that were included as possibly related to myeloma in each study and the definitions for the time points and intervals used in each study. In total, 19 different symptoms were reported with significant variability in which and how many symptoms each study was using. Studies reported 3 to 12 symptoms although some studies were including multiple symptoms under more general categories. The CPRD studies for example included multiple musculoskeletal symptoms under a more general bone pain category, while Howell et al. used a general pain category that included abdominal, chest and other types of pain. Three studies did not report the symptoms used, quite likely due to them reporting intervals for multiple cancers rather than just focusing on myeloma.

Myeloma-related symptoms were used in all studies to define illness onset or first presentation to primary care. In three studies, the first occurrence of a myeloma-related symptom was limited to a year before diagnosis, in three studies to two years before diagnosis and in one study to three years before diagnosis. One study used patient recorded dates to identify the onset of first symptom and date of first visit to a GP with a myeloma related symptom. Goldschmidt N et al (22) did not use the first symptom as the start of the measurement period but they used the first combination of symptom and laboratory result (i.e. the earliest of blood test + pain complaint or two blood tests within a month or two pain complaints within 1-3 months).

Table 2: Symptoms and interval definitions

Symptoms used	Onset of first symptom	Date of first Presentation in healthcare services	Date of first referral	Date of diagnosis
Friese et al 2009, (USA)				
Anaemia Packed red blood cell transfusion Back pain	N/A	Up to1 year before Diagnosis	N/A	SEER cancer diagnosis date
Howell et al 2013, (UK)				
Tiredness Pain Shortness of breath Infections Joint problems/Fractures Stomach/bowel symptoms Other	Patient reported	Patient reported	N/A	Date provided by the Haematological Malignancy Diagnostic service
Lyratzopoulos et al 2013, (UK)				
Not reported	Estimated based on patient's clinical records.	Up to 2 years before Diagnosis	Date that the referral letter was sent	Clinical records and hospital correspondence
Varga et al 2014, (Hungary)				
Bone symptoms Anaemia Renal failure General symptoms Other Tumour presence Metastatic bone disease	N/A	Up to 3 years before Diagnosis	N/A	Tertiary haematology centre
Neal et al 2014, (UK)				
Bleeding Bone Pain Bruising Anaemia Fatigue Anorexia Weight loss	N/A	Up to 1 year before Diagnosis	N/A	First occurrence of a myeloma Read Code in the patient's record in CPRD database.
Din et al 2015, (UK)				
Bleeding Bone Pain Bruising Anaemia Fatigue Anorexia Weight loss	N/A	Up to 1 year before Diagnosis	N/A	First occurrence of a myeloma Read Code in the patient's record in CPRD database.
Lyratzopoulos et al 2015, (UK)				
Not reported	N/A	Up to 2 years before Diagnosis	Date that the referral letter was sent	Clinical records and hospital correspondence
Goldscbmidt et al 2016, (Israel)				
Pain (back, cervical spine, musculoskeletal, non-specific) Infection	N/A	Up to 2 years before diagnosis	N/A	Israel National Cancer Registry

Weight loss Fatigue Peripheral edema Constipation Presyncope Syncope Dizziness				
Swann et al 2017, (UK)				
Not reported	NA	Up to 2 years before diagnosis	Date that the referral letter was sent	Hospital Episode Statistics (HES)

2.3.3 Risk of bias assessment

Overall there was low risk of bias for most of the studies (Appendices 5 and 6). All studies clearly defined what the start and finish of the measured intervals were and how they collected the information for each time point. Most studies discussed the limitations and the biases affecting the collected data especially around the collected dates and in what way and to which extent it would affect the analysis and the generalisability of the results. Most common sources of bias mentioned included misclassification bias for studies using medical records like CPRD or SEER (15,16,20) and recall bias for studies using patent interviews to determine the start and finish of the measuring period as well as symptoms (23).

Misclassification bias tends to occur when individuals are assigned to a different group than the one they should be in. For example the starting point for these studies was defined as the first myeloma related symptom based on pre-defined symptom lists. It is possible thought that the recorded symptom was not caused by cancer but another condition or patients were presenting with symptoms not included in the pre-defined symptom lists. This is similar with cancer diagnosis. It is possible that some cancers might have been under-recorded which can potentially lead to exclusion of certain cases from the sample. All the above can also affect the measurement of the different intervals under investigation. The effect of this misclassification in diagnosis thought is expected to be small especially in databases like CPRD which has been shown that accuracy of the diagnostic codes is approximately 95% (24)

Recall bias usually refers to errors caused by participants not remembering accurately previous events i.e. the date of first symptom, date they first visited GP with that symptom etc. Patients might be reporting symptoms that are bothering them but are not necessarily related to myeloma or they might not remember the exact date the symptoms started which can lead to the overestimation or underestimation of the intervals under investigation. This type of bias might even be more severe in the case of myeloma if we take into consideration the advanced age of myeloma patients, the non-specific nature of the symptoms and the multiple comorbidities they might be suffering from. In this case it is reasonable to assume that this may have affected the estimate of the intervals in the survey due to patient recall.

Few studies provided an adequate discussion about the definition of the date of diagnosis which was the category with the highest risk of bias. According to the Aarhus statement there should be a hierarchy for defining the date of diagnosis in terms of the how the diagnosis was done (date of first histological confirmation of cancer, date of hospital admission due to cancer, date of death etc.). For example, is the date of diagnosis the day that the patient the date of the biopsy or the date of the pathology report on the biopsy? The recorded date of diagnosis can affect the estimation of the length of the interval. Although all the studies were clearly defining where the date of diagnosis was taken from it was not very clear what the date of diagnosis was corresponding to, thus the risk of bias was high for this domain.

2.3.4 Estimation of the length of intervals

Seven papers reported the diagnostic interval (15,16,19–23), two the patient interval(18,23), two the primary care interval(17,19), and one the onset of symptom to diagnosis (23). No studies reported the referral to diagnosis interval. In addition the primary care, patient and onset of symptom to diagnosis intervals were reported only by UK studies while the majority of studies included for the diagnostic interval were also from the UK (4 out of 7 studies). The length of the different intervals can be seen in table 3 and the fitted log normal distributions in figure 3 along with the parameters used to fit them.

For the diagnostic interval, the pooled weighted mean of the 50th percentile is 108.6 days (n=5398) and the IQR is from 33.3 to 241.7 (n=5288). Removing the study with the largest risk of bias(21) based on the Aarhus statement checklist did not alter the results (107.9 days, IQR: 31.3-242.2). An additional sensitivity analysis was conducted by excluding the Goldschmidt N et al (22) study as it was an outlier but the results were not affected (pooled median of 103.8). While all the studies reported a median diagnostic interval less than 5 months this study had an interval of 11.2 months. The IQR was not estimated for this sensitivity analysis as it was not reported by the authors.

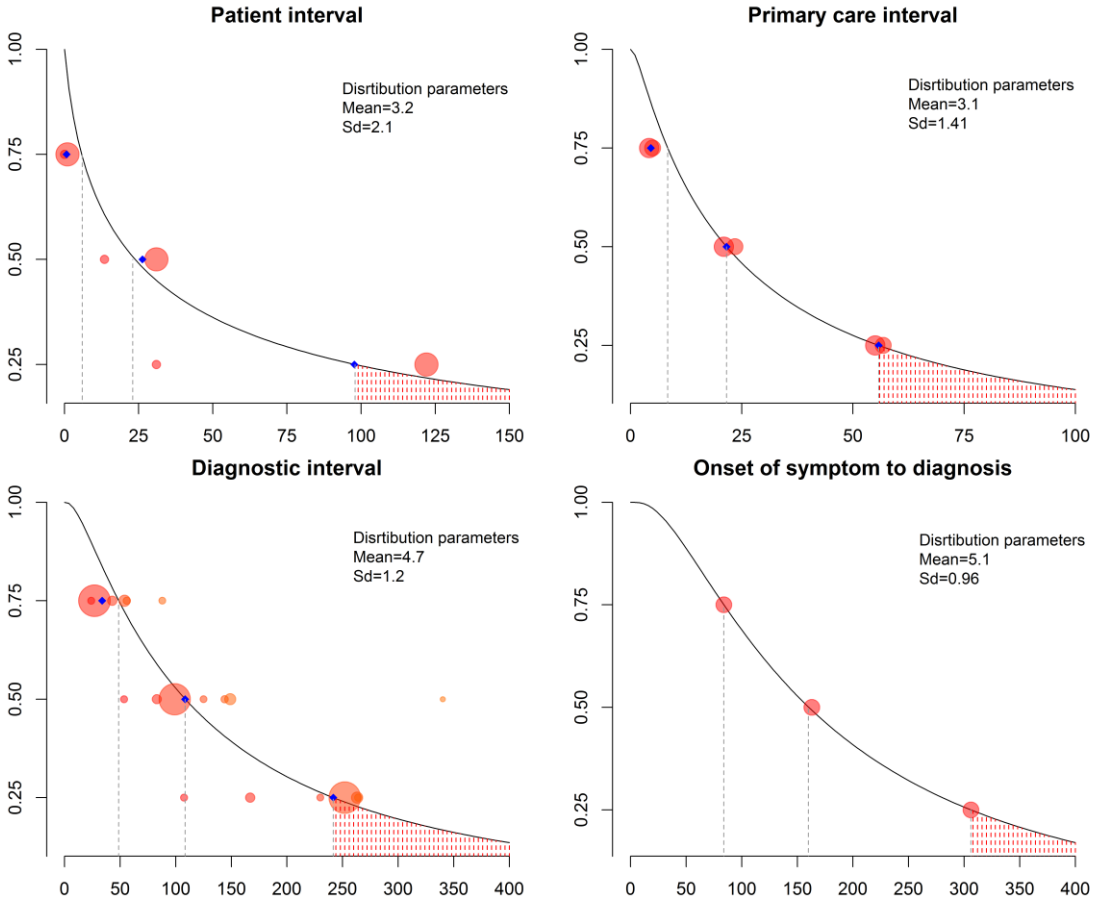
The pooled estimate of the 50th percentile of the patient interval is 26.3 days (IQR: 0.7-97.7, n=465). The primary care interval was reported by two studies (17,19) with a median of 21.6 days (IQR: 4.6-55.8, n=326) and the onset of symptom to diagnosis was also reported by one study with a median of 163 days (IQR: 84-306, n=341) (23).

No study reported the referral to diagnosis interval but it can be inferred by subtracting the median length of the primary care and patient intervals from the diagnostic interval or the onset of symptom to diagnosis which suggests that the median length of the referral to diagnosis interval can range from 60.7 to 115.1 days depending on whether we use the diagnostic or the onset of symptom to diagnosis for the inference.

Table 3: Length of intervals

Percentile	N	25th	50th	75th
Patient Interval				
Howell et al 2013	341	1	31	122
Lyratzopoulos et al 2015	124	0	13.5	31
Weighted estimate	465	0.7	26.3	97.7
Primary care Interval				
Lyratzopoulos et al 2013	176	5	21	55
Swann et al 2017	150	4.2	23.5	56.8
Weighted estimate	326	4.6	21.6	55.8
Referral to Diagnosis interval				
No papers reporting this interval				
Diagnostic interval				
Friese et al 2009	3831	27	99	252
Howell et al 2013	341	34	83	167
Varga et al 2014	193	88	125	230
Neal et al, 2014	221	56	144	264
Din et al 2015	500	54	149	263
Goldschmidt et al 2016	107	NR*	341	NR*
Swann et al 2017	202	24	53.5	107.5
Weighted estimate	5395	33.3	108.6	241.7
Onset of symptom to diagnosis interval				
Howell et al, 2013	341	84	163	306
*Not reported				

Figure 3: Distribution of intervals. Each circle corresponds to one study and the size is proportional to the total sample size. The blue diamond corresponds to the weighted estimate. For intervals with only one study (onset of symptom to diagnosis) no weighted estimates were calculated. Y-axis corresponds to 1-percentile. For example the value of 0.25 corresponds to the 75th percentile and 0.75 to the 25th percentile.



2.4 Discussion

2.4.1 Summary of findings

The results confirm that myeloma patients tend to experience long diagnostic intervals before they are diagnosed. The median length of the patient interval is approximately one month; for 25% of patients, the time is longer than three months (98 days). The median length of the diagnostic interval is approximately three months (109 days) and up to eight months for 25% of patients. Symptoms reported in the different studies are mainly bone-related pain especially back pain, myeloma defining events such as anaemia or reduced renal function, and general signs of malignant disease such as fatigue and weight loss. The referral to diagnosis interval was estimated to be longer than the primary care. The reasons that might explain this finding are of course speculative, but could include time spent on the wrong organ-specific investigations in secondary care aimed at confirming a specific diagnosis. It should be noted that the referral to diagnosis interval was estimated based on the length of the other intervals as no studies were quantifying it thus the quality of evidence for this interval is very low. Any inferences for the referral to diagnosis interval should be interpreted with caution. Variability was observed for the definitions of the different intervals, time points and symptoms.

2.4.2 Strengths and weaknesses

This is the first systematic review that quantified the patient pathway to diagnosis of myeloma from onset of first symptom until confirmation of diagnosis. The search strategy used did not have any restrictions in terms of study design or health care systems. The focus of the search strategy was on two medical databases that were the most likely to contain papers about the length of the diagnostic process. In order to compliment the databases search I searched in the reference list of similar systematic reviews and key references. Grey literature (i.e. conference abstracts) was not included although several were identified that were addressing the review question. The reason they were not included is the fact that the length of the different intervals is affected by design decisions including the start of the measurement period, the initial symptoms used, data collection methods etc. These are not usually included in detail in

conference abstracts (if mentioned in the first place) and thus they were not included in the review.

One of the main strengths of the review is that the focus was not just on the measures of central tendency like the mean and the median but also on measures that summarize the tails of the distribution like the 25^{th} and 75^{th} percentiles. This is of importance as time interval data tend to be right skewed so patients at the end of these types of distributions would be experiencing much longer delays than what the median would indicate so probably these patients would benefit the most from an intervention that aims to shorten time to diagnosis.

A potential weakness to this analysis is that there are, at present no agreed methods for pooling medians and quartiles which also does not allow us to estimate heterogeneity statistics as information on the confidence intervals around the reported quartiles is not available. A sensitivity analysis was done by meta-analysing the means rather the quartiles. This also required to transform the median and IQR into mean and standard deviation as not all papers reported them. I decided to use this approach as the sensitivity rather than the primary analysis method for the following reasons:

1. All the information about the tails of the distribution would have been lost by making use of the mean and standard deviation. As explained previously, patients at the tails of the distribution experience the longest delays and are more likely to experience the worst outcomes. This skewness of the time intervals becomes very apparent by how well the log-normal distribution fits the through the three pooled estimates.
2. The mean tends to overestimate measures of central tendency when used on right skewed distributions and it would make comparison with other cancer intervals or myeloma intervals reported in future studies difficult as most of them are quantified using the median rather than the mean.
3. Three studies did not report the mean and standard deviation for the different intervals so they had to be approximated by making use of the median and IQR which was

reported. This approximation thought does not seem to work very well with skewed distributions which also became evident after the transformations were applied on studies for which all the measures were reported. The results can also be affected depending on the type of transformation that we use (3 scenarios which vary according to the data reported in the paper).

4. The analysis in this review is closer to a fixed effects meta-analysis which means that more weight is assigned on the bigger study. Specifically in this dataset the length of the diagnostic interval might be underestimated but this does not change the clinical message of the paper that myeloma patients experience delays of more than 3 months.

The sensitivity analysis based on the means and standard deviations (appendix 7) showed that the mean diagnostic interval is 156 days compared to a 109 days when using the medians. The difference could be potentially explained by some of the reasons described above but the clinical message of the review does not change which is that myeloma patients experience long diagnostic intervals.

It would also be of interest to investigate the differences in the diagnostic interval between the NICE qualifying and NICE-non-qualifying symptoms since most of the studies retrieved were UK based. Unfortunately this was not possible as amongst the UK studies only 3 report the diagnostic interval and two used the same database with same symptoms.

I did not conduct a subgroup analysis by country or healthcare system as the majority of studies included in the systematic review are from the UK and in certain cases some intervals were quantified only by UK studies (primary, patient and onset of symptom to diagnosis). While the results greatly generalise to the UK population this is not the case for other countries, especially for ones that use different healthcare systems. Caution should be exercised when extrapolating the results of this review to different healthcare systems in different countries.

In addition the referral to diagnosis interval was estimated based on the length of the other intervals as no studies were reporting it. Any inferences made with respect to this interval should be interpreted with caution.

2.4.3 Sources of Heterogeneity

As mentioned in the above section no formal methods of estimating heterogeneity exist when pooling medians or quartiles. In the sensitivity analysis (appendix 7), using only studies that reported means and standard deviations, the heterogeneity in the diagnostic interval was 98.6% (diagnostic interval). Similar values were observed for the patient interval. No heterogeneity was observed for the primary care interval which was expected since the two studies reporting the primary care interval came from the National Audit of Cancer Diagnosis in Primary care in 2010 and 2014 thus exactly the same design and data collection methods were used. For the patient and diagnostic intervals heterogeneity was expected due to the various design decisions that are going to be discussed later but the I^2 in this case should still be interpreted with caution as it might be an overestimation. This might be the case due to the small amount of uncertainty for each within-study estimate which results in very narrow confidence intervals. Also for certain studies the mean, standard deviations and confidence intervals had to be approximated which could result in the introduction of bias when estimating the I^2.

In order to accurately estimate the length of the different intervals, the definition of the beginning and the end of the interval is crucial. There was variability in how studies defined starting points, especially for the first symptom and the first presentation to healthcare.

Another important factor that might affect the length of the measured intervals is the time used to detect myeloma related symptoms prior to diagnosis. In the studies included this time spanned from one to three years. Although a lot of the included studies restricted the intervals to one year before diagnosis, there is some evidence to suggest that cancer symptoms might be present for more than one year which may have led to an underestimation of intervals in these studies (25). On the other hand, the more you extend the symptom period the more likely you

are to detect symptoms that are unrelated to myeloma and overestimate the length of the intervals. In order to explore this, a sensitivity analysis was conducted where the length of the diagnostic interval was stratified according to the time used to define the presenting symptoms (one year back vs. three years back). Studies that identified the symptom at 1 year before diagnosis had a median diagnostic interval of 105 days compared more than 1 year before diagnosis which had a median of 142.3 days. This could explain some of the observed variability but the difference is small and it is mainly driven by the Goldschmidt study which reported a diagnostic interval of approximately one year. Overall it seems that one of the main sources of heterogeneity comes from studies using different definitions of the pre-diagnostic period. The studies in this review used pre-diagnostic intervals from one to three years but most limited the pre-diagnostic period to one year. It is likely that one year might not be enough and to what extent this leads to biased estimates of diagnostic intervals is something that should be investigated in future research.

As described there are various sources of heterogeneity between the studies but none of the sensitivity analyses conducted changed the trend of the results or the clinical interpretation as 70% of the included studies (5 out of 7) reported an interval longer than three months.

2.4.4 Comparison of findings with existing literature

Our estimate for the patient interval is in line with the findings of Howell et al. (13) which reported that 15% of myeloma patients delay their presentation more than three months before they go to the doctor after the onset of a myeloma related symptom. This was a study that was excluded due to outcomes not reported in a numerical format. This study explained this delay in terms of patients' lack of understanding of the seriousness of their symptoms because of their non-specific nature.

The diagnostic interval (the time from first presentation to referral) is longer for myeloma than for many other cancers. Cancers that have more specific symptoms and are easier to suspect tend to have shorter intervals. For breast cancer the median diagnostic interval for symptomatic

patients is 14 days (26), testicular with a median diagnostic interval of 44 days and melanoma with a median interval of 32 days (16,19). Other cancers with a similarly long median diagnostic interval include lung cancer which has a diagnostic interval of 88 days and leukaemia with a median of 102 days (15,26).

2.4.5 Implications for future research and clinical practice

Reducing the time to diagnosis is important because longer intervals are associated with more advanced disease and worsening of patients' quality of life (27,28); the National Audit of Cancer diagnosis in Primary Care (2010-2011) reported that 22% of myeloma patients had metastatic disease at diagnosis compared to 8.7% of breast cancer patients. Myeloma patients expressed greater levels of dissatisfaction compared with other cancer patients and felt that their health had got worse while waiting to see a specialist (28).

The primary care interval may be influenced by time spent waiting for the results of further investigations prior to referral (29) . Although investigations such as blood tests may prolong the primary care interval, the diagnostic interval could still be shorter if it these investigation result in referral to the appropriate specialist (haematology), reducing the secondary care interval by avoiding an inappropriate referral which potentially can take more time than the time it takes for a blood result to come back. A systematic review on colorectal cancer showed that delay in diagnosis was mostly associated with initial misdiagnosis, inadequate examination and inaccurate investigations (30). Interventions targeted on improving the above could results in significant reductions in diagnostic delays.

Although I was not able to estimate the referral to diagnosis interval directly, it is reasonable to believe that it is longer than the primary care interval. As in other cancers (31), suboptimal referrals, i.e. to the wrong specialty or with an insufficient level of urgency, prolong the secondary care interval. The choice of referral route has been shown to be a strong predictor of the length of the diagnostic interval (32). Future studies should not only estimate the duration of primary and referral to diagnosis intervals, but also investigate the impact of one setting on the

other. In addition a recent qualitative study of 20 patients identified that only 25% of myeloma get a first referral to a haematologist which supports to some extent the argument made above (33). Due to the fact that no studies were identified to estimate the referral to diagnosis interval the quality of evidence regarding this interval is very low. This highlights a big gap in the literature when exploring the link between the primary and secondary care settings. Any inferences and conclusions drawn should be interpreted with caution as the generalisability of the findings regarding this interval is very limited and likely to be biased. More studies are required to properly estimate the duration of this interval and how is it linked with the primary care interval.

2.5 Conclusion and implications for next chapter

Some important conclusions arise from this systematic review. Firstly it is confirmed that myeloma patients experience long diagnostic intervals up to more than eight months for certain patients thus there is room for significant reductions in time to diagnosis. The review also confirmed that the primary care interval is long and that there is some evidence that GP's recognise a problem but don't necessarily suspect myeloma. This initial works also suggests that the referral to diagnosis interval may be longer than the primary care interval which might be affected by the type and urgency of referral. This cannot be said with certainty due to the lack of studies quantifying the referral to diagnosis interval and lack of studies that properly examine the type of referral that patients get after their first presentation with a myeloma related symptom. Some of these questions which were not definitively answered by the systematic review will be explored even further in the third chapter of the thesis. One of the key messages in terms of the overall thesis aims is the longest waiting times seem to be occurring within the healthcare pathway after the patients visit their doctor suggesting that a prediction rule can potentially have a positive impact in the reduction of delays.

References

1. Kariyawasan CC, Hughes D a., Jayatillake MM, Mehta a. B. Multiple myeloma: causes and consequences of delay in diagnosis. Qjm. 2007;100(10):635–40.

2. Ong F, Hermans J, Noordijk EM, Wijermans PW, Kluin-Nelemans JC. Presenting signs and symptoms in multiple myeloma: High percentages of stage III among patients without apparent myeloma-associated symptoms. Ann Hematol. 1995;70(3):149–52.

3. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. Br J Cancer. 2012;106(7):1262–7.

4. Olesen F, Hansen RP, Vedsted P. Delay in diagnosis: the experience in Denmark. Br J Cancer. 2009;101 Suppl(S2):S5–8.

5. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. Br J Cancer. 2015;112 Suppl:S92-107.

6. Lyratzopoulos G, Wardle J, Rubin G. Rethinking diagnostic delay in cancer: how difficult is the diagnosis? BMJ. 2014;349:g7400–g7400.

7. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14(1):135.

8. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. Br J Gen Pr. 2015;65(631):e106-13.

9. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel G a. Variation in number of general practitioner consultations before hospital referral for cancer: Findings from the

2010 National Cancer Patient Experience Survey in England. Lancet Oncol. 2012;13(4):353–65.

10. Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. Br J Cancer. 2012;107(8):1220–6.

11. Rubin G, Elliott K, McPhail S (2011) National Audit of Cancer Diagnosis in Primary Care. Royal College of General Practitioners: London. http://www.rcgp.org.uk/policy/rcgp-policy-areas/national-audit-of-cancer- diagnosis-in-primary-care.aspx

12. Keeble S, Abel GA, Saunders CL, McPhail S, Walter FM, Neal RD, et al. Variation in promptness of presentation among 10,297 patients subsequently diagnosed with one of 18 cancers: Evidence from a National Audit of Cancer Diagnosis in Primary Care. Int J Cancer. 2014;135(5):1220–8.

13. Howell DA, Warburton F, Ramirez A-J, Roman E, Smith AG, Forbes LJL. Risk factors and time to symptomatic presentation in leukaemia, lymphoma and myeloma. Br J Cancer. 2015;113(7):1114–20.

14. Li N, Lv Y, Zeng H, He F, Yao Y, He X. Renal impairment in multiple myeloma: Presenting features in different departments. J Huazhong Univ Sci Technol Medical Sci. 2012;32(1):65–8.

15. Din NU, Ukoumunne OC, Rubin G, Hamilton W, Carter B, Stapley S, et al. Age and Gender Variations in Cancer Diagnostic Intervals in 15 Cancers: Analysis of Data from the UK Clinical Practice Research Datalink. PLoS One. 2015;10(5):e0127717.

16. Neal RD, Din NU, Hamilton W, Ukoumunne OC, Carter B, Stapley S, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. Br J Cancer.

2014;110(3):584–92.

17. Lyratzopoulos G, Abel G a, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. Br J Cancer. 2013;108(3):686-90.

18. Lyratzopoulos G, Saunders CL, Abel GA, McPhail S, Neal RD, Wardle J, et al. The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. Br J Cancer. 2015;112 Suppl(s1):S35-40.

19. Swann R, McPhail S, Witt J, Shand B, Abel GA, Hiom S, et al. Diagnosing cancer in primary care: results from the National Cancer Diagnosis Audit. Br J Gen Pract. 2018;68(666):e63–72.

20. Friese CR, Abel G a., Magazu LS, Neville B a., Richardson LC, Earle CC. Diagnostic delay and complications for older adults with multiple myeloma. Leuk Lymphoma. 2009;50(3):392–400.

21. Varga G, Mikala G, Andrikovics H, Masszi T. How long does a myeloma patient currently wait for the diagnosis in Hungary?. Orv Hetil. 2014;155(39):1538–43.

22. Goldschmidt N, Zamir L, Poperno A, Kahan NR, Paltiel O. Presenting Signs of Multiple Myeloma and the Effect of Diagnostic Delay on the Prognosis. J Am Board Fam Med. 2016;29(6):702–9.

23. Howell DA, Smith AG, Jack A, Patmore R, Macleod U, Mironska E, et al. Time-to-diagnosis and symptoms of myeloma, lymphomas and leukaemias: a report from the Haematological Malignancy Research Network. BMC Hematol. 2013;13(1):9.

24. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract. 2010;60(572):e128-36.

25. Corner J, Hopkinson J, Fitzsimmons D, Barclay S, Muers M. Is late diagnosis of lung

cancer inevitable? Interview study of patients' recollections of symptoms before diagnosis. Thorax. 2005; 60(4):314–9.

26. Redaniel MT, Martin RM, Ridd MJ, Wade J, Jeffreys M. Diagnostic intervals and its association with breast, prostate, lung and colorectal cancer survival in England: Historical cohort study using the Clinical Practice Research Datalink. PLoS One. 2015;10(5):1–17.

27. Howell D, Smith A, Appleton S, Bagguley T, Macleod U, Cook G, et al. Multiple myeloma: routes to diagnosis, clinical characteristics and survival - findings from a UK population-based study. Br J Haematol. 2017;177(1):67–71.

28. Galinsky J, Richard S, Morris M. Myeloma patients' self-reported experiences of care and treatment. Cancer Nurs Pract. 2016;15(7):22–6.

29. Rubin GP, Saunders CL, Abel GA, Mcphail S, Lyratzopoulos G, Neal RD. Impact of investigations in general practice on timeliness of referral for patients subsequently diagnosed with cancer : analysis of national primary care audit data. Br J Cancer. 2015;112(4):676–87.

30. Mitchell E, Macdonald S, Campbell NC, Weller D, Macleod U. Influences on pre-hospital delay in the diagnosis of colorectal cancer: a systematic review. Br J Cancer. 2008;98(1):60–70.

31. Barrett J, Hamilton W. Pathways to the diagnosis of lung cancer in the UK: a cohort study. BMC Fam Pract. 2008;9:31.

32. Jensen H, Tørring ML, Olesen F, Overgaard J, Vedsted P. Cancer suspicion in general practice, urgent referral and time to diagnosis: a population-based GP survey and registry study. BMC Cancer. 2014; 14:636.

33. Howell DA, Hart RI, Smith AG, Macleod U, Patmore R, Cook G, et al. Myeloma:

Patient accounts of their pathways to diagnosis. Harris F, editor. PLoS One. 2018 ;13(4):e0194788.

Chapter 3: Identifying opportunities for early diagnosis of myeloma: a case-control study based on Electronic health care records (CPRD)

3.1 Chapter Introduction

The main aims of this chapter is to identify early symptoms and biomarkers of myeloma, in particular the investigation of which blood test combinations can be used as a red flag for myeloma when they are abnormal or rule it out when normal. The other main aim is to identify which potential diagnostic markers manifest early and which late with respect to time to diagnosis and thus distinguish which ones can be used to expedite diagnosis. It can also help explain why some of them can better rule out myeloma compared to others. Features that manifest early in the diagnosis can be useful for ruling out the disease while features that manifest late and suddenly can be useful at ruling in the disease.

In chapter 2, I conducted a systematic review of the studies reporting diagnostic intervals for patients with myeloma. I estimated the primary and referral to diagnosis intervals, the interval across primary and secondary care, and the total time to diagnosis. One of the findings of the systematic review suggested that the referral to diagnosis interval is longer than the primary care interval. However caution should be exercised when interpreting this finding as no studies were identified to quantify the referral to diagnosis interval, thus it had to be inferred from the other intervals. This highlights the strong limitation of the literature to estimate the referral to diagnosis interval. For this reason a secondary aim of this chapter is to further explore and expand the findings of the systematic review with an emphasis on the referral to diagnosis interval.

The work in this chapter extends on a case control study conducted by the University of Exeter using CPRD data in which the risk of myeloma symptoms in primary care was quantified (1). The study identified eleven symptoms and five investigations associated with myeloma within the year before diagnosis (back pain, chest pain, chest infections, shortness of breath, nausea, fracture, Joint pain, bone pain, weight loss, rib pain, nosebleeds, cytopenia, raised inflammatory

markers, raised creatinine, raised mean corpuscular volume, hypercalcaemia). The predictive ability of dual symptom combinations and the combination of single symptom plus blood test investigation was examined but blood test combinations were not examined. In addition the inflammatory markers were grouped together as a single predictor indicating abnormalities in any inflammatory marker so it is not clear which one is the best and should be used which is something that I am going to explore in this chapter.

3.2 Research questions

The three main research questions of this chapter are the following

1. Distinguish between early and late features of the disease (symptoms and blood test investigations) going back up to five years before diagnosis
2. Identify which blood tests and blood test combinations are the best for ruling in/out myeloma (C-reactive protein, Erythrocyte sedimentation rate, plasma viscosity, full blood count, calcium, and creatinine).
3. A more in-depth quantification of the diagnostic pathway of myeloma patients where the primary care, referral to diagnosis and diagnostic intervals will be explored. The primary care interval will be stratified by presenting symptoms and referrals will be furthered explored in order to examine how multiple referrals, type of referral and department can affect the length of the referral to diagnosis interval.

3.3 Methods

3.3.1 Study design

The study had a case control design and it used patient records from CPRD (Clinical Practice Research Datalink) which is a large anonymised database that contains data collected from primary care. 2,703 cases and 12,157 controls were matched based on age, gender and general practice. Cases were selected if they were over 40 years old with a myeloma diagnosis between January 2000 and December 2009. Date of diagnosis was defined as the first myeloma Read

code which was also used as the index-date for their matched control. Exclusion criteria included cases or controls with less than one year of records, cases without controls, controls with myeloma and controls that did not seek medical care after registration.

3.3.2 Variable definitions

All symptoms and investigations were identified using libraries of Read codes and can be seen in the appendix 8. The following are the definitions of the variables that have been used throughout the analysis.

Date definitions:

Date of first presentation: Date of first READ code corresponding to a myeloma related symptom within the year before diagnosis

Date of first referral: Date of first READ code corresponding to a referral after date of first presentation

Date of diagnosis: Defined as the first myeloma READ code in the database.

Intervals

Diagnostic intervals are defined as per the systematic review in chapter 2 and follow those given in the Aarhus statement (2).

Primary care interval: Time from first myeloma related symptom to first referral.

Referral to diagnosis: Time from first referral after initial presentation to diagnosis

Diagnostic interval: Time from first myeloma related symptom to diagnosis.

First referral

Referral type: Defined as referral to a haematologist or other

Referral urgency: Defined as 1) 2 week wait 2) Routine 3) Immediately 4) Urgent

Number of referrals: Classified as one, two, more than two

Diagnostic parameters

Symptoms: back pain, chest pain, chest infections, shortness of breath, nausea, fracture, Joint pain, bone pain, weight loss, rib pain, nosebleeds,

Laboratory investigations: Cytopenia (low white cell count, low haemoglobin or low platelets), raised inflammatory markers, raised creatinine, raised mean corpuscular volume, hypercalcaemia

Abnormal blood test results were defined as results above or below the local laboratory's normal range. Implausible test values were excluded. Combinations of blood tests, such as FBC and calcium, were defined as positive if at least one of the test results was abnormal. For the combination to be defined as negative, all tests in the combination were required to be normal. For the FBC, which has multiple constituent parts (e.g. WCC, haemoglobin, platelets), a result was defined abnormal if at least one of the parts was indicating cytopenia (low haemoglobin, WCC or platelets).

3.3.3 Statistical analysis

Analysis for aim 1: Distinguishing between early and late features

To distinguish between the early and late features of myeloma I compared the occurrence of symptoms and blood test results in myeloma cases with matched controls in the five-years before diagnosis. Time before diagnosis was broken down into three-month intervals (90 days), and symptom occurrence and blood test results in these periods were identified. For each three month interval I calculated the percentage of cases and controls reporting symptoms of back pain, chest pain, chest infections, shortness of breath, nausea, fracture, Joint pain, bone pain, weight loss, rib pain and nosebleeds. Patients with multiple symptoms were handled in the following ways:

1. Same symptom within the same three month interval: Patients were coded as symptom positive if they had at least one consultation with that symptom within the same three month interval. This means that patients that were coded as symptom positive could have multiple occurrences of the same symptom in the same interval.
2. Same symptom at different three month intervals: Each three month interval included all the patients that were experiencing a particular symptom irrespectively of whether the same patients were also appearing at a different three month interval. This means that patients could appear in multiple three month intervals.
3. Multiple different symptoms at different three month intervals: Patients with 2 or more myeloma related symptoms occurring at different times or in the same period were retained. E.g. A patient with symptoms of back pain and fracture at the same visit would appear in both the back-pain and fracture analyses, similarly if these symptoms presented at different times they would also contribute to both analyses.

For the same symptom and time period patients appear only once to avoid double counting but they could appear in different time periods as this presents opportunities for earlier diagnosis. The analysis was not limited to patients that had only one symptom as this could over-represent the symptoms that appear early.

The results for individual symptoms are presented as grouped bar charts showing the percentage of cases and controls with symptoms over time. An additional plot was also generated in which the percentage of patients with two or more different symptoms in the same three month interval is presented.

Same principles were used for patients with blood tests. For each three-month interval, I calculated the mean (with 95% confidence intervals) ESR, CRP etc. for both myeloma cases and controls. For patients who had multiple occurrences of the same blood test within the same three month period the mean value amongst all tests was estimated. Mean blood test results are shown as "trajectories" over time with 95% confidence intervals.

Analysis for aim 2: Establish rule in and rule out properties of blood tests and combinations.

To determine which tests or combinations of tests can be used to rule in or rule out myeloma the positive and negative likelihood ratios were used. Positive likelihood ratios indicate how many times more likely a positive test occurs in subjects with the disease compared to subjects without the disease and negative likelihood tests represent how much less likely a negative result will occur in a subject with the disease compared to a healthy subject. The positive likelihood ratio was calculated by using the sensitivity and specificity values for each test or test combination (LR^+= sens/(1-spec) and LR^-=(1-sens)/spec). The different estimates of sensitivity and specificity for the different test and test combinations were estimated. Sensitivity measures the proportion of positives that are correctly identified and specificity measures the proportions of negatives that are correctly identified. To identify the best combinations for ruling out myeloma, the negative likelihood ratio was estimated for each test individually and in combination. Following definitions used in other studies I defined a useful rule in test as one with a LR+ of over 5 and a useful rule out test one with a LR- less than 0.2 which has been also been used in other diagnostic accuracy studies and they are usually considered moderate to strong evidence for ruling in or ruling out the disease (3,4). A likelihood ratio of 0.2 usually corresponds to a reduction in probability from pre-test to post-test of approximately 30% which is based on an approximation (change in probability≈0.19 * logLR) (5). This approximation does not work well for diseases with prevalence more than 90% or less than 10% which is the case for myeloma so in order to help with their interpretation I illustrate an example of how to transition from pre-test to post –test probabilities depending on the likelihood ratio in the results section.

Each combination included a different inflammatory marker as it is quite unlikely for a GP to order more than one at the same time. In addition to the measures mentioned above, each inflammatory marker was also examined by using conditional logistic regression in order to establish their association with myeloma. The constituent parts of the FBC were examined

either separately or all together. The analysis was restricted to patients who had been tested (or had combination of tests).

Analysis for aim 3: Estimation of intervals to diagnosis

For patients that had only one symptom the date of that symptom was used as the date of the first consultation with a myeloma related symptom. Similarly for patients who had multiple consultations with the same symptom the date of the first consultation was selected as the starting point. For patients who had multiple different symptoms, the first myeloma related symptom was identified and used as the symptom onset date. For example if a patient had back pain and fracture symptoms in their clinical record (both myeloma related) the one that manifested first was used as the first symptom. The distributions of the different intervals were summarised using median interquartile range and 90th percentile, due to time variables not being normally distributed. The length of the primary care interval was stratified by initial presentation and the referral to diagnosis was stratified by referral type, referral urgency and number of referrals. The following hypotheses were examined: 1) is there a difference in the length of the primary care interval between the different presenting symptoms? 2) is there a difference in the length of the referral to diagnosis interval between the different referral types? 3) is there a difference in the length of the referral to diagnosis interval by referral urgency? and 4) is there a difference in the length of the referral to diagnosis interval according to number of referrals? The above hypotheses were examined by the use of the non-parametric Kruskal-Wallis test with p.value<0.05.

3.4 Results

The median age in the whole sample was 73 years old with 53% being male. 1591 (59%) cases were identified to have one of the eleven symptoms in the year before diagnosis of which 1074 (68%) patients were identified to have a referral to secondary care after their first presentation with a symptom. Amongst those referred 549 (51%) were males and 915 (85%) were over the

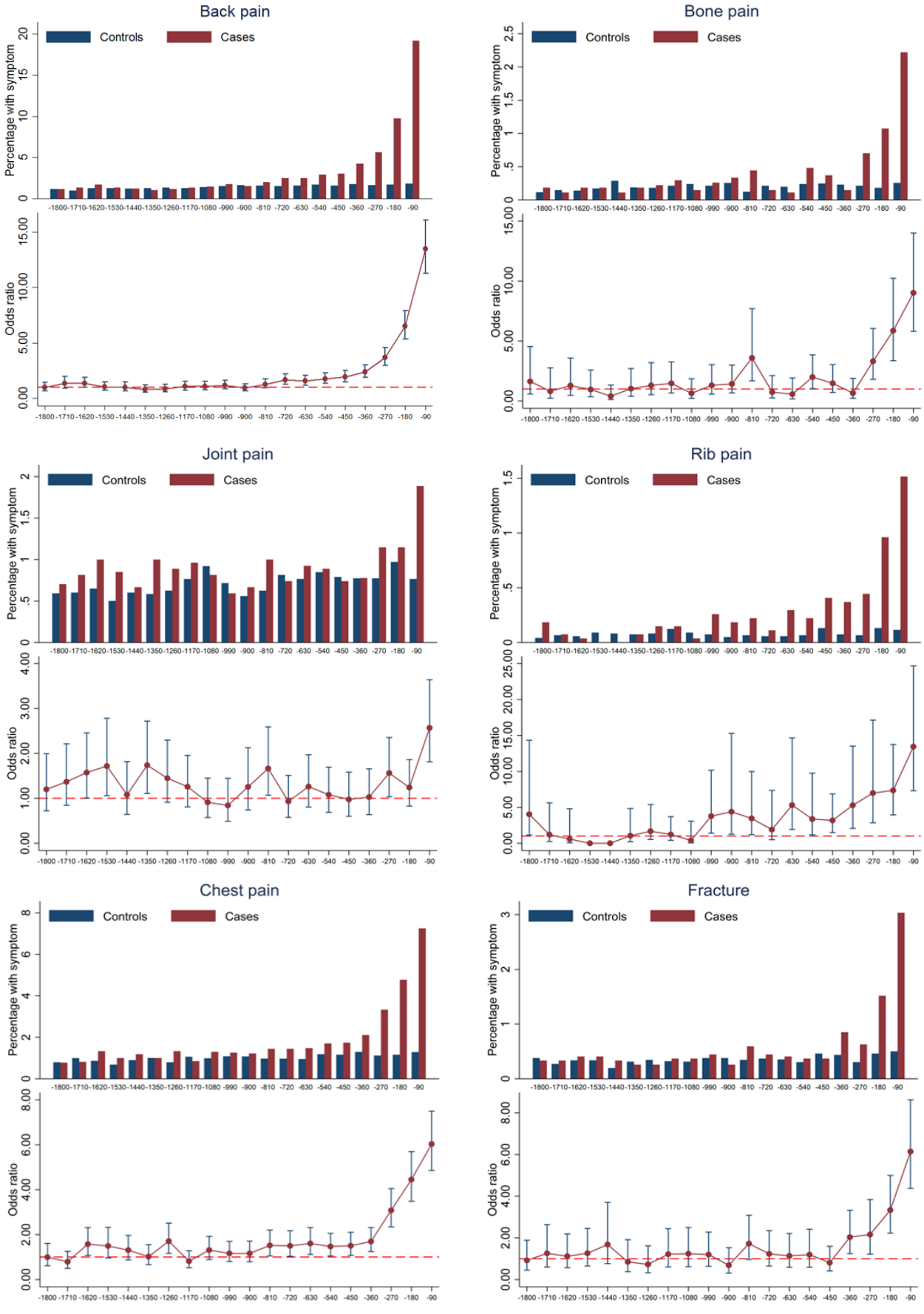
age of 60. Within the year before diagnosis myeloma patients had more consultations compared to controls (median of 16 compared to 8).

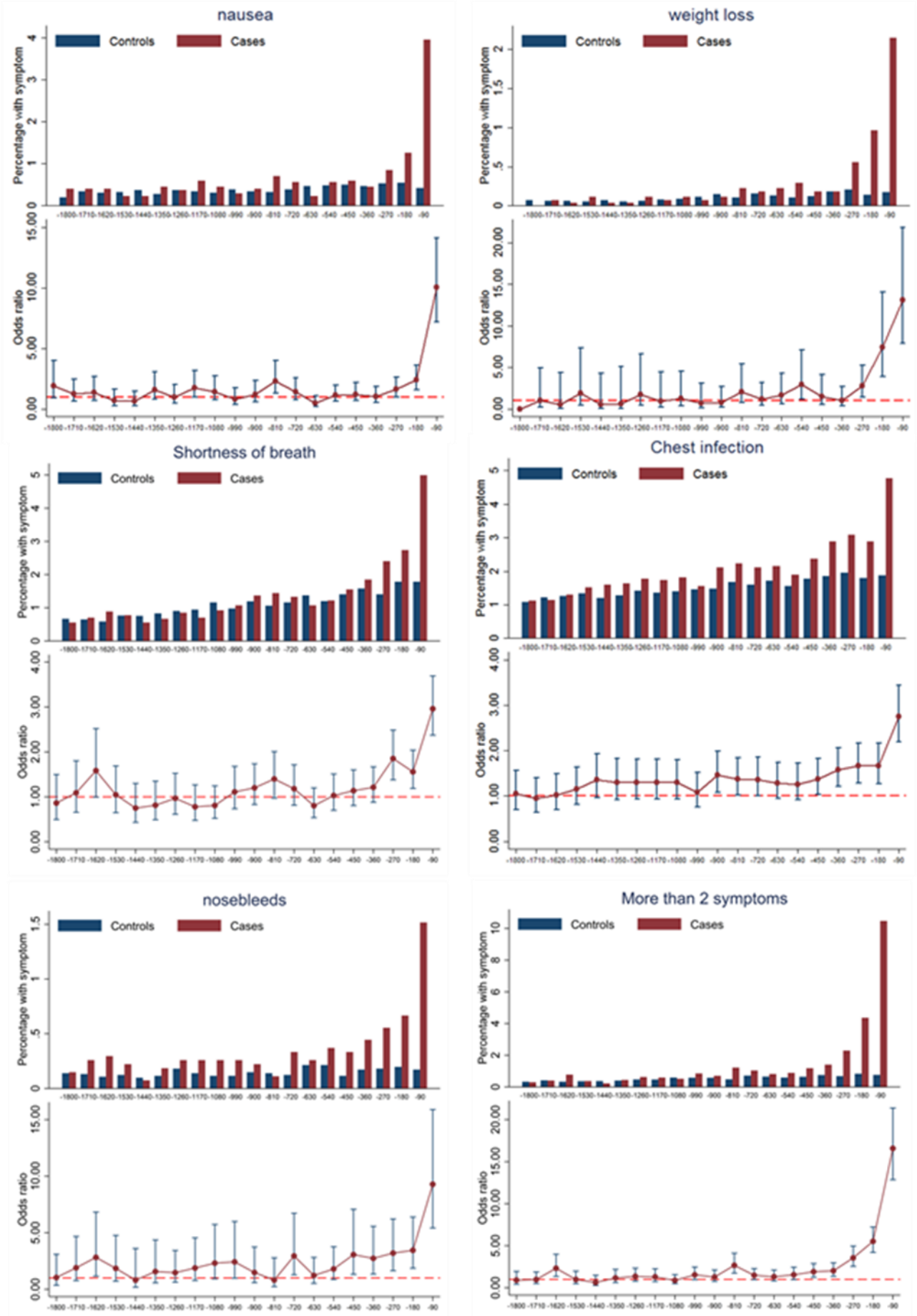
3.4.1 Timings of symptoms and investigations

Figure 1 illustrates the symptom prevalence trajectory for both cases and controls up to 5 years before diagnosis. Symptom occurrence for the controls is stable at all time points while the symptom prevalence for cases follows an upward trend. The time point at which the OR for a symptom becomes significant and remains significant was defined as the time at which myeloma starts to manifest. Symptoms of back pain, rib pain, chest infections, chest pain and nosebleeds, become more common in myeloma patients compared to controls approximately at 2 years before diagnosis although their prevalence is quite low at that time.

Symptoms suggesting advanced disease include fractures, weight loss and nausea, are more common in cases compared to controls up to one year before diagnosis, especially within the last 3 months. In addition approximately 10% of myeloma patients (compared to 1% of controls) were experiencing more than one different symptom within the last three months before diagnosis. Myeloma patients were presenting with the same symptom in more three month intervals compared to controls especially for back and chest pain symptoms. For example within the year before diagnosis 30% of myeloma patients had a consultation for back pain in at least two different three month intervals compared to 14% of controls. Similarly for chest pain 19% of myeloma patients had a consultation for chest pain in at least two different three month intervals compared to 13% of controls. For the rest of the symptoms very few patients had consultations for the same symptom in different three month intervals (less than 10%).

Figure 1: Symptom trajectories up to 5 years before diagnosis. The x-axis corresponds to time before diagnosis in 3month intervals.




nausea
Controls
Cases
Percentage with symptom
Odds ratio
weight loss
Controls
Cases
Percentage with symptom
Odds ratio
Shortness of breath
Controls
Cases
Percentage with symptom
Odds ratio
Chest infection
Controls
Cases
Percentage with symptom
Odds ratio
nosebleeds
Controls
Cases
Percentage with symptom
Odds ratio
More than 2 symptoms
Controls
Cases
Percentage with symptom
Odds ratio
Time in three month intervals

Figures 2, 3 and 4 illustrate the mean values for blood tests for each three month period in the five years leading up to the date of myeloma diagnosis. In myeloma patients, but not in control patients, haemoglobin values start to decrease approximately three years before diagnosis in both males and females and the mean haemoglobin value meets the World Health Organization (WHO) criteria for anaemia throughout the final 24 months in males (13 g/dL) and final 15 months for females (12 g/dL). Mean corpuscular volume (MCV) starts to rise approximately one and a half years before diagnosis. White cell count (WCC) and platelets remain stable within the normal range at all time points although at 3 months before diagnosis the mean is lower for cases compared to controls. The mean calcium and creatinine values rise rapidly within the year before diagnosis although both tests reach the upper abnormal thresholds only in the last three months (cal 2.6 and creat 120). ESR and PV values start to increase at approximately 2 years before diagnosis while CRP showed no difference between cases and controls at any point.

Figure 2: Inflammatory marker trajectories up to five years before diagnosis

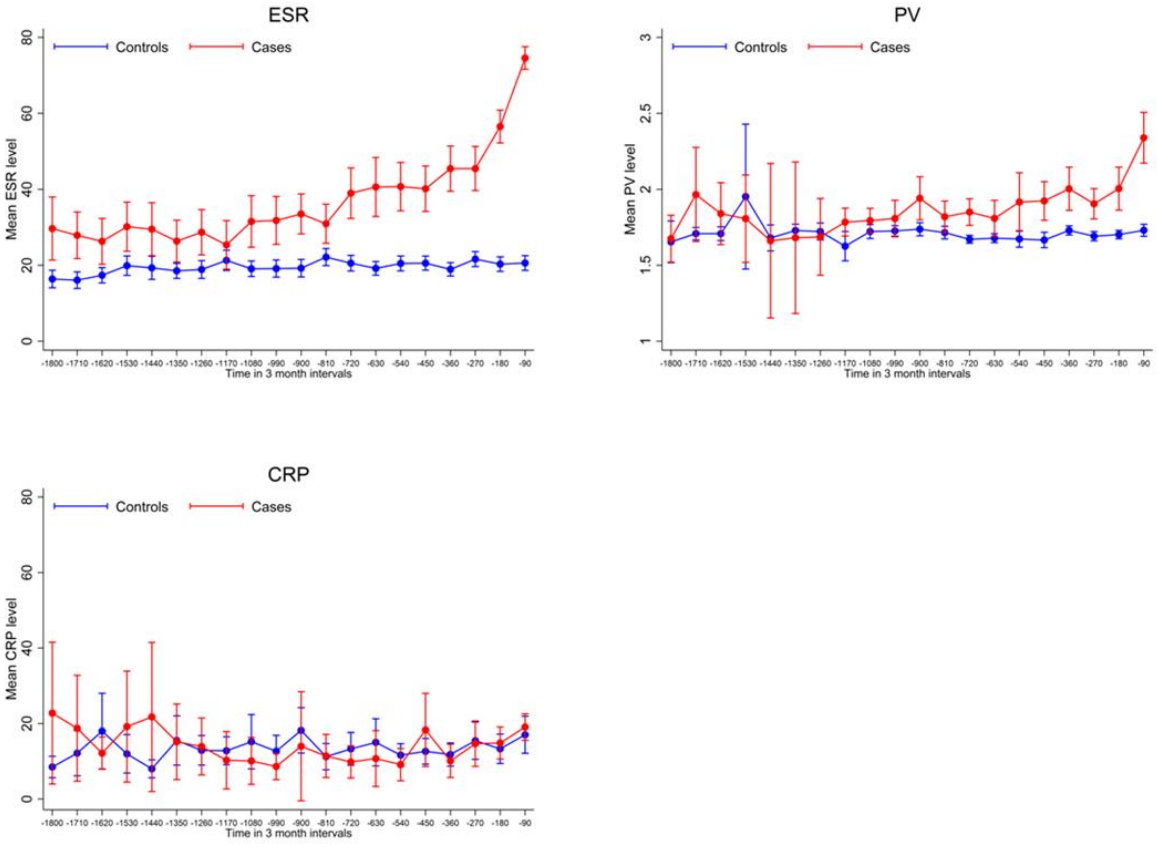


Figure 3: Full blood count component trajectories up to 5 years before diagnosis

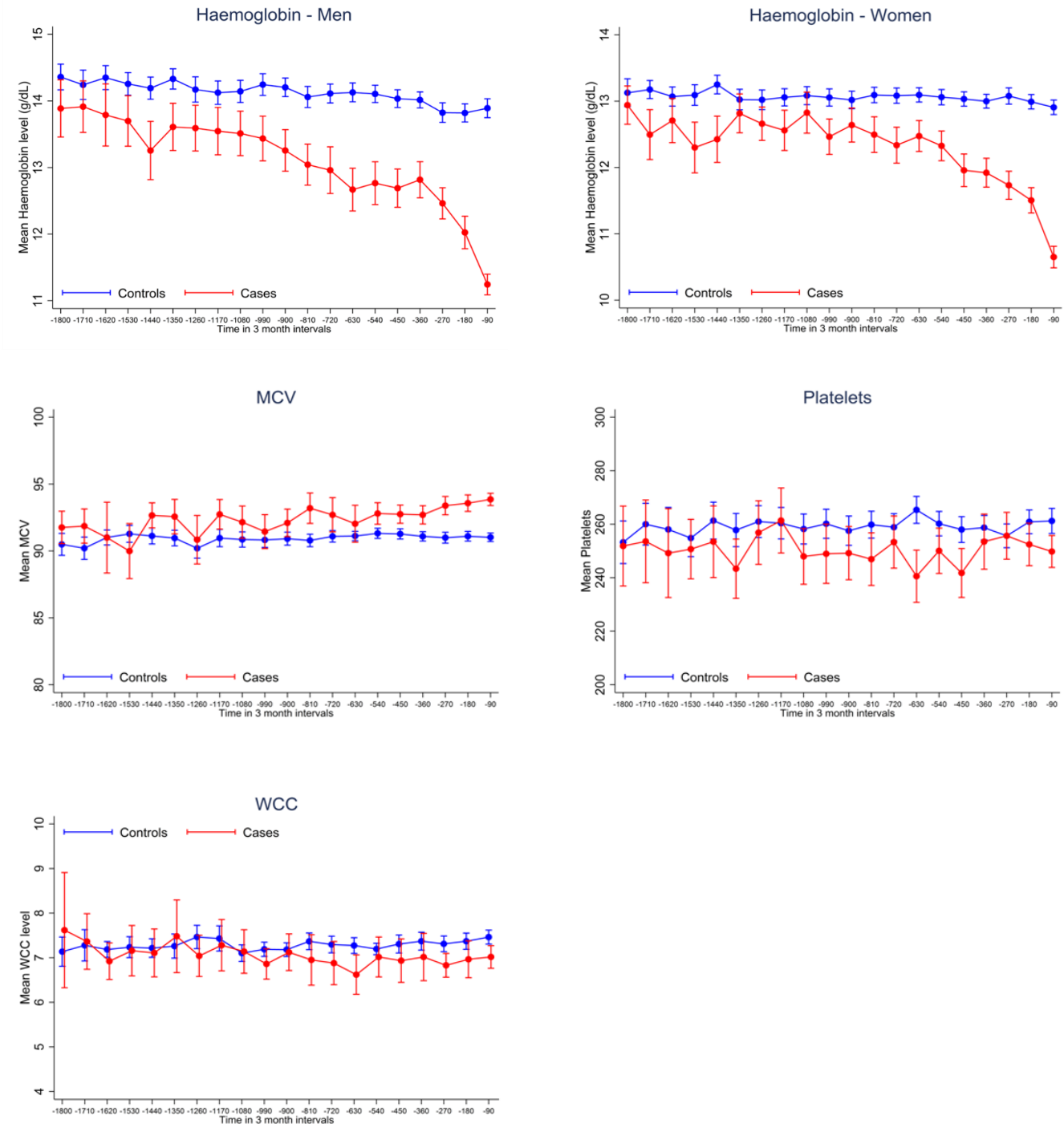
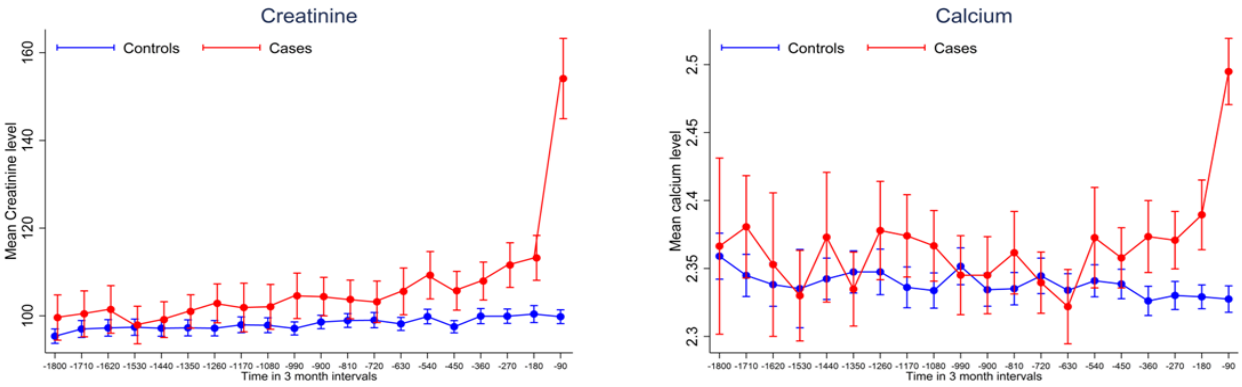


Figure 4: Creatinine and calcium



3.4.2 Rule in and rule out tests

Within the year before diagnosis 8221 (55%) participants had at least one blood test (either FBC, calcium, creatinine or one of the inflammatory markers) with testing being more prevalent in cases compared to controls (78% in cases and 50% in controls). The most common tests were creatinine and the FBC (40% and 50% respectively); the least common test was calcium (19%). Amongst the inflammatory markers the most common test was the ESR, followed by CRP and plasma viscosity. The most common three test combination was FBC, creatinine and ESR (n=1860, 13%). At three years before diagnosis 84% of myeloma patients had at least one blood test compared to 66% in the controls and at 5 years before diagnosis 85% of the cases had a test compared to 70% of the controls.

Associations between raised inflammatory markers and a myeloma diagnosis are displayed in table 1. During the year before diagnosis 85% of myeloma patients had an abnormal ESR compared to 46% of controls (OR: 5.7 - 95% CI: 4.1-8.1; p<0.0001), 81% had an abnormal PV compared to 41% of controls (OR: 4.4 – 95% CI: 2.2-8.8; p<0.0001) and 46% had an abnormal CRP compared to 37% of controls (OR: 1.5 – 95% CI: 1.0-2.2; p=0.05). The rule in properties for the rest of the tests which were published in the original report included cytopenia (OR: 5.4 – 95% CI: 4.6-6.4), raised creatinine (OR: 1.8 – 95% CI: 1.5-2.2), raised MCV (OR: 3.1 – 95% CI: 2.1-4.1) and hypercalcaemia (OR: 11.4 – 95%: 7.1 to 18) (9).

Table 1: Comparison of inflammatory markers

Test	Number Tested	Cases		Controls		OR (95% CI)	LR+ (95% CI)
		Abnormal (%)	Normal (%)	Abnormal (%)	Normal (%)		
ESR	2192	844 (85)	154 (15)	546 (46)	648 (54)	5.7 (4.1-8.0; p<0.0001)	1.9 (1.7-2.0)
CRP	1274	294 (46)	341 (54)	238 (37)	401 (63)	1.5 (1.0-2.2; p=0.05)	1.2 (1.1-1.4)
PV	433	162 (81)	39 (19%)	95 (41)	137 (59)	4.4 (2.2-8.8; p<0.0001)	2.0 (1.7-2.3)

No single blood test is able to rule out myeloma (Table 2), although confidence intervals for each marker approach our pre-specified cut-off of LR of 0.2. A normal haemoglobin performs much better as a rule out test (LR-: 0.42, 95% CI: 0.39-0.45) compared to calcium (LR-: 0.81, 95% CI: 0.78-0.83) and creatinine (LR-: 0.80, 95% CI: 0.77-0.83). For inflammatory markers, the best ones for ruling out the disease are the ESR (LR-: 0.28, 95% CI: 0.24-0.33) and PV (LR-: 0.32, 95% CI: 0.24-0.44). Ruling out is achieved by most of the test combinations, with the lowest LR- being when haemoglobin, calcium and PV are all normal (LR-: 0.06, 95% CI: 0.02-0.18). When adding creatinine to the 3-test combinations it did not change the negative likelihood ratio suggesting that it does not add much value to the diagnostic performance of the combination. Replacing creatinine with calcium in the three test combinations results in very similar negative likelihood ratios but calcium as an individual marker is stronger than raised creatinine. Appendix 9 also shows the LR-s when the whole full blood count is used but results were very similar when using just haemoglobin. This can be explained by the fact that very few patients had an abnormal FBC without abnormal haemoglobin. Low platelets without low haemoglobin occurred only in 3% of cytopenic patients and low white cell count without low haemoglobin occurred only on 2% of cytopenic patients.

Below I illustrate an example of the change between the pre-test to the post-test probability depending on the LR and in appendix 10 I present the methods used for doing the above

calculation in addition with a table showing the post-test probability for a range of values for both the positive and negative likelihood ratio.

Example: assuming that the pre-test probability of myeloma is 0.03% (based on figures from cancer research UK) and a negative likelihood ratio of 0.2 then this corresponds to a post-test probability of 0.006%. With the same pre-test probability and a positive likelihood ratio of 10 the post-test probability becomes 0.3%. If a higher pre-test probability is assumed i.e. 20% then a likelihood ratio of 0.9 corresponds to a post-test probability of 18.4% and a likelihood ratio of 10 corresponds to a post-test probability of 71%

Table 2: The rule-out value of individual and combinations of investigations for myeloma. Green corresponds to a LR- of ≤0.2 and red corresponds to LR- > than 0.8.Abbreviations: Haemoglobin (Hb), calcium (CAL), creatinine (CREAT), Sedimentation rate (ESR), C-reactive protein (CRP), plasma viscosity (PV), true positives (TP), false positives (FP), false negatives (FN), true negatives (TN).

Investigations	N	TP	FP	FN	TN	Sensitivity (%)	Specificity (%)	LR-
Individual								
ESR	2192	844	546	154	648	85 (82-87)	54 (51-57)	0.28 (0.24-0.33)
PV	433	162	95	39	137	81 (74-86)	59 (52-64)	0.32 (0.24-0.44)
Hb	5934	1243	912	596	3183	68 (65-70)	78 (76-79)	0.42 (0.39-0.45)
CAL	2809	246	43	893	1627	22 (19-24)	97 (97-98)	0.81 (0.78-0.83)
Creatinine	7407	648	1021	1202	4536	35 (33-37)	82 (81-83)	0.80 (0.77-0.83)
CRP	1274	316	281	319	358	46 (42-50)	63 (59-67)	0.87 (0.78-0.94)
Combinations (2 tests)								
Hb + PV	396	175	119	10	92	95 (90-97)	44 (37-50)	0.12 (0.07-0.23)
Hb + ESR	2071	884	603	72	512	93 (91-94)	46 (43-49)	0.16 (0.13-0.21)
Hb + CRP	1193	491	311	111	280	84 (80-86)	47 (43-52)	0.39 (0.32-0.47)
Combinations (3 tests)								
Hb + CREAT + PV	356	165	111	6	74	97 (93-99)	40 (33-47)	0.09 (0.04-0.20)
Hb + CAL + PV	213	128	49	3	33	98 (94-100)	40 (30-52)	0.06(0.02-0.18)
Hb + CREAT + ESR	1864	823	599	54	388	94 (92-95)	39 (36-42)	0.16 (0.12-0.21)
Hb + CAL + ESR	1146	590	286	45	225	93 (91-95)	44 (40-49)	0.16 (0.12-0.22)
Hb + CREAT + CRP	1115	505	341	70	199	88 (85-90)	37 (33-41)	0.33 (0.26-0.42)
Hb + CAL + CRP	722	363	160	64	135	85 (81-88)	46 (40-52)	0.33 (0.25-0.42)
Combinations (4 tests)								
Hb + CAL + CREAT + PV	211	127	50	3	31	98 (93-100)	38 (28-50)	0.06 (0.02-0.19)
Hb + CAL + CREAT+ ESR	1103	569	309	39	186	94 (91-95)	38 (33-42)	0.17 (0.12-0.24)
Hb + CAL + CREAT + CRP	706	373	189	44	100	89 (86-92)	35 (29-40)	0.31 (0.22-0.42)

3.4.3 Quantification of diagnostic pathway

Table 3 illustrates the length of the different intervals on the myeloma pathway. The primary care interval had a median length of 37 days (IQR: 7-115); no differences were detected by gender (m: 37 vs. f: 37; p=0.88) or age (>60: 38 vs. <=59: 38; p=0.3). The length of the secondary care interval was 75 days (IQR: 30-163 days). Women had a longer secondary care interval than men (80 vs. 68; p=0.08) and younger patients had a longer secondary care interval compared to older patients (>60: 77 vs. <=59: 61; p=0.2) but none of these differences were found to be statistically significant. When stratified by referral speciality 13% of the patients were referred first to a haematologist with a median secondary interval length of 25 days (8-62.5). 67% of patients were referred to other specialities with a median length of 89 (IQR: 40-183). 62% were routine referrals with a secondary care interval of 85.5 days (IQR: 40-175). 22% of the patients had more than 3 referrals which were associated with an interval length of 160.5 days (IQR: 84-259). The total diagnostic interval spanning across primary and secondary care had a length of 148 days (IQR: 65-260) with younger patients having a longer interval (>60: 153 vs. <=59: 126.5; p=0.04).

Table 3: Length of intervals

Total length of intervals	N	25th	50th	75th	90th	p.value
Primary care interval	1074	7	37	115	210	
Secondary care interval	1074	30	75	163	254	
Diagnostic interval	1591	65	148	260	325	
Primary care interval						
Symptoms						
Back pain	380 (35%)	7	26.5	92.5	194.5	
Chest pain	160 (15%)	6.5	40	106.5	179	
Chest infection	143 (13%)	24	77	165	257	
Shortness of breath	107 (10%)	6	39	127	215	
Joint pain	67 (6%)	3	37	137	220	
Fracture	38 (4%)	29	92.5	159	286	P<0.00001*
Nausea	40 (4%)	2	42	64.5	192.5	
Weight loss	43 (4%)	0	14	28	96	
Nosebleeds	36 (3%)	2	26.5	90	167	
Rib pain	34 (3%)	8	52.5	115	197	
Combined bone pain	26 (2%)	1	19.5	127	210	
Referral to diagnosis interval						
Referral type						
Haematologist	144 (13%)	8	25	62.5	127	
Other	718 (67%)	40	89	183	267	P<0.00001**
Missing	212 (20%)	30.5	76.5	161	259	
Referral urgency						
Missing	140 (13%)	37	88	186	184.5	
2 week wait	66 (6%)	14	39.5	120	227	P<0.00001***
Routine	662 (62%)	40	85.5	175	263	

Immediately	47 (4%)	14	43	144	250	
Urgent	159 (15%)	14	43	99	219	
Number of referrals						
1	542 (50%)	15	42	92	177	
2	294 (27%)	50	101.5	179	263	P<0.00001****
>3	146 (22%)	84	160.5	259	321	

*difference in the length of the primary care interval by presenting symptoms
** difference in the length of the referral to diagnosis interval length by referral type
***difference in the length of the referral to diagnosis interval length by urgency of referral
****difference in the length of the referral to diagnosis interval length by number of referrals

3.5 Discussion

3.5.1 Summary

The individual inflammatory markers which have the best rule in and rule out properties for myeloma are ESR and PV, both as a red flag when abnormal, and a rule-out test when normal. This is in contrast to CRP which has very little value in myeloma diagnosis. Both calcium and creatinine rise on average before a diagnosis of myeloma which is a surrogate measure for calcium and creatinine on the individual level. A negative calcium or creatinine alone should not be used to rule it out because they only start to rise late in the disease process. Conversely, haemoglobin starts to decline on average much earlier, making a normal FBC a more useful rule out test.

In the year before diagnosis, 95% of myeloma patients who had plasma viscosity and full blood count tested, had at least one abnormality (95% confidence intervals 90-97 %). The combination of a normal ESR or PV and normal haemoglobin can be used to rule out myeloma in patients presenting with symptoms such as back pain.

The median referral to diagnosis interval is longer than the primary care interval but that depends on the type and urgency of referrals. Referrals to the haematology department results in fairly short referral to diagnosis intervals while being referred to other departments results in

longer waiting times. Multiple referrals and inappropriate urgency of referral further lengthen the interval with patients having three or more referrals having the longest intervals.

3.5.2 Strengths and limitations

The strengths include the large sample size, representative of the UK, and the primary care setting: these increase the generalisability of the results (6). This is also the first study which describes the progression of symptoms and blood tests over time. Most studies usually provide symptom prevalence and blood test abnormalities only at a specific time point. Although useful it does not enable the progression of disease to be studied (7). Similarly, with blood tests it is the first time that the diagnostic accuracy has been examined in combination and not as stand alone. I chose to focus on the ability of the test combinations to rule out the disease and not just rule it in which is the main focus of the lot of diagnostic research conducted in cancer. Tests that have good rule out properties can be very useful in primary care to rule out rare diseases such as myeloma with sufficient confidence. Other conditions where the rule out properties of different strategies in primary care include pulmonary embolism and coronary heart disease (8,9). I have also explored the full diagnostic pathway from first presentation to primary care with a myeloma related symptom up until confirmation of diagnosis including the primary and referral to diagnosis interval. Previous studies on the same topic mainly concentrated on the length of the diagnostic interval but not the whole pathway (10,11).

The choice of a case-control design could be criticised as they can over-estimate the accuracy of diagnostic tests. This limitation is alleviated by the fact that this study is using a nested-case control design and the overestimation is not as much of a problem compared to conducting an unmatched case-control design (12). Another very important important limitation of the study is missing data for the different blood tests, because patients were not tested, which is applicable for single tests and test combinations. If patients that are tested are systematically different from the untested patients (one reason could potentially be that sicker patients are more likely to be tested) this may lead to under or over estimation of the sensitivity, specificity and the likelihood ratios we report (13). This means that these findings are generalizable to the

patients currently being tested in primary care but caution should be exercised when extrapolating them to patients not currently being tested. The combinations recommended in this study should not be used as a screening tool for all patients visiting their GP but they should be targeted on patients who experience symptoms that might indicate myeloma such as back pain, rib pain, chest pain and repeated infections. It is reassuring thought that in another nested case control study of 110 myeloma patients conducted in Israel they identified the same sensitivity values for haemoglobin and ESR (68% for haemoglobin and 82% for ESR) (14). Finally as with most clinical databases, there will be some under-reporting or under-recording of symptoms which can influence the diagnostic accuracy (6). Not all patients with a symptom had a referral which might be due to under-recording of referrals in the CPRD database or this might be a reflection of the fact that many myeloma patients are diagnosed through emergency presentations.

3.5.3 Comparison with existing literature and findings from the previous chapter

This work is an extension of the Shephard et al. study in which the predictive ability of single symptoms, dual symptom combinations and the combination of single symptom plus single abnormal blood test were examined (1). This analysis builds further on these findings by quantifying the predictive ability of blood test combinations and by examining the "trajectories" of symptoms and blood tests over time before a myeloma diagnosis. In addition a more detailed investigation of the referral to diagnosis interval is conducted in this analysis.

A Scandinavian study of 167 blood donors who developed myeloma has previously identified declining haemoglobin for three years before diagnosis, similar to our results(15). I could find no similar reports for the timing of hypercalcemia , reduced kidney function and the other components of the full blood count, though hypercalcemia is common in patients with advanced cancer stage (16). A common reason for myeloma patients to present as emergency admissions is renal failure and this work suggests that deterioration happens rapidly as creatinine and

calcium spike in the last three months before diagnosis. Identifying myeloma patients at the early stages of renal insufficiency can lead to survival benefits and to better treatment options as mentioned in chapter 1 (17). Creatinine and calcium have been traditionally considered as high risk manifestations for myeloma patients and although they can be quite predictive they are not necessarily helpful when it comes to early diagnosis (18). No other studies were identified that examined how the tests perform in combination.

NICE guidelines currently recommend the use of PV or ESR when myeloma is suspected (19); the results of this study strongly support this recommendation. ESR has been reported as a good prognostic marker for myeloma with higher values of ESR associated with a more advanced cancer stage (20). In contrast, CRP is not useful as an initial investigation if myeloma is suspected. Previous reports have concentrated on its use in prognosis rather than diagnosis and concluded that CRP measurements are not valuable as a monitor of disease activity in myeloma patients(21,22).

The results in this chapter indicate that 50% of myeloma patients experience a diagnostic interval of more than 148 days while the systematic review showed the median diagnostic interval to be approximately 109 days. Although they are slightly different both indicate intervals longer than three months which is quite substantial.

Similarly the median primary care interval is longer in this study than the systematic review (median length of the primary care interval in the review is 22 days while in this study is 37 days). The only other studies that report the length of the primary care interval for myeloma are based on the National Audits of cancer diagnosis in primary care that took place in 2009-2010 and 2014 and used the same methods of data collection (23,24). The study design for the audits included GPs and other health professionals entering primary care data on patient's characteristics, place of presentation and symptoms, pre-referral consultations etc. on patients who were diagnosed with cancer. The clinicians responsible for collecting the above data were also determining the date of first presentation with a myeloma related symptoms based on their

clinical judgement. For this study the first presentation was the first code that indicated a myeloma related symptom from a pre-specified list. This difference in the way that the date of first presentation with a myeloma related symptom is defined might be the reason for the differences observed in the length of the primary care interval. Variation in symptom definitions could also explain some of the differences.

The referral to diagnosis interval is longer than the primary care interval. This was also identified in the systematic review and the results in this study confirm that. The main reason for the very long referral to diagnosis interval is multiple referrals. Two previous studies described the referral pathway that resulted in a diagnosis of myeloma. These studies report that approximately 20% of myeloma patients are diagnosed through the two week wait referral route which is higher than the estimate in this chapter which is approximately 6% (25,26). In Howell et al 53% of the patients first referral was to a haematologist which is far higher than the 13%. These discrepancies can be attributed again to the way the data were collected. Both studies were based in secondary care and therefore only had data from the time the patient presented to hospital. The referral date reported in the studies does not necessarily correspond to the first referral after presentation in primary care which is the definition I used in this study. Approximately 50% of the patients in this study experience more than one referral but multiple referrals are not explored in the studies mentioned above. In a much smaller qualitative study of 20 myeloma patients in which data were also collected for the all the referrals they had before diagnosis, 25% of the patients were send straight to haematology which is more in-line with the results in this analysis and 66% of the patients felt that they had a delayed diagnosis (27).

3.5.4 Implications for practice

My results indicate that a substantial proportion of the delays patients are experiencing are due to multiple referrals and patients not being sent to the haematology department. Better targeted referrals could reduce the time to diagnosis and thus potentially improve the outcomes of myeloma patients.

The results of this study indicate that more appropriate referrals can be achieved by ordering simple blood tests in patients with myeloma related symptoms like back pain, rib pain, chest pain and recurrent chest infections. The 95% sensitivity of the combination of an abnormal haemoglobin or abnormal inflammatory marker (PV or ESR) makes the chance of myeloma extremely low. Adding CRP or creatinine adds little additional sensitivity. Conversely, in a patient reporting myeloma suggestive symptoms, with abnormalities in any of the inflammatory markers (PV or ESR) or in the full blood count, an approach that may have potential is for the clinician to request protein electrophoresis plus urinary Bence Jones protein urgently. This could potentially expedite myeloma diagnosis as low haemoglobin or elevated ESR/PV can manifest very early during disease progression. In addition clinicians would feel more confident to rule out myeloma on the basis of these simple tests which can results in fewer haematology referrals that can lead to over-diagnosis of MGUS patients (28). Hypercalcaemia is a high-risk marker of possible myeloma. Normal calcium levels do not make myeloma sufficiently less likely for the calcium test to be used as a safe 'rule out' test alone. Similar results were observed for high creatinine. This represents the fact that hypercalcaemia and reduced renal function are late features of the disease. Addition of calcium can also be used in the above combination in the presence of certain symptoms like back pain, rib, joint pain and fracture since abnormal calcium has a relatively high rule-in potential (PPV >10%) when combined with the pre-mentioned symptoms (1). Combinations when all tests are abnormal were not examined as it is important to investigate further abnormalities even if one of the components of the combination is abnormal. It has been shown that patients with more than 2 myeloma defining events (hypercalcaemia, renal failure, anaemia, and bone lesions) are more likely to be diagnosed via emergency presentation and also have worse survival (20). Thus abnormalities in more than 2 components in any of the combinations should result in very urgent investigation.

Other symptoms appear to represent more advanced disease, and thus are only reported late in its course. Nausea is one of these, possibly arising from hypercalcaemia or kidney failure; similarly, breathlessness may relate to severe anaemia; and weight loss suggesting advanced,

multifocal, disease. The results of this study also indicate that in the case of myeloma the pre-diagnostic interval should be longer than 2 years which is quite important for studies who want to establish the onset of the first symptom.

3.6 Conclusion and implications for next chapter

Myeloma is a difficult disease to identify – especially in its early stages and hence this is why we see long diagnostic intervals which can be a factor of multiple referrals and multiple consultations so implementation of a strategy which can result in fewer consultations and better targeted referrals is essential.

Testing with FBC, ESR/PV and calcium in certain cases appears to be sufficient to select those patients who may have the disease. Those with test abnormalities can proceed to more specific primary care blood testing, and if these tests are abnormal, they warrant haematology expertise. In contrast, if the initial tests are normal, myeloma becomes extremely unlikely, and another cause can be sought for the patient's symptoms.

The work in this chapter has helped determine which test could enable the biggest reductions in the diagnostic interval .One such test is the full blood count which is a very common test in primary care, is ordered for multiple reasons and as I have shown in this chapter it can become abnormal in myeloma even at 2 years before diagnosis (29) .

A prediction rule based on routine tests could help GP's suspect myeloma earlier. This approach is going to be explored in more detail in the next chapters.

References

1. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. Br J Gen Pract. 2015; 65(631):e106–13.

2. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. Br J Cancer. 2012; 106(7):1262–7.

3. Jaeschke R, Guyatt GH, Sackett DL, Guyatt G, Bass E, Edwards PB, et al. Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test B. What Are the Results and Will They Help Me in Caring for My Patients? JAMA J Am Med Assoc. 1994; 271(9):703–7.

4. Thompson M, van den Bruel A, Verbakel J, Lakhanpaul M, Haj-Hassan T, Stevens R, et al. Systematic review and validation of prediction rules for identifying children with serious infections in emergency departments and urgent-access primary care. Health Technol Assess (Rockv). 2012;16(15):1–99.

5. McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002; 17(8):647–50.

6. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T Van, et al. Data Resource Profile : Clinical Practice Research Datalink (CPRD). Int J Epidemiol 2015; 44(3):827–36.

7. Talamo G, Farooq U, Zangari M, Liao J, Dolloff NG, Loughran TP, et al. Beyond the CRAB Symptoms: A Study of Presenting Clinical Manifestations of Multiple Myeloma. Clin Lymphoma Myeloma Leuk. 2010;10(6):464–8.

8. Haasenritter J, Bösner S, Vaucher P, Herzig L, Heinzel-Gutenbrunner M, Baum E, et al. Ruling out coronary heart disease in primary care: external validation of a clinical

prediction rule. Br J Gen Pract. 2012; 62(599):e415-21.

9. Bösner S, Haasenritter J, Becker A, Karatolios K, Vaucher P, Gencer B, et al. Ruling out coronary artery disease in primary care: development and validation of a simple prediction rule. CMAJ. 2010; 182(12):1295–300.

10. Din NU, Ukoumunne OC, Rubin G, Hamilton W, Carter B, Stapley S, et al. Age and Gender Variations in Cancer Diagnostic Intervals in 15 Cancers: Analysis of Data from the UK Clinical Practice Research Datalink. PLoS One. 2015;10(5):e0127717.

11. Neal RD, Din NU, Hamilton W, Ukoumunne OC, Carter B, Stapley S, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. Br J Cancer. 2014;110(3):584–92.

12. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, Van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006;174(4):469–76.

13. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ. 2016; 353:i3139.

14. Goldschmidt N, Zamir L, Poperno A, Kahan NR, Paltiel O. Presenting Signs of Multiple Myeloma and the Effect of Diagnostic Delay on the Prognosis. J Am Board Fam Med. 2016;29(6):702–9.

15. Edgren G, Bagnardi V, Bellocco R, Hjalgrim H, Rostgaard K, Melbye M, et al. Pattern of declining hemoglobin concentration before cancer diagnosis. Int J Cancer. 2010;127(6):1429–36.

16. Mirrakhimov AE. Hypercalcemia of Malignancy: An Update on Pathogenesis and Management. N Am J Med Sci. 2015; 7(11):483–93.

17. Wirk B. Renal failure in multiple myeloma: a medical emergency. Bone Marrow

Transplant. 2011;46(6):771–83.

18. Bird JM, Owen RG, D'Sa S, Snowden JA, Pratt G, Ashcroft J, et al. Guidelines for the diagnosis and management of multiple myeloma 2011. Br J Haematol. 2011; 154(1):32–75.

19. Multiple myeloma - NICE CKS [Internet]. [cited 2017 Sep 1]. Available from: https://cks.nice.org.uk/multiple-myeloma#!scenario

20. Alexandrakis MG, Passam FH, Ganotakis ES, Sfiridaki K, Xilouri I, Perisinakis K, et al. The clinical and prognostic significance of erythrocyte sedimentation rate (ESR), serum interleukin-6 (IL-6) and acute phase protein levels in multiple myeloma. Clin Lab Haematol. 2003; 25(1):41–6.

21. Hsu DC, Wilkenfeld P, Joshua DE. Multiple myeloma. BMJ. 2012; 344:d7953–d7953.

22. Brown RD, Snowdon L, Uhr E, Gibson J, Joshua D, Snowdon L, et al. C-Reactive Protein (CRP) Levels Do Not Reflect Disease Status in Patients with Multiple Myeloma C-Reactive Protein (CRP) Levels Do Not Reflect Disease Status in Patients with Multiple Myeloma. Leuk Lymphoma. 1993; 9(6):509-12.

23. Lyratzopoulos G, Abel G a, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. Br J Cancer. 2013;108(3):686-90.

24. Swann R, McPhail S, Witt J, Shand B, Abel GA, Hiom S, et al. Diagnosing cancer in primary care: results from the National Cancer Diagnosis Audit. Br J Gen Pract. 2018;68(666):e63–72.

25. Howell D, Smith A, Appleton S, Bagguley T, Macleod U, Cook G, et al. Multiple myeloma: routes to diagnosis, clinical characteristics and survival – findings from a UK population-based study. Br J Haematol. 2017;177(1):67–71.

26. Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. Br J Cancer. 2012;107(8):1220–6.

27. Howell DA, Hart RI, Smith AG, Macleod U, Patmore R, Cook G, et al. Myeloma: Patient accounts of their pathways to diagnosis. Harris F, editor. PLoS One. 2018; 13(4):e0194788.

28. Go RS, Swanson KM, Sangaralingham LR, Habermann EB, Shah ND. Clinical prevalence (diagnosed cases) of monoclonal gammopathy of undetermined significance in the US: estimating the burden on health care. Leukemia. 2016;30(6):1443–6.

29. Busby J, Schroeder K, Woltersdorf W, Sterne JAC, Ben-Shlomo Y, Hay A, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. Br J Gen Pract. 2013; 63(609):e256-66.

Chapter 4: Derivation of clinical prediction rules for early diagnosis of myeloma: a retrospective cohort study based on CPRD data

4.1 Chapter Introduction

In the previous chapter I identified that one the earliest indications of myeloma is changes in haemoglobin which can start becoming abnormal up to two years before diagnosis. I also identified some useful test combinations which can be used to identify people with a higher risk of myeloma. These include the full blood count (more specifically haemoglobin), an inflammatory marker (Erythrocyte sedimentation rate or plasma viscosity, but not CRP) and calcium in certain cases. An abnormality in any of these tests requires further investigation with more definitive tests (protein and urine electrophoresis). The above combination of haemoglobin, ESR and calcium in certain cases is a good combination of tests that can be used when myeloma is suspected.

A significant challenge in diagnosing myeloma is the fact that it is a rare disease and a GP is unlikely to suspect myeloma and thus not order all tests. An additional problem is that symptoms are non-specific for myeloma. With this in mind, I wanted to investigate whether a commonly used test such as the full blood count could be used to improve the diagnosis of myeloma (1). In addition taking advantage of all the components of the full blood count (haemoglobin, platelets, white cell count etc.) along with symptoms and risk factors within a clinical prediction seems like a good way to identify patients that might have myeloma when the disease becomes symptomatic but is not yet suspected. This is particularly important as myeloma is one of the few cancers in which abnormalities can appear in all components of the full blood count. In addition, using clinical prediction rules also means that we make more efficient use of the available information potentially making our prediction rules more accurate and improving decision making. It has been shown in colorectal cancer that using scoring systems like prediction rules perform significantly better than NICE referral guidelines (2).

The aim of this study is the development of a series of prediction rules with a focus on the full blood count which can help with expediting the diagnosis of myeloma. This chapter will just focus on the development of the prediction rules. The validation, examination of clinical utility and implementation will be examined in the next chapter.

4.2 Objectives and aims

General objective: The development of clinical prediction rules based on risk factors, symptoms and laboratory investigations.

This study mainly focused on models that make use of the full blood count

- Model 1: demographic characteristics, risk factors, symptoms and full blood count test results
- Model 2: demographic characteristics, risk factors, symptoms and changes from previous full blood count tests
    - Sensitivity analysis for model 2: demographic characteristics, risk factors, symptoms, full blood count test results and changes from previous full blood count tests.
- Model 3: demographic characteristics, risk factors, symptoms and a broader range of blood tests which include full blood count, calcium, creatinine and inflammatory markers (ESR/CRP/Plasma viscosity)

4.3 Methods

4.3.1 Study design

This study has been approved by the Independent Scientific Advisory Committee (ISAC) with protocol number 17_088R (appendix 11) The study had an open cohort study design including patients over the age of 40 with no previous diagnosis of myeloma and with 2 full blood counts

within a year period. A cohort study was used rather than a case-control because the latter has been shown to overestimate measures of diagnostic accuracy (3).

Patients were eligible to enter the cohort if they were over the age of 40, were registered at "up-to standard" (UTS) CPRD practices and had 2 full blood counts within a period of 1 year and a minimum follow-up of two years after the last full blood count test. The results of the previous chapter showed that certain symptoms and abnormalities in blood tests can manifest up to two years before diagnosis so patients were followed up for 2 years after index date and myeloma patients were identified. The index date was set as the date of a second full blood count result; patients who had a diagnosis of multiple myeloma before the index date were excluded. Patients were identified between 1st January 2000 and 1st January 2014. Eligible patients had to be registered with their practice for a minimum of 1 year prior to the index date to ensure adequate recording of baseline covariates.

4.3.2 Sample size considerations

It has been shown that 20 or more events per variable are usually enough to eliminate overfitting in Cox proportional hazard models in the presence of many low prevalence predictor variables (4). With a maximum of 25 candidate predictor variables in any model (including potential interactions) and with an event rate of 20 events per coefficient, we estimated 500 events to be an adequate sample size for the derivation dataset.

4.3.3 Variable definitions

Outcomes: Myeloma diagnosis (Y/N) within two years of index date. Myeloma diagnosis will be identified through the use of Read codes similarly to how it was done for the case-control study in chapter 3 (READ codes can be seen in appendix 8).

Predictors: Defined within the year before index date.

1. Demographic characteristics: Identified from the patient file
    a. Age (continuous; years)

b. Gender (Binary; Y/N)

2. Risk factors: MGUS will be identified based on Read codes and BMI will be generated by the use of the entity codes 13 (Weight) and 14 (Height). These variables will be identified by using the clinical file.
    a. BMI (continuous; kg/m^2)
    b. MGUS (Binary; Y/N)
3. Symptoms: Symptoms will be defined based on Read code lists and will be identified within the year before the 2^{nd} full blood count (index test). Symptoms can be recorded before or after the 1^{st} full blood count.
    a. Back pain (Binary; Y/N)
    b. Chest pain (Binary; Y/N)
    c. Bone pain (Binary; Y/N)
    d. Rib pain (Binary; Y/N)
    e. Joint pain (Binary; Y/N)
    f. Shortness of breath (Binary; Y/N)
    g. Chest infections (Binary; Y/N)
    h. Fatigue (Binary; Y/N)
    i. Bleeding (Binary; Y/N)
    j. Bruising (Binary; Y/N)
    k. Fracture (Binary; Y/N)
    l. Weight loss (Binary; Y/N)
    m. Nausea (Binary; Y/N)
4. Laboratory investigations: Laboratory investigations will be identified through the use of Read Codes and the numerical value of the different tests will be identified by the use of entity codes which are located in the Test files. Tests identified from Read codes and do not have a numerical value in the entity type will be classified as missing data.
    a. Full blood count
        i. Haemoglobin (continuous; g/dL): Entity code 173.

ii. White blood cells (continuous; 10*9/L): Entity code 207.

iii. Platelets (continuous; 10*9/L): Entity code 189

iv. Mean corpuscular volume (continuous; fL): Entity code 182

b. Creatinine (continuous; umol/L): Entity code 165

c. Calcium (continuous; mmol/L): Entity code 159

d. Inflammatory markers

i. Erythrocyte sedimentation rate (continuous; mm/h) : Entity code 273

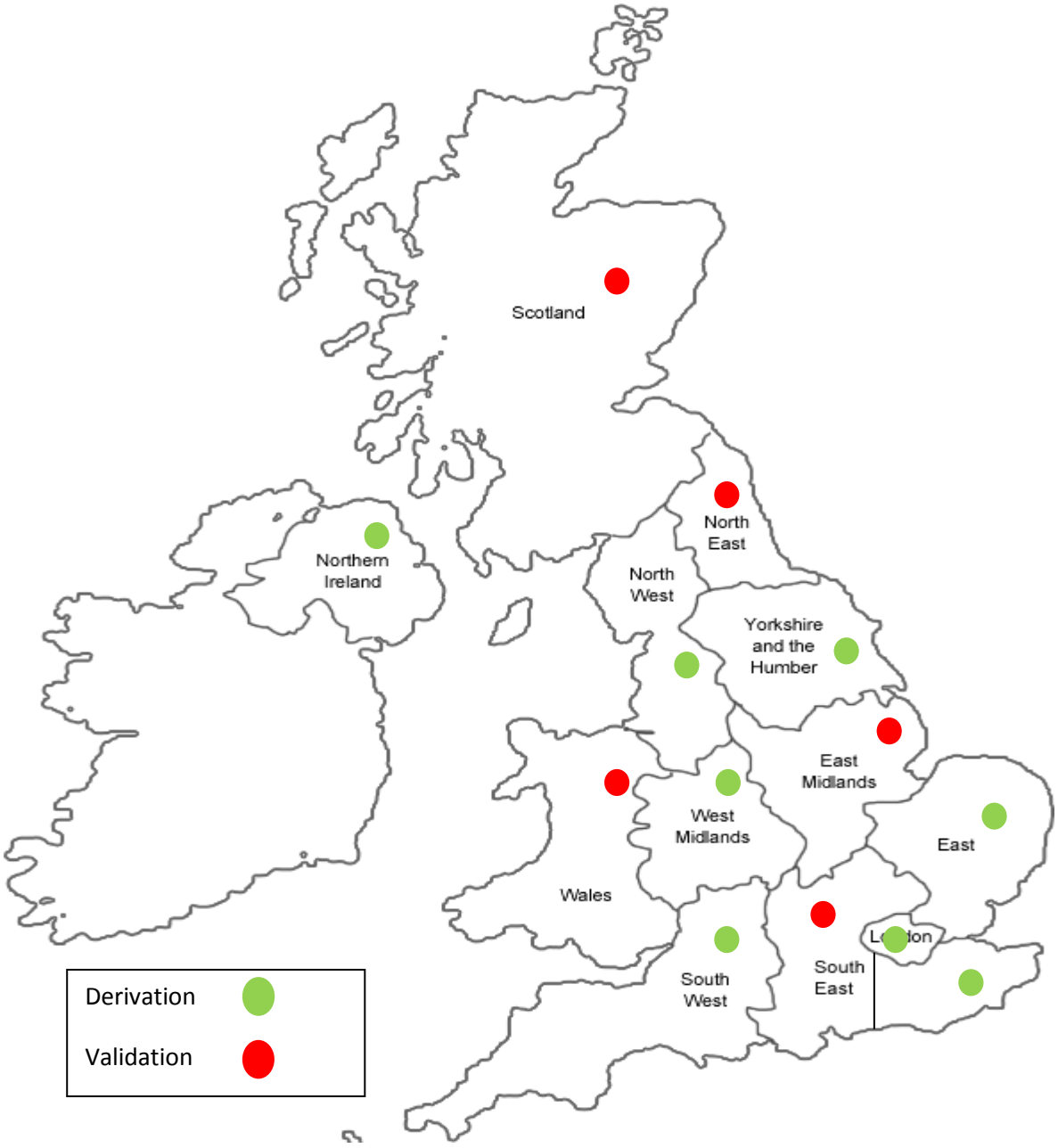
ii. C-Reactive protein (continuous; mg/L): Entity code 280

iii. Plasma viscosity (continuous; mPas): Entity code 374

4.3.4 Statistical analysis

The database was randomly split in derivation and validation sets based on geographical region to allow for geographic validation. I selected this approach as a geographical split is a more robust form of validation compared to a random split into derivation and validation datasets as the two sets in the random split can be very similar and vary only by chance (5). An additional approach would have been to use temporal validation which examines the generalizability of the prediction rule over time. I have decided to use the geographic split as it allows me to investigate the generalisability of the model to other regions and practices within the UK. The regions were randomly assigned to the two sets with more regions in the derivation set due to requiring a larger sample size (6). Figure 1 illustrates the UK map along with the regions that were assigned in the derivation and the validation set.

Figure 1: UK map illustrating the derivation and validation datasets



Descriptive statistics and univariate analysis

Descriptive statistics of baseline characteristics in the derivation and validation cohorts were produced. I have included the basic demographics for the derivation dataset in this chapter rather as the comparison is important to show that the two populations are similar. Binary and categorical variables were presented as percentages and mean/standard deviation or median/interquartile range were used for continuous variables depending on data distribution. The continuous nature of the blood test variables was retained in order to avoid any loss of information and reduction in power but an additional analysis was conducted where the blood test results were classified as normal/abnormal depending on the reference range provided by the local laboratory. Univariate associations between predictors and outcome were examined by the use of Cox proportional hazards model and model assumptions like the proportional hazards assumption were examined. The proportional hazards assumption assumes that the observed effects are constant over time i.e. the hazard ratios are constant over time.

Model derivation

Multivariate models were fitted by the use of Cox proportional hazards models. Time to diagnosis was estimated by subtracting the date of diagnosis minus the index date. Continuous variables were centered and rescaled which can help with the convergence of the models. For models which include inflammatory markers I included the marker with the highest hazard ratio from the univariate analysis as it is unlikely that all 3 tests are ordered in practice at the same time. Interactions of predictor variables with age were also examined

Linearity of predictors

Transformations or fractional polynomials were used in order to identify the correct functional form of the continuous variables such as the BMI and blood tests (7). One of the most important assumptions of the Cox regression model is the linearity assumption which assumes that the predictors are linearly related with the log-hazard. Violation of this assumption can be corrected with the use of transformations. Fractional polynomials are used when traditional

transformations fail to model the true relationship between the outcome and the predictor under investigation. The usual set of powers we tend to use when trying transformations include {-2, -1, -0.5, 0, 0.5, 1, 2, 3} with 0 corresponding to the logarithmic one, 2 to the quadratic, -1 to the reciprocal etc. These transformations are also known as FP1 polynomials (i.e. degree 1 polynomials). FP2 polynomials include taking dual combinations of the powers illustrated above including repeated powers (i.e. 3 and 3). Together FP1 and FP2 result in 44 models and all of them are fitted when the relationship between continuous predictor and the outcome is examined and the best FP1 and FP2 terms are retained. The FP2 and FP1 models are then compared to the linear, the null model (model without the predictor) and between them, with the use of formal statistical tests. The selection of the best FP2 and FP1 is done by taking the models with the lowest deviance. Deviance is a goodness of fit statistic for a statistical model and it can be seen as measure of lack of fit between a model and the data (Lower deviance corresponds to better models). Polynomials offer a high degree of flexibility when modelling continues variables and they can be very useful when modelling non-standard trends (8).

Model selection

For model selection a backwards elimination procedure was used in which all predictors are included in the model. The alpha value was set at 5%. The predictor with the highest p-value greater than the pre-selected alpha-value is dropped. The model is refitted and the same procedure is repeated until all the remaining variables have a p-value less than the pre-defined alpha value. The backwards elimination procedure was used as the sample size calculations presented in section 4.3.2 indicated that it would be possible to include all the variables in the model without causing overfitting problems. Due to the use of fractional polynomials the final model selection was done by using the multivariate fractional polynomials procedure (mfp) which is a method that combines the backward elimination procedure and the fractional polynomials that was described in the previous section. For binary/categorical variables that do not require transformation the procedure used is the usual backwards elimination selection process as described above. Results are presented as hazards ratios with 95% confidence

intervals. This method is a formal procedure used by most statistical software including STATA and R and in short the mfp procedure follows the following steps for each continuous variable in the model

1. Model inclusion: The best FP2 transformation is compared against the null model at a pre-determined a-value (5% in this analysis). If FP2 is significant then the predictor is retained in the model otherwise is dropped. The null model in this case corresponds to the intercept only model i.e. a model without any predictors.
2. Examining non-linearity: The FP2 is examined against the linear model (i.e. the model in which the predictor is included as a linear term). If FP2 is not significant when compared to the linear model then the linear term is selected for the predictor otherwise the FP2 is carried forward.
3. Simplifying the model: The FP2 is tested against the best FP1 (i.e. the best single transformation). If the FP2 is significant then this is the final form of the predictor otherwise the FP1 is selected.

Below I present some STATA output based on a help dataset in STATA where I illustrate how the procedure described above works.

Variable	Model	(vs.)	Deviance	Dev diff.	P	Powers	(vs.)
x5	null	FP2	3503.610	61.366	0.000*	.	.5 3
	lin.		3471.637	29.393	0.000+	1	
	FP1		3449.203	6.959	0.031+	0	
	Final		3442.244			.5 3	

In the above output x5 is a continuous variable and the mfp procedure will be used to assess whether the variable is going to be included in the model and whether a transformation is required. First the best FP2 polynomial is selected (which is the one that has the highest deviance). In this case the FP(0.5 3) was selected. Similarly the best FP1 polynomial is selected which in this case is the logarithmic transformation.

Step 1: Model inclusion

The FP2 polynomial is compared with the null model in terms of their deviance. If the p.value is significant then this means that the FP2 model is better than the null model thus the variable is retained in the model. If the p.value was not significant then the variable x5 would have been dropped from the model.

Step 2: Examination of non-linearity

Given that the x5 variable is going to be retained in the model then we need to examine whether the FP2 model is statistically better than the linear model. Similarly to before the deviances are compared and based on the p.value we can see that the FP2 is better than the linear model. Thus it can be concluded that the linearity assumption is not satisfied

Step 3: Simplification

This is the last step which examines whether we can simplify the model i.e. rather than using an FP2 polynomial use an FP1 polynomial (in this case a simple logarithmic transformation). Similarly to before, the deviances are compared but still the FP2 model is better than the logarithmic transformation and thus the FP2 polynomial was selected for the final model.

For variables that were modelled through the use of fractional polynomials I have produced plots that show the relationship between outcome and predictor as the hazard ratios cannot be interpreted in the usual way. Y-axis corresponds to the component (linear predictor function – the set of coefficients and predictor variables that were identified to predict the outcome). Shaded areas correspond to the 95% confidence intervals. Methods for the model validation and examination of clinical utility will be presented in the next chapter

<u>Dealing with missing data:</u>

Missing data patterns were investigated and multiple imputation methods were used in the presence of 20% total missing data within each model. Multiple imputation is a statistical technique which generates different plausible imputed data sets and then combines the results from each one (9). In order for imputation techniques to be used some assumptions need to be

made with respect to the data and more specifically they need to be at least MAR (Missing at random) i.e. any systematic differences between the missing values and the observed values can be explained by differences in observed data meaning that we can predict the value that is missing based on other data. I discuss the plausibility of this assumption in the results and discussion sections.

In short multiple imputation can be summarised in three phases

Phase 1 (Imputation): Missing entries of the datasets are imputed which results in the creation of different versions of the dataset. Depending on the type of variable we want to impute we will use the equivalent model. For example if we have missing data on a binary variable then we would fit a logistic regression model.

Phase 2 (Analysis): The analysis of interest is conducted in each individual imputed dataset.

Phase 3 (Pooling): The estimates from the models are pooled together into final results by using Rubin's rule. According to Rubin's rule the pooled estimate is the average of the statistics estimated in each imputation dataset and the corresponding standard error is estimated by taking into account the within and between imputation variation.

I carried out ten imputations using chained equations. This has been shown to provide sufficient accuracy under specific conditions which usually include the fraction of missing data in the dataset (10) . Inefficient estimates tend to have more sampling variability than necessary. Some models might have required more imputations due to having large fractions of missing data but ten was a realistic number after taking into account the size of the datasets, the computational requirement, the imputation models and the time required for these models to be fitted. These are explained more extensively in the discussion section. The imputation models in this study contained all the predictors, the binary indicator for the outcome and the cumulative baseline hazard (the hazard or failure rate obtained when all model predictors are set to 0) estimated by the Nelson – Aalen estimator (11). For the imputation models in this study the predictive mean matching procedure (PMM) was used in place of linear regression for variables

that were not normally distributed (12). Predictive mean matching is a method which combines linear regression and closest neighbour approaches. Linear regression is used to obtain linear predictions and by using this linear prediction as a distance measure it identifies a set of values (amongst the complete values) which are called nearest neighbours. The imputed value is selected at random amongst that set of nearest neighbours. For this analysis I used a set of 5 values as the set of nearest neighbours (13). Imputation diagnostics were performed by comparing the distributions of the imputed and the observed data (14).

One of the most challenging components of this analysis was the combination of model selection methods, fractional polynomials and multiple imputation which created a very complicated modelling strategy (15). In order to do this, I used the mfpmi command in STATA 14 which is an extension of the mfp procedure mentioned above within a multiple imputation setting. This procedure selects the best multivariate fractional polynomial for multiply imputed data based on likelihood-ratio tests. This procedures works by treating the M imputed datasets as one big dataset of size n x M (which is known as "stacking") where n is the number of observations in each dataset and each observation is given a different weight depending on the fraction of missing data of the variable (16). The above method is an improvement from more traditional methods which typically include fitting the model in each individual imputed dataset and selecting the predictors that appear in any of these models or the predictors that appear in at least half of the models or select the ones that appear in all models.

Multiple imputation results were compared with the complete case analysis to examine the robustness of the procedure and to detect any problematic inconsistencies (i.e. associations changing directions etc.).

Baseline survival

In order for the Cox proportional hazards model to be used for individual patient predictions the baseline survival probability at time t needs to be estimated (at two years in this case). Another option would have been to use parametric survival models which do not require the estimation

of the baseline survival probability. The reason I did not use parametric survival models is that they require more assumptions than the semi-parametric Cox model. More specifically for parametric survival models the distribution of the survival times needs to be specified. This distribution can take many shapes like the normal, log-normal, Weibull etc. and by assuming that the survival time follows any of these distributions then a maximum likelihood parametric approach can be used. The problem is that the shape of the distribution is affected by censoring and covariates so a lot of the times it is not enough to just generate the histogram of the survival times and select the best fitting distribution as it might be very different in another dataset. Given that this is the first time that clinical prediction rules are generated for myeloma diagnosis no information is available in the literature about the shape of the survival functions which could inform the validity of the selected distribution thus I considered it safer to use a semi-parametric approach.

To calculate this on the imputed data the final model was fitted in each imputed dataset, the survival probability at two years was then estimated by using the Kaplan-Meir estimate and combined by using Rubin's rule.

4.4 Results

The total number of patients included in the study is 1,283,042 with 770 incident myeloma cases within two years after study inclusion. Eight regions were randomly assigned to the derivation dataset and five regions to the validation dataset with 516 myeloma patients in the derivation and 254 cases in the validation.

4.4.1 Baseline characteristics of derivation and validation datasets

Table 1 illustrates the baseline characteristics for the derivation and the validation datasets. The incidence of myeloma within 2 years in the derivation dataset was 0.06%. The mean age is 63.7 years old (SD 13.8) with 41% being male. MGUS was also very rare (0.08%). Within the year before diagnosis BMI was recorded for approximately 35% of the patients. Although the cohort

was selected based on patients who had a full blood count, there were missing data for some of the components of the full blood count ranging from 5% to 12% as the selection was based on at least one component. The median time between the two full blood count tests was 5 months (1.9-8.6). Variables with the most missing data included calcium (available in 24% of patients), creatinine (46%), ESR (26%), CRP (19%) and PV (4%). The missing data patterns for blood tests can be seen in appendix 12 but briefly one of the most striking findings is the fact that approximately 32% of the patients were tested with a full blood count only and approximately 7% of the patients were tested with all tests. There was an increasing trend in the recording of blood tests over time. Approximately 18% of the full blood count tests were recorded within the first 5 years of the study compared to 41% recorded in the last 5 years. Similar patterns were observed for all blood tests. Very similar results can be observed in the validation dataset making the patients in the two datasets comparable in terms of their age, gender, risk factors, symptoms and blood tests distributions independently of UK region.

Table 1: Baseline demographics of patients in derivation and validation datasets

	Derivation cohort (n=836056)	Validation cohort (n=446986)
Myeloma	516 (0.06%)	254 (0.06%)
Demographics		
Male (%)	343467 (41%)	181124 (41%)
Female (%)	492589 (59%)	262862 (59%)
Mean age (SD)	63.7 (13.8)	63.6 (13.7)
Risk factors		
BMI recorded prior to index date (%)	35 %	36%
Median BMI (IQR)	27.7 (24.3-31.8)	28.1 (24.6-32.3)
MGUS (%)	652 (0.1%)	464 (0.1%)
Symptoms*		
Any myeloma symptom (%)	324336 (39%)	169700 (38%)
Back pain (%)	78472 (9.4%)	39519 (8.8%)
Chest pain (%)	53366 (6.4%)	28479 (6.4%)
Bone pain (%)	12956 (1.6%)	6364 (1.4%)
Rib pain (%)	3821 (0.46%)	2211 (0.49%)
Joint pain (%)	35404 (4.2%)	16984 (3.8%)
Shortness of breath (%)	66194 (7.9%)	37698 (8.4%)
Chest infections (%)	54310 (6.5%)	28468 (6.4%)
Fatigue (%)	66970 (8.0%)	34495 (7.7%)
Nosebleeds (%)	7042 (0.84%)	3472 (0.78%)

Bruising (%)	8693 (1.0%)	4305 (0.96%)
Fracture (%)	13395 (1.6%)	8156 (1.8%)
Weight loss (%)	10915 (1.3%)	5725 (1.3%)
Nausea	25174 (3%)	13696 (3%)
Blood tests		
Full blood count (2nd test-index)		
Haemoglobin recorded prior to index date (%)	785727 (94%)	419259 (94%)
Mean Haemoglobin (SD)	13.54 (1.57)	13.56 (1.59)
White cell count recorded prior to index date (%)	773153 (92%)	408370 (91%)
Mean White cell count (SD)	7.3 (5.2)	7.3 (3.3)
Platelets recorded prior to index date (%)	768774 (92%)	406374 (91%)
Mean platelets (SD)	265.4 (80.2)	266.9 (81.1)
MCV recorded prior to index date (%)	763401 (91%)	390441 (87%)
Mean MCV (SD)	90.6 (5.98)	91.0 (5.95)
Full blood count (1st test)		
Haemoglobin recorded prior to index date (%)	778401 (93%)	408565 (91%)
Mean Haemoglobin (SD)	13.5 (1.7)	13.5 (1.7)
White cell count recorded prior to index date (%)	767489 (92%)	399925 (89%)
Mean White cell count (SD)	7.4 (5.3)	7.3 (3.4)
Platelets recorded prior to index date (%)	763070 (91%)	398151 (89%)
Mean platelets (SD)	267 (82.4)	268.7 (83.3)
MCV recorded prior to index date (%)	757827 (91%)	382028 (85%)
Mean MCV (SD)	90.5 (6.4)	90.9 (6.4)
Other tests		
Calcium recorded prior to index date (%)	200699 (24%)	101409 (23%)
Mean Calcium (SD)	2.3 (0.12)	2.3 (0.13)
Creatinine recorded prior to index date (%)	381913 (46%)	206775 (46%)
Mean Creatinine (SD)	89.3 (30.2)	87.6 (30.9)
ESR recorded prior to index date (%)	214440 (26%)	103872 (23%)
Median ESR (IQR)	12 (6-24)	12 (6-24)
CRP recorded prior to index date (%)	155697 (19%)	79119 (18%)
Median CRP (IQR)	5 (2-10)	5 (3-11)
Plasma viscosity recorded prior to index date (%)	37147 (4%)	19609 (4%)
Mean Plasma viscosity (SD)	1.70 (0.16)	1.68 (0.17)

* Symptoms were recorded within the year before the 2nd full blood count which was the starting date of the follow up period. Symptoms could be recorded before or after the 1st full blood count.

4.4.2 Univariate analysis

Table 2 displays the results of the univariate analysis and table 3 contains the results of the blood test parameters when stratified as normal/abnormal rather than continuous. Females were less likely to get myeloma (HR: 0.57; 95% CI: 0.48-0.68) than males. MGUS was very predictive of myeloma (HR: 53.8; 95% CI: 34.4-84.1). The most common symptom was back pain with prevalence 19% in myeloma patients followed by chest pain with a prevalence of

11.2% in myeloma patients. Predictive symptoms included back pain (HR: 2.2; 95% CI: 1.7-2.2), chest pain (HR: 1.8; 95% CI: 1.4-2.4), rib pain (HR: 4.0; 95% CI: 2.1-7.7), shortness of breath (HR: 1.4; 95% CI: 1.03-1.8), nosebleeds (HR: 3.1; 95% CI: 1.8-5.4), fracture (HR: 1.8; 95% CI: 1.04-3.0) and weight loss (HR: 1.8; 95% CI: 1.04-3.3). Approximately 51% of myeloma patients had any of the pre-defined symptoms compared to 39% of the controls.

Most of the continuous variables had non-linear relationships with the outcome while the model assumes that there should be a linear relationship between the log-hazard and the different predictors. The different transformations used in order to deal with the non-linear relationships will be shown in a later section. Proportional hazards assumptions were examined visually with the use of Kaplan-Meier plots for binary variables and Schoenfeld residual plots for continues variables. The plots can be seen in appendix 13. Overall the proportional hazards assumption was satisfied by most variables. When the investigations were classified as normal/abnormal, all the components of the full blood count were predictive of myeloma (anaemia, leukopenia, low platelets and high MCV). The most common blood test abnormality within the full blood count was anaemia with a prevalence of 57% in myeloma patients, followed by an elevated MCV with a prevalence of 20%. Raised calcium and creatinine levels were predictive of myeloma, as were an abnormal ESR or PV whereas CRP was not predictive. When examining the change in the full blood count parameters, only a change in haemoglobin was predictive of myeloma (HR: 0.78; 95% CI: 0.71-0.85) suggesting a decrease in the risk of myeloma when the change from one to the other is positive (i.e. haemoglobin has increased since the last test).

Table 2: Univariate analysis for variables that are going to be included in the model

	Non-myeloma (n=835540)	Myeloma (n=516)	HR (95% CI) Non transformed
Demographics			
Female (%)	492357 (41%)	232 (45%)	0.57 (0.48-0.68)
Mean age (SD)	63.7 (13.8)	70.1 (10.4)	1.04 (1.03-1.05)*
Risk factors			
Mean BMI (SD)	28.5 (6.2)	26.5 (4.5)	0.94 (0.91-0.97)
MGUS (%)	631 (0.08%)	21 (4.1%)	53.8 (34.4-84.1)
Symptoms			
Any myeloma symptom (%)	324072 (39%)	264 (51%)	1.7 (1.4-2.0)
Back pain (%)	78376 (9.4%)	96 (18.6%)	2.2 (1.7-2.7)
Chest pain (%)	53308 (6.4%)	58 (11.2%)	1.8 (1.4-2.4)
Bone pain (%)	12944 (1.6%)	12 (2.3%)	1.6 (0.9-2.8)
Rib pain (%)	3812 (0.5%)	9 (1.7%)	4.0 (2.1-7.7)
Joint pain (%)	35383 (4.2%)	21 (4.1%)	1.0 (0.6-1.5)
Shortness of breath (%)	66139 (7.9%)	55 (10.7%)	1.4 (1.03-1.8)
Chest infections (%)	54268 (6.5%)	42 (8.1%)	1.2 (0.9-1.7)
Fatigue (%)	66935 (7%)	35 (6.8%)	0.9 (0.6-1.2)
Nosebleeds (%)	7028 (0.8%)	14 (2.7%)	3.1 (1.8-5.4)
Bruising (%)	8687 (1%)	6 (1.2%)	1.14 (0.5-2.6)
Fracture (%)	13381 (1.6%)	14 (2.7%)	1.8 (1.04-3.0)
Weight loss (%)	10903 (1.3%)	12 (2.3%)	1.8 (1.04-3.3)
Nausea	25152 (3.0%)	22 (4.26%)	1.4 (0.94-2.2)
Blood tests			
Full blood count (2nd test-index)			
Mean Haemoglobin (SD)	13.5 (1.6)	12 (1.9)	0.62 (0.59-0.65)*
Mean White cell count (SD)	7.3 (5.2)	6.5 (3.6)	0.87 (0.83-0.92)*
Mean platelets (SD)	265.43 (80.2)	247.8 (87)	0.996 (0.995-0.998)*
Mean MCV (SD)	90.6 (5.98)	93.6 (6.03)	1.08 (1.07-1.09)*
Difference in FBC parameters			
Mean Haemoglobin diff (SD)	0.01 (1.0)	-0.25 (1.06)	0.78 (0.71-0.85)
Mean White cell count diff (SD)	-0.03 (5.3)	0.11 (2.3)	1.0 (0.99-1.02)
Mean platelets diff (SD)	-1.8 (54.9)	-3.0 (60.5)	0.99 (0.99-1.00)
Mean MCV diff (SD)	0.19 (3.1)	-0.04 (2.4)	0.98 (0.95-1.00)
Other tests			
Calcium (range: 2.2 -2.6)	2.3 (0.12)	2.4 (0.17)	2.4 (1.8-3.3)*
Creatinine (range: 80-110)	89.2 (30.2)	100 (44.7)	1.007 (1.004-1.009)
ESR (range: 0-15)	18.6 (19.5)	54.8 (41.4)	1.04 (1.03-1.04)
CRP (range: 0-5)	p50: 5 (2-10)	p50: 5 (2-17)	1.00 (0.99-1.007)
Plasma viscosity (range:1.5-1.72)	1.7 (0.16)	1.96 (0.7)	4.7 (3.0-7.3)

*Although in this table I report the coefficients when a linear relationship is assumed these variables have been identified to require transformations.

Table 3: Univariate analysis for blood test variables defined as normal/abnormal rather than continuous

	Non-myeloma (n=835540)	Myeloma (n=516)	HR (95% CI)
Blood tests			
Full blood count (index-test)			
Anaemia (%)	120412 (15.3%)	277 (57.4%)	7.4 (6.2-8.9)
Leukopenia (%)	22447 (2.9%)	60 (13%)	5.0 (3.8-6.6)
Low platelets (%)	25205 (3.3%)	45 (9.7%)	3.2 (2.3-4.3)
High MCV (%)	52446 (6.9%)	88 (20%)	3.4 (2.7-4.3)
Other tests			
Abnormal calcium (%)	4071 (2%)	14 (8%)	4.1 (2.4-7.1)
High creatinine (%)	56203 (14.7%)	65 (26.6%)	2.1 (1.6-2.8)
Abnormal ESR (%)	89851 (41.9%)	134 (80.2%)	5.6 (3.8-8.2)
Abnormal CRP (%)	51044 (33%)	47 (41%)	1.4 (0.98-2.1)
Abnormal PV (%)	14109 (38%)	16 (70%)	3.7 (1.5-9.1)

4.4.3 Multiple imputation

There is no formal way to evaluate whether the data are MAR or MNAR (Missing not at random). MNAR means that even if the observed data are taken into account, systematic differences between the missing values and the observed values still remain.

In this case, it is reasonable to assume that the data are MAR for the models with the full blood count components because the test was done but for some reason not all the results are available. This might be due to non-clinical reasons, such as a broken tube or an insufficient volume of blood was collected. In addition, it is possible that in the early 2000's automatic systems for information entering were not that advanced so probably there were practices that were manually entering the results in the system which might result in omissions or mistakes. All of the above suggest that the MAR assumption is quite reasonable for the first two models. With respect to the final model which uses all selected tests I cannot be certain whether the data are MAR since the tests were not ordered in the first place, most likely depending on the patient's clinical presentation.

Imputation diagnostics plots are displayed in appendix 14. The distribution of the imputed values matches very well the distribution of the completed variables suggesting that the process worked quite well even for variables for which there was a big proportion of missing data.

4.4.4 Full blood count model (model 1)

Table 4 displays the results for model 1 along with the different sensitivity analyses. The complete case analysis suffers from a very large reduction in power due to the missingness in the BMI variable. In total 144 cases are observed in the complete case analysis sample (out of 516). The symptom variables that were retained in the model included back pain (HR: 2.3; 95% CI: 1.9-2.9), chest pain (HR: 1.7; 95% CI: 1.3-2.3), rib pain (HR: 3.1; 95% CI: 1.6-6.1), and nosebleeds (HR: 2.2; 95% CI: 1.3-3.8). Lower BMI was associated with myeloma (HR: 0.97; 95% CI: 0.94-0.99) which might be related to weight loss although the weight loss predictor (based on read codes for weight loss) itself was not selected for the final model. MGUS was still predictive for myeloma (HR: 26.9; 95% CI: 17.3-41.9) although the relationship was much weaker compared to the univariate analysis. All the components of full blood count were predictive of myeloma but all of them required a fractional polynomial transformation along with age as the linearity assumption of the Cox proportional hazards model was not met. No interactions with age were included in the model as none of them was found to be significant. The complete case analysis and the imputation model produced very similar results. The largest contributor to missing data in this particular model was BMI so I conducted a sensitivity analysis by removing the BMI variable and the results were again very similar with the complete case analysis and imputation.

Hazard ratios for the fractional polynomials cannot be interpreted in the same way as hazard ratios for linear relationships and thus visualisation is better for explaining them. The relationships of the variables that require fractional polynomial transformations are visually illustrated in Figure 2. The confidence intervals at the plot correspond to the amount of information available for the estimation at those particular values. For example, in haemoglobin the confidence intervals are extremely narrow around the values of 9 to 15 suggesting that most

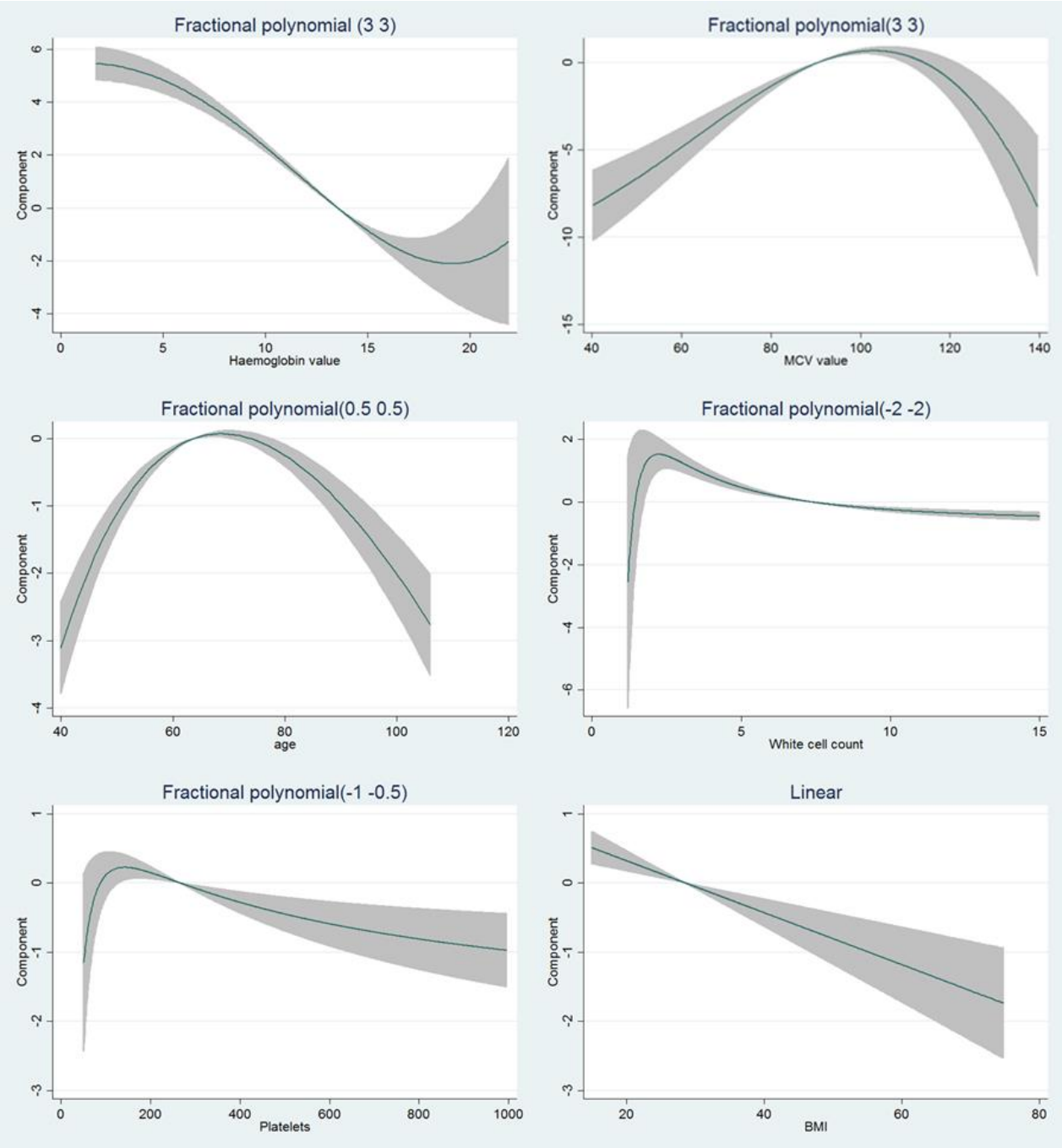
of the data were in that range. For values over 18, the confidence intervals are very wide suggesting that very few data points were available for these values. In addition, a plot is provided in appendix 15 where the predicted probabilities for the fractional polynomial transformations are compared with probabilities obtained assuming a linear relationship. Decreasing haemoglobin is associated with increased risk of myeloma although the risk starts to plateau at very low levels (lower than 5 g/dL); for MCV the risk is increased up until a value of approximately 110 fL and starts to decrease for lower values. Decreasing WCC is associated with higher risk of myeloma and the risk starts decreasing at very low values (less than 2). A similar pattern is observed for platelets where the risk peaks at approximately 130 10*9/L and then starts to drop off. The risk also increases with age which maximises around the age of 75 years after which it decreases again.

Table 4: Multivariate analysis for model 1 (demographic, risk factors, symptoms, full blood count) for complete case and imputed version. The table shows the variables selected in each case, the hazard ratios for the selected variables and the fractional polynomial selected for the continuous variables. For the continuous variables not requiring a transformation the hazard ratio is presented.

	Complete case MFP method (n=266399)	**Complete case MFP method -excluding BMI (n=744031)***	**Imputation-MFP method (n=836056)**
	HR (95% CI)	**HR (95% CI)**	**HR (95% CI)**
Demographics			
Female	0.40 (0.29-0.57)	0.44 (0.36-0.55)	0.44 (0.36-0.53)
Age	FP(0.5 3)	FP(0.5 0.5)	FP(0.5 0.5)
Risk factors			
BMI	0.95 (0.92-0.99)	NA**	0.97 (0.94-0.99)
MGUS	38.7 (18.8-79.41)	27.7 (16.9-45.1)	26.9 (17.3-41.9)
Symptoms			
Back pain	2.21 (1.43-3.41)	2.32 (1.81-2.96)	2.33 (1.86-2.91)
Chest pain	Not selected	1.45 (1.1-2.0)	1.73 (1.32-2.28)
Bone pain	2.71 (1.19-6.2)	Not selected	Not selected
Rib pain	5.91 (2.41-14.49)	3.40 (1.68-6.84)	3.13 (1.61-6.07)
Nosebleeds	Not selected	Not selected	2.22 (1.30-3.79)
Full blood count			
Haemoglobin	FP(3 3)	FP(3 3)	FP(3 3)
White cell count	FP(-2 -2)	FP(-2 -2)	FP(-2 -2)
Platelets	Not selected	Linear	FP(-1 -0.5)
MCV	FP(3 3)	FP(3 3)	FP(3 3)

*Sensitivity analysis model: For this model BMI was excluded as it was the variable with the most missing data so we examined the robustness of the imputation model by excluding it.
**Not applicable

Figure 2: Relationships between the transformed variables and the outcome. These relationships were generated only in the first imputation dataset for illustration purposes.



4.4.5 Difference in blood test parameters models (model 2)

Model 2 fitted differences in the full blood count parameters, symptoms, risk factors and demographics. Table 5 displays results for the complete case analysis, complete case excluding

BMI and the imputation model. Amongst the full blood count parameters, only the difference in haemoglobin was found to be predictive (HR: 0.88, 95% CI: 0.80-0.95) i.e. a one-point increase in haemoglobin from the first to the second test (index) is associated with a 12% decrease in the risk of myeloma. By inverting the hazard ratio (which becomes 1.14) the interpretation changes to one unit decrease in haemoglobin from the first test to the second is associated with a relative increase in the risk of myeloma of approximately 14%. In this model, the same symptoms were retained as in the previous model (back pain, chest pain, rib pain, and Nosebleeds) with very similar hazard ratios. Other variables included in the model were gender, BMI, MGUS and age which required a fractional polynomial transformation (similar to the one in the previous model). Sensitivity analysis included the absolute values plus the change but these were either not significant or suffered from serious collinearity problems. An example would be haemoglobin change and the absolute value of haemoglobin which resulted in change in the direction and significance of the association for haemoglobin change suggesting that an increase in haemoglobin levels over time would increase the risk of myeloma which doesn't make sense from a clinical point of view (HR for absolute haemoglobin value: 0.61, 95% CI: 0.58-0.64; HR for change in haemoglobin: 1.06, 95% CI: 0.99-1.13). Similar problems were observed for the other full blood count parameters which included the white cell count, platelets and MCV (HR for absolute white cell count: 0.83, 95% CI: 0.79-0.87; HR for white cell count change: 1.15, 95% CI: 1.08-1.23; HR for absolute platelet value: 0.998, 95% CI: 0.997-0.999; HR for platelet change: 0.999, 95% CI: 0.997-1.00; HR for MCV absolute value: 1.06, 95% CI: 1.05-1.07; HR for MCV change: 0.96, 95% CI: 0.94-0.99). The above associations remained unchanged irrespectively of whether the blood test variables were being included as binary or as fractional polynomials. For the above reasons only the change variables were included in the primary model in order to examine if they can be of any use without the absolute values. More details on why this might be occurring will be described more extensively in the strengths and limitations section.

Table 5: Multivariate analysis for model 2 (demographic, risk factors, symptoms, difference in full blood count parameters) for complete case and imputed version. The table shows the variables selected in each case, the hazard ratios for the selected variables and the fractional polynomial selected for the continuous variables. For the continuous variables not requiring a transformation the hazard ratio is presented.

	Complete case MFP method (n=249187)	**Complete case MFP method -excluding BMI (n=693773)****	**Imputation-MFP method (n=836056)**
	HR (95% CI)	**HR (95% CI)**	**HR (95% CI)**
Demographics			
Female	0.59 (0.42-0.84)	0.66 (0.54-0.81(	0.61 (0.52-0.73)
Age	FP(3 3)	FP(0.5 0.5)	FP(0.5 1)
Risk factors			
BMI	0.93 (0.90-0.97)	NA*	0.95 (0.92-0.97)
MGUS	51.1 (22.43-116.6)	34.8 (20.0-60.7)	38 (24.5-59.0)
Symptoms			
Back pain	2.18 (1.40-3.37)	2.39 (1.86-3.1)	2.28 (1.83-2.86)
Chest pain	Not selected	Not selected	1.70 (1.28-2.24)
Bone pain	2.63 (1.15-6.0)	Not selected	Not selected
Rib pain	7.83 (3.2-19.2)	4.2 (2.1-8.49)	3.2 (1.63-6.13)
Nosebleeds	Not selected	2.27 (1.17-4.39)	2.67 (1.57-4.54)
Full blood count			
Haemoglobin diff	0.71 (0.61-0.84)	0.76 (0.69-0.84)	0.88 (0.80-0.95)

*Sensitivity analysis model: For this model BMI was excluded as it was the variable with the most missing data so we examined the robustness of the imputation model by excluding it.
**Not applicable

4.4.6 All tests model (model 3)

The final model was the one including all tests (FBC, ESR, calcium and creatinine) in addition to symptoms and risk patient characteristics (table 6). Amongst the inflammatory markers only ESR and PV were eligible for inclusion in the model as CRP was not found to be associated with myeloma in the univariate analysis. I decided to include ESR as the inflammatory marker in the final model as PV had a much higher percentage of missing data which reflects clinical practice in which PV is much less commonly requested. The complete case analysis model was

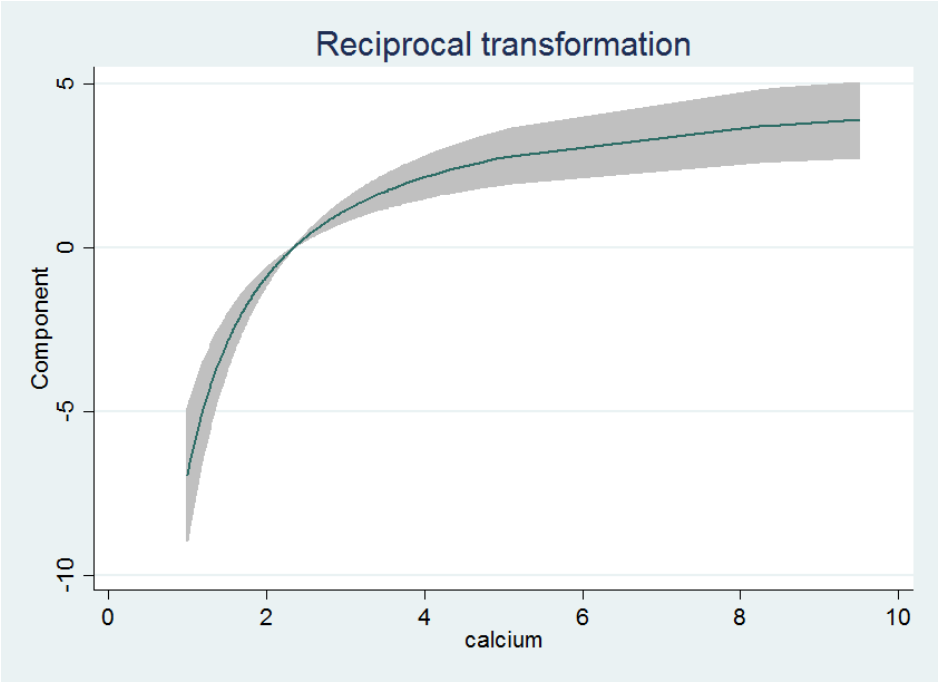
based on only 16,193 patients with calcium, creatinine and ESR tests contributing the most to missing data. This lead to a very low number of myeloma cases, significantly reducing power hence most of the variables were not selected. In the imputation model, the selected symptoms are the same as the ones selected in the previous models (back pain, chest pain, rib pain and nosebleeds). The full blood count components were all included in the model with the same fractional polynomials as in model 1. ESR was linearly associated with the outcome (HR: 1.03 95% CI: 1.026 – 1.034) suggesting that for each unit increase in ESR the risk of myeloma is increased by 3% which is very similar to the univariate analysis. Calcium was not linearly associated with the outcome and a reciprocal transformation was required (Figure 3). Creatinine, despite being significant in the univariate analysis, was not selected for the final model.

Table 6: Multivariate analysis for model 3 (demographic, risk factors, symptoms, full blood count, ESR, calcium and creatinine) for complete case and imputed version. The table shows the variables selected in each case, the hazard ratios for the selected variables and the fractional polynomial selected for the continuous variables. For the continuous variables not requiring a transformation the hazard ratio is presented.

	Complete case MFP method (n=16193)	**Imputation-MFP method (n=836056)**
	HR (95% CI)	**HR (95% CI)**
Demographics		
Female	Not selected	0.42 (0.34-0.52)
Age	Not selected	FP(0.5 0.5)
Risk factors		
BMI	Not selected	0.96 (0.94-0.99)
MGUS	Not selected	16.3 (9.72-27.3)
Symptoms		
Back pain	Not selected	2.42 (1.93-3.03)
Chest pain	Not selected	1.79 (1.35-2.36)
Bone pain	13.6 (4.4-41.5)	Not selected
Rib pain	Not selected	2.93 (1.49-5.79)
Nosebleeds	Not selected	2.14 (1.25-3.67)
Full blood count		
Haemoglobin	Linear	FP(3 3)
White cell count	FP(-2 -2)	FP(-2 -2)
Platelets	Not selected	FP(-0.5 0)
MCV	Linear	FP(3 3)
Other tests		
ESR	Linear	1.03 (1.026-1.034)
Calcium	Not selected	FP(-1)
Creatinine	Not selected	Not selected

Coefficient/Hazard ratios for fractional polynomials cannot be interpreted in the same way as linear terms.

Figure 3: Relationship of calcium with risk of myeloma



4.5 Discussion

4.5.1 Summary of results

This retrospective cohort based on CPRD was used for the development of three clinical prediction rules to assist with the diagnosis of multiple myeloma. The cohort consisted of patients who had had at least two full blood count tests within a year, were over the age of 40 and did not have a prior diagnosis of myeloma. All the algorithms included basic demographic characteristics, symptoms, and risk factors like MGUS and laboratory tests: the first model included a single full blood count only, the second included changes in full blood count parameters between the first and second test and the third and final model included all tests currently being done when myeloma is suspected including the full blood count, calcium, inflammatory markers like the ESR and creatinine.

The final models all included gender and age, back pain, chest pain, rib pain and nosebleeds. MGUS was also always included in all models. In model 1, all full blood count parameters (haemoglobin, platelet, MCV, WCC) were included. Model 2, which examined changes in full blood count parameters, indicated that only changes in haemoglobin are predictive of myeloma. Model 3 included ESR and the full blood count parameters. Overall model 1 included 12 predictors, model 2 included 9 predictors and model 3 included 14 predictors.

4.5.2 Strengths and limitations

To my knowledge this is the largest retrospective cohort study to date that has set out to develop a clinical prediction rule for diagnosing myeloma in primary care. The derivation dataset consisted of 835 540 patients from 8 different regions in the UK and 430 practices selected on the basis of having full blood count tests. This makes the study representative of people attending General Practice in the UK. A study of this kind can only be done with a sample of this size which is perhaps one of the reasons there aren't any clinical prediction rules for myeloma. A prospective study on myeloma was able to identify 71 incident cases but in order to do that they had to recruit almost 78,000 patients which further highlights the point

mentioned above (17). In addition, there is a good mixture of regions in the derivation cohort ranging from south to north (figure 1).

The selection was based on patients that had two full blood counts within a period of one year. This is a different design compared to other derivation studies of clinical prediction rules for cancers as these studies mainly focused on symptoms (18,19). A symptom-only approach is not working for myeloma due to how common these symptoms are in non-diseased patients (20). This is further backed up by the results of this study which indicate that 50% of myeloma symptoms did not have a record of myeloma related symptoms. This suggests that a symptom plus blood test approach might be more useful than a symptom only based approach. Another design that I considered initially was selecting patients that had a full blood count, calcium, creatinine and an inflammatory marker i.e. all tests that are ordered when myeloma is suspected. This approach though would probably be more prone to selection bias as I would be selecting a very high risk population, potentially patients who are already suspected of having myeloma. This is further highlighted by the high rate of missing data in the other tests as only a very small fraction of patients had all tests ordered. Other reasons for selecting the full blood count and not the other tests are described in the introduction but briefly include the early manifestation of abnormalities (compared to the other tests) and the fact that all the components can be predictive.

Two full blood counts were selected in order to investigate change from the previous full blood in this test. It could be argued that this might lead to selection bias but feasibility counts indicated that the prevalence of myeloma was very similar independently of selecting patients on one or two full blood counts. This suggests that any generated algorithm could be applied on patients who have at least one full blood count. Moreover from the data, it seems that the change from the previous FBC is not as predictive as the absolute values of the test at least when considering two full blood counts within a one year period. It is possible that changes over time need to be examined using more than two full blood counts and in a period longer than one year. The two vs one full blood count within one year argument will be investigated

even further in the validation analysis where the full blood count only algorithm will be applied on the first test and examine how it performs.

One of the major strengths of the study is the robustness of the analysis methods used including multiple imputation and fractional polynomials. The continuous nature of variables was retained resulting in more accurate predictions and more power. Although from a clinical point of view it makes more sense to know positive predictive values (i.e. for anaemia or leukopenia) a clinical prediction rule works better with continuous variables. For example the positive predictive value of anaemia is approximately 0.2% which is similar to the positive predictive value identified in other studies (20). Using continuous haemoglobin can result in a much wider range of probabilities ranging from 0.006 to 2.7% making predictions much more accurate. More complex and specific scenarios will be shown in the validation chapter. Non-linear relationships were identified so fractional polynomials were used to model these relationships as the usual transformations were not adequate. A disadvantage of using this method is that these polynomials might be prone to overfitting when several variables are considered but this will be investigated in the next chapter when the calibration of the models will be examined in the validation dataset (21). Initial assessment in the derivation set indicated that the assumption of linearity could lead to over-prediction. Estimated probabilities based on linear relationships ranged from 0.0003% to 10% (haemoglobin) which was much higher than the positive predictive value (see appendix 14).

The greatest limitation of this study is the extent of missing data in the test results. Each of the three models were affected by this this to some extent but model 3 was particularly susceptible. With the exception of the full blood count, tests had less than 50% complete data. Although the imputation diagnostics indicated the imputation procedure worked well, it is not clear whether the data for these tests were missing at random or missing not at random. In addition 10 imputations were used which are not enough for model 3 with the amount of missing data present. Doing more than 10 imputations though would be quite inefficient both in terms of time and computing power as already three weeks were required for generating 10 imputed

datasets using the pmm procedure. Model 3 will be explored in the validation dataset but all inferences regarding this model should be interpreted with a lot of caution.

An additional limitation that could have affected the imputation procedure is that the year of the test was not included in the imputation models thus the trends in missing data for laboratory tests over time were not taken into account which could have biased the imputed values. Test recording seems to have increased over time as shown in section 4.4.1. This could potentially be attributed to two reasons. First is that test results in the early 2000's required manual entry into the systems so recording was not so good early in the study and second is that the overall test usage has been increasing over time (1). Inclusion of this variable in the imputation model would have made the MAR assumption more plausible as it is linked with the missing data mechanism. Due to the very slow imputation procedure which took substantial time it was not possible to repeat the process by including the test year in the imputation models but the imputation diagnostics indicated that the procedure still worked quite well.

The other limitation is related to the coding of symptoms, similarly to the limitations in the previous chapter, which is about under-reporting of symptoms from patients or under-recording from clinicians which can influence measures of association (6). Also some of the symptoms that were predictive of myeloma in the case-control study were not selected either in the univariate or the multivariate analysis which could be attributed to not having enough power to detect some effects especially when the differences are too small.

4.5.3 Comparison with existing literature and finding from previous chapter

The two year incidence of myeloma in this cohort was 0.06% which is far higher than the incidence 0.0055% reported in Velez (22). The may be due to the selection that was based on patients who having at least one full blood counts within a year.

ESR and PV were the only inflammatory markers that were found to be predictive of myeloma while CRP was not. This is the same conclusion as the previous chapter and although it could

be argued that this is happening because of using the same database (CPRD) the fact that it was confirmed both in a case-control and a cohort design reinforces the results.

With the exception of BMI and ESR, none of the continuous variables was found to have a linear relationship with myeloma. In the cohort study, the risk of myeloma peaked at approximately age 75 and then dropped off, which is something that has been observed in other studies for myeloma (23). This pattern was also observed in most of the blood test variables. In most cases the risk reaches a peak and then starts going down, usually for the very extreme values. This might be because these very extreme values, e.g. extremely low haemoglobin, are not related to myeloma but to other conditions such as severe bleeding or even other cancers (24). These most extreme values were occurring in less than 1 % of the total sample for most tests although population-wise this can translate to a lot of patients.

In terms of changes in the full blood count parameters over time, only a change in haemoglobin was predictive for myeloma. In most cases, based on the magnitude of the coefficients for the different blood tests, it seems that the absolute value of the test is more important than the change. Inclusion of absolute values and change was leading to serious collinearity problems.

4.5.4 Developed clinical prediction rules

Table 7 illustrates the variables included in the final models and boxes 1-3 present the mathematical expressions of the developed prediction rules. The baseline risk values which correspond to the risk or failure rate obtained when all model predictors are set to 0 were estimated to be 0.9994327 for model 1, 0.9992894 for model 2 and 0.999572 for model 3.

Table 7: Final models

Model 1 *	Model 2**	Model 3***
Patient characteristics		
Age	Age	Age
Gender	Gender	Gender
BMI	BMI	BMI
MGUS	MGUS	MGUS
Symptoms		
Nosebleeds	Nosebleeds	Nosebleeds
Back pain	Back pain	Back pain
Chest pain	Chest pain	Chest pain
Rib pain	Rib pain	Rib pain
Blood test		
Haemoglobin	Change in Haemoglobin	Haemoglobin
White cell count		White cell count
Platelets		Platelets
Mean corpuscular volume		Mean corpuscular volume
		Calcium
		ESR

* Model 1 included a single FBC within a year
** Model 2 included the change between 2 full blood counts within a year
*** Model 3 included all tests (FBC, calcium, creatinine calcium, ESR)
Only the predictors selected in the final models are displayed

Box 1: Clinical prediction rule including demographics, risk factors, symptoms and single full blood count components

$$\begin{aligned}\Pr(MM) = 1 - 0.9994327^{\wedge}\exp\Bigg(& (0.845718 \cdot back\ pain) + (0.550535 \cdot chest\ pain) + (1.141665 \cdot rib\ pain) + (0.798659 \cdot Nosebleeds) \\ & + (-0.826708 \cdot gender) + (3.292089 \cdot MGUS) + (-0.034105 \cdot (BMI - 28.54623254)) \\ & + \left(78.11354 \cdot \left(\left(\frac{age}{10}\right)^{.5} - 2.523984838\right)\right) + \left(-19.89036 \cdot \left(\left(\frac{age}{10}\right)^{.5} \cdot ln\left(\left(\frac{age}{10}\right)\right) - 4.673606883\right)\right) \\ & + \left(11.50466 \cdot \left(\left(\frac{MCV_{value2}}{100}\right)^{3} - .7441717089\right)\right) + \left(-31.70556 \cdot \left(\left(\frac{MCV_{value2}}{100}\right)^{3} \cdot ln\left(\left(\frac{MCV_{value2}}{100}\right)\right) + .0732968152\right)\right) \\ & + \left(0.009802 \cdot \left(\left(\frac{WCC_{value2}}{100}\right)^{-2} - 185.5122661\right)\right) + \left(0.002337 \cdot \left(\left(\frac{WCC_{value2}}{100}\right)^{-2} \cdot ln\left(\left(\frac{WCC_{value2}}{100}\right)\right) + 484.4765069\right)\right) \\ & + \left(-4.373265 \cdot \left(\left(\frac{plat_{value2}}{100}\right)^{-1} - .3763220186\right)\right) + \left(7.489624 \cdot \left(\left(\frac{plat_{value2}}{100}\right)^{-.5} - .6134509097\right)\right) \\ & + \left(-3.332689 \cdot \left(\left(\frac{hgb_{value2}}{10}\right)^{3} - 2.484184196\right)\right) + \left(3.474869 \cdot \left(\left(\frac{hgb_{value2}}{10}\right)^{3} \cdot ln\left(\left(\frac{hgb_{value2}}{10}\right)\right) - .7534897614\right)\right)\Bigg)\end{aligned}$$

Box 2: Clinical prediction rule including demographics, risk factors, symptoms and changes in full blood count parameters

$$\Pr(MM) = 1 - 0.9992894\text{^}\exp\Bigg((0.826839 * back\ pain) + (0.529109 * chest\ pain) + (1.149765 * ribpain) + (0.982094 * Nosebleeds) + (-0.486149 * gender) + (3.638081 * MGUS) + \left(32.92934 * \left(\left(\frac{age}{10}\right)^{.5} - 2.523984838\right)\right) + \left(-6.010851 * \left(\left(\frac{age}{10}\right) - 6.370499464\right)\right) + (-0.056842 * (BMI - 28.54623254)) + \left(-0.134034 * \left(hgb_{diff} - .0135999507\right)\right)\Bigg)$$

Box 3: Clinical prediction rule including demographics, risk factors, symptoms, full blood count components, Erythrocyte sedimentation rate (ESR) and calcium

$$\Pr(MM) = 1 - 0.999572\text{^}\exp\Bigg((.8899983 * back\ pain) + (.5855513 * chest\ pain) + (1.107998 * rib\ pain) + (.752799 * Nosebleeds) + (-.7600712 * gender) + (2.797007 * MGUS) + (-.0384539 * (BMI - 28.54623254)) + \left(76.09003 * \left(\left(\frac{age}{10}\right)^{.5} - 2.523984838\right)\right) + \left(-19.50114 * \left(\left(\frac{age}{10}\right)^{.5} * ln\left(\left(\frac{age}{10}\right)\right) - 4.673606883\right)\right) + \left(11.05669 * \left(\left(\frac{MCV_{value2}}{100}\right)^{3} - .7441717089\right)\right) + \left(-31.08975 * \left(\left(\frac{MCV_{value2}}{100}\right)^{3} * ln\left(\left(\frac{MCV_{value2}}{100}\right)\right) + .0732968152\right)\right) + \left(.0125102 * \left(\left(\frac{WCC_{value2}}{100}\right)^{-2} - 185.5122661\right)\right) + \left(.0030235 * \left(\left(\frac{WCC_{value2}}{100}\right)^{-2} * ln\left(\left(\frac{WCC_{value2}}{100}\right)\right) + 484.4765069\right)\right) + \left(-8.134492 * \left(\left(\frac{plat_{value2}}{100}\right)^{-.5} - .6134509097\right)\right) + \left(-3.554435 * \left(ln\left(\left(\frac{plat_{value2}}{100}\right)\right) - .9773100698\right)\right) + \left(-2.500354 * \left(\left(\frac{hgb_{value2}}{10}\right)^{3} - 2.484184196\right)\right) + \left(2.688383 * \left(\left(\frac{hgb_{value2}}{10}\right)^{3} * ln\left(\left(\frac{hgb_{value2}}{10}\right)\right) - .7534897614\right)\right) + \left(-14.33565 * \left(cal_{value}^{-1} - .4278569248\right)\right) + (.0287225 * (ESR_{value} - 18.63286954))\Bigg)$$

4.5.5 Steps for the next chapter

The three algorithms that were generated in this chapter are going to be carried forward to the next chapter where they are going to be validated in terms of discrimination and calibration. Their clinical utility will be examined and compared to existing strategies like the ones currently being used by NICE guidelines. If a useful algorithm is identified, then implementation processes will be explored for that algorithm. In addition to the three main derived algorithms the performance of an algorithm based only on symptoms and demographic characteristics will be examined as a proof of principle to what other studies have shown that symptoms are not enough to diagnose myeloma.

References

1. Busby J, Schroeder K, Woltersdorf W, Sterne JAC, Ben-Shlomo Y, Hay A, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. Br J Gen Pract. 2013;63(609):e256-66.

2. Marshall T, Lancashire R, Sharp D, Peters TJ, Cheng KK, Hamilton W. The diagnostic performance of scoring systems to identify symptomatic colorectal cancer coampared to current referral guidance. Gut. 2011;60(9):1242–8.

3. Rutjes AWS, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PMM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005;51(8):1335–41.

4. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016;76:175–82.

5. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691–8.

6. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827–36.

7. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol. 1999;28(5):964–74.

8. Wong ES, Wang BCM, Garrison LP, Alfonso-Cristancho R, Flum DR, Arterburn DE, et al. Examining the BMI-mortality relationship using fractional polynomials. BMC Med Res Methodol. 2011;11:175.

9. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in

predictive models. J Clin Epidemiol. 2007;60(9):979.

10. Schafer JL. Multiple imputation: A primer. Vol. 8, Statistical Methods in Medical Research. 1999. p. 3–15.

11. White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med. 2009;28(15):1982–98.

12. Little RJA. Missing-data adjustments in large surveys. J Bus Econ Stat. 1988;6(3):287–96.

13. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Med Res Methodol. 2014;14:75.

14. Eddings W, Marchenko Y. Diagnostics for multiple imputation in Stata. Stata J. 2012;12(3):353–67.

15. Morris TP, White IR, Carpenter JR, Stanworth SJ, Royston P. Combining fractional polynomial model building with multiple imputation. Stat Med. 2015;34(25):3298–317.

16. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? Stat Med. 2008;27(17):3227–46.

17. Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, Hayes RB, et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. Blood. 2009;113(22):5412–7.

18. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: Independent and external validation of QCancer?? (Ovarian) prediction model. Eur J Cancer Care (Engl). 2013;22(4):423–9.

19. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify men with suspected cancer in primary care: Derivation and validation of an algorithm. Br J Gen Pract. 2013;63(606):1–10.

20. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: A large case-control study using electronic records. Br J Gen Pract. 2015;65(631):e106–13.

21. Royston P, Sauerbrei W. Stability of multivariable fractional polynomial models with selection of variables and transformations: A bootstrap investigation. Stat Med. 2003;22(4):639–59.

22. Vélez R, Turesson I, Landgren O, Kristinsson SY, Cuzick J. Incidence of multiple myeloma in Great Britain, Sweden, and Malmö, Sweden: The impact of differences in case ascertainment on observed incidence trends. BMJ Open. 2016;6(1):e009584.

23. Turesson I, Velez R, Kristinsson SY, Landgren OLA. Patterns of multiple myeloma during the past 5 decades: Stable incidence rates for all age groups in the population but rapidly changing age distribution in the clinic. Mayo Clin Proc. 2010;85(3):225–30.

24. Hamilton W, Lancashire R, Sharp D, Peters TJ, Cheng KK, Marshall T. The importance of anaemia in diagnosing colorectal cancer: a case-control study using electronic primary care records. Br J Cancer. 2008;98(2):323–7.

Chapter 5: Statistical validation, clinical utility and implementation of diagnostic prediction rules for early diagnosis of myeloma

5.1 Chapter Introduction

The clinical prediction rules that were developed in the previous chapter aim to alert the clinician to the possibility of myeloma as a potential cause of the patient's symptoms (back pain, mild anaemia), so that either a protein electrophoresis is ordered or, when the risk is high enough, refer the patient direct to haematology, and to identify disease progression in people with the pre-cursor condition MGUS. The performance of the models appears sufficient to be considered for further evaluation.

When developing a clinical prediction rule whether it is prognostic or diagnostic it is not sufficient to demonstrate that it works well in the dataset that was used for the development (1,2). The performance of the model should also be examined in new datasets in order to avoid estimation of over-optimistic performance measures. There are various reasons that might lead to over optimistic estimates of performance which might include overfitting or absence of importance predictors from the model. Overfitting occurs when there are very few event outcomes relative to the number of candidate predictors that need to be estimated (3,4). Absence of an important predictor is hard to know but sometimes it can occur due to improper use of predictor selection methods (5).

There are two main types of validation procedure, internal and external validation. Internal validation involves splitting the dataset into two parts called the "development" and "validation" datasets(6). The model is generated using the development dataset and validated in the validation dataset. Ideally a non-random splitting should be done in order to reduce the similarities between the two datasets and avoid optimistic estimates. This could include splitting the data by geographical region or by study centre if the sample size is sufficiently large. Another option is to split the dataset by time period which is also known as temporal validation. If the data are limited and a split sample approach is not possible then other techniques like

cross-validation and bootstrapping can be used. The second type of validation, external validation, is arguably the most important type of validation as it examines the performance of the model in a new dataset, ideally in a different country or a different setting (7).

5.2 Aims

The aim of this chapter is to examine the performance of the prediction rules that were generated in chapter 4 with the use of data splitting internal validation. Throughout the chapter the terms prediction rule and model will be used interchangeably.

The three clinical prediction rules are the following

- Model 1 (M1): symptoms + single Full blood count (all components)
- Model 2 (M2): symptoms + change between two full blood count tests (haemoglobin)
- Model 3 (M3): symptoms + Full blood count (all components) + calcium + ESR

The above models will be examined in terms of discrimination, calibration and clinical utility and compared with strategies that involve single symptoms and abnormalities in tests. Based on the results a final model will be selected. Analyses will be conducted that will explore whether application of the models can potentially reduce time to diagnosis and more specifically the diagnostic interval which was explored in detail in chapters 2 and 3.

5.3 Methods

The overall study design was described in the previous chapter but as a reminder, the cohort consisted of people over the age of 40 who had two full blood count tests within a year, with the second test being the start of a two years' follow-up to identify new myeloma diagnoses. Rather than randomly splitting patients into derivation and validation datasets the splitting was done by geographical region to examine the reproducibility (the prediction rule is accurate in patients that were not included in the development stage) and the geographic robustness (the prediction rule is accurate when tested in data from different locations) (8). The derivation dataset consisted of 8 UK regions and the validation set of 5.

Similarly to the derivation set, the validation also contained missing data. The methods employed for imputation in the validation dataset are identical to those described in chapter 4 but were conducted separately for the derivation and the validation set. In both cases 10 imputation sets were generated. The models that were developed in chapter 4 were applied in each imputed dataset and the statistics of interest were estimated in each set and combined. In addition to the models the performance of the individual parameters was also examined in the validation dataset (symptoms and blood tests). The statistical methods used for combining them are described in the following sections.

5.3.1 Discrimination and calibration

Discrimination is one of the most important properties that a good clinical prediction rule needs to satisfy in order to be useful. It refers to the ability of the model to separate individuals who develop the events of interest from those who do not (9). When models are developed using the cox-proportional hazard model the most common measures of discrimination include Harrel's C statistic (10) which is also known as the area under the curve, the D-statistic and R^2(11–13). Harrell's C-statistic ranges from 0.5 to 1 with higher values corresponding to better discrimination. In this study, it denotes the probability that among two patients of whom one is diagnosed with myeloma within two years and the other is not, the model assigns a higher probability of myeloma to the patient with myeloma. The D-statistic is a measure of discrimination developed specifically for prognostic models and it quantifies the separation in observed survival curves between subgroups of patients with differing predicted risk. D statistic can take any values and the higher the value the better the separation between the curves and thus better discrimination. For the C-statistic certain rough thresholds exist which help with the interpretation i.e. a value between 0.9 and 1 is considered excellent discrimination. Unfortunately, for the D-statistic no such thresholds exist which makes interpretation less straightforward. Despite that the D-statistic is a discrimination statistic that has been increasingly used in recent years due to certain properties that make it very applicable and appealing such as robustness to outliers, independence from censoring, an easily calculated

standard error and easily computed in external validation settings (14). In addition new methodology has been developed recently which extrapolates from the D-statistic in order to estimate the sample size for clinical prediction rules that use time to event data thus it is important to report it (15). The R^2 is a measure of overall performance and it quantifies how much variation in the outcome is explained by the model factors. It is expressed as a percentage and the higher the value the better the variability is explained by the model.

A second important aspect of performance evaluation is calibration. Calibration refers to whether the probabilities derived from the model agree with the observed rates in the validation dataset (16). If for example we predict a 3% risk of developing myeloma for a particular set of patients then we would hope that the observed myeloma frequency would be around 3 per 100 patients. If a model has poor calibration then it may not be clinically useful as the model can result in overestimation (over-predicting) or underestimation (under-predicting) of the predicted probabilities. Both can affect subsequent analysis and decision-making thus it is important for models to be well calibrated. One of the most common ways to examine calibration is the use of calibration plots. In order to generate calibration plots the following steps need to be followed:

1. Individuals are put into deciles (tenths) of risk – from lowest to highest – according to their predicted probability which is obtained from the prediction rule.
2. In each decile the observed and the predicted probability of the event is estimated. The observed probability is the event rate of the outcome in that particular decile while the predicted probability is the average predicted probability in that decile.
3. To generate the calibration plot the observed probability is plotted against the predicted probability usually in a scatter plot which sometimes incorporates a smoothing line.
4. For a model to be well calibrated the predicted probabilities should be as close to the observed probabilities as possible. If the predicted probabilities are higher than the observed then the model is over-predicting and when the predicted probabilities are lower than the observed probabilities then the model under-predicts.

In the thesis I present the two types of calibration plots which are usually encountered in the literature and are going to be explained in more detail when the calibration plots are presented. These plots have similar interpretations and they results in the same conclusion.

5.3.2 Clinical utility

Sensitivity, specificity, negative and positive predictive values and likelihood ratios were estimated for each symptom and investigations. The components of the full blood count were examined separately and for specific combinations (i.e. haemoglobin and MCV which are usually interpreted together in clinical practice). In addition, combinations of back pain and blood test abnormalities like leukopenia, high ESR and high calcium were investigated as they are cited by NICE guidelines as high risk combinations (17). The combination that was identified in chapter 3 as a good rule out approach (high ESR or low haemoglobin) was also validated.

Each model generates probabilities of myeloma within the next two years. But a rule/strategy is required to make a clinical decision. This usually involves a single decision threshold (although two or more can be useful). When the predicted probability is lower than the decision threshold, myeloma would be ruled out, but if the predicted probability is high enough then further investigations (such as serum/urine protein electrophoresis) should be requested. Models such as the ones being validated in this chapter produce a wide range of predicted probabilities and they have a number of potential decision thresholds.

The performance of the models was examined at different threshold probabilities which include

1. Prevalence of the disease in the dataset
2. Relative risk thresholds, i.e. probabilities that correspond to the risk percentiles of 90%, 95%, 99.5% and 99.9%. These cut offs were selected based on other similar studies that were validating prediction rules for the diagnosis of cancer (18,19).
3. Absolute risk threshold of 3% which is the cut-off that NICE uses to send patients to the hospital through the 2-week cancer referral route (20).

Based on the probability thresholds the patients were being classified in traditional 2x2 tables and the measures mentioned above were also estimated. Similar rules as in chapter 3 are used in order to determine whether a test/symptoms can be of use (LR+ >=5 for ruling in and LR- <=0.2 for ruling out).

5.3.3 Net benefit and Decision curve analysis

Traditional measures of model performance which have been described above are useful but further work is required in order to assess the clinical value of the model. For example is an AUC curve of 0.7 good enough for a model/biomarker/test to be used in practice? Will clinicians apply a dichotomous threshold, e.g. refer patients to a fast-track cancer pathway when the risk is higher than 3%? How about sensitivity and specificity? If a certain strategy has very high sensitivity but very low specificity it means that it will be picking up a lot of false positives. How will this translate into clinical practice? Are the extra false positives that are detected by using this approach causing more harm than benefit? Or it is so important to pick up all potential cases that we are willing to accept the harm. The following simulated example illustrates the problem described above (table 1).

Table 1: Simulated example for demonstrating the net benefit approach

	D (+)	**D (-)**		**D (+)**	**D (-)**
T1 (+)	53	67	**T2 (+)**	58	99
T1 (-)	20	60	**T2 (-)**	15	28
	Sens=73%	Spec=47%		Sens=79%	Spec=22%

Making a decision based on just the values of sensitivity and specificity is not straightforward. From table 1 it can be seen that T2 has better sensitivity than T1 but worse specificity. So T2 picks up more true positives in exchange for more false positives. It is not very clear which approach is more beneficial.

Traditional decision analytic techniques can be used to answer these questions but the problem is that most of these approaches require additional data like costs or QALYS (quality-adjusted life-years) and these are usually not collected in a validation dataset (21). So when validating a model in a database the analysis usually stops after estimating calibration and discrimination.

In recent years a method called decision curve analysis (DCA) has been developed which allows for a particular strategy to be evaluated within the dataset in which the model is being validated without the use of any additional information such as patient outcomes, costs etc. (22–24). In short this type of analysis tries to examine whether a particular diagnostic (or prognostic) approach can be useful in clinical practice by trying to put the harms and benefits of that approach in the same scale where the benefits are the true positives and the harms the false positives. In the case of myeloma the benefit is to get as many true positive diagnoses as possible to expedite diagnosis and subsequent treatment, and the harms are the false positives which might result in over-testing and over-diagnosing MGUS patients.

Before I expand any further on this particular method I am going to proceed first with defining some important concepts which are necessary for the application of this technique. The first one is the definition of the probability or decision threshold which determines which patients are high risk and thus require either further investigation or treatment. An example would be a doctor ordering a biopsy if the risk of cancer for a patient is more than 5%. In the case of myeloma, a reasonable argument would be that it is important to detect the disease as early as possible meaning a haematologist would be willing to perform bone marrow biopsy on 100 patients to find one case of myeloma. This means the haematologist in this scenario is considering that missing a myeloma patient is much more harmful than doing unnecessary biopsies by a factor of 99 (in every 100 people having a bone marrow a myeloma case is equal to 99 unnecessary biopsies). This exchange rate or weight described above can be obtained by the following formula:

$$w = \frac{p}{1-p} \; where\ p\ is\ the\ threshold\ probability\ and\ w\ is\ the\ weight$$

Table 2 illustrates how the weight is modified depending on the threshold/decision probability used and how it can be interpreted.

Table 2: Weight definition

Probability thresholds	Weight p/(1-p)	Interpretation	
1%	0.01	Value of 1 TP is equal to 99 FP	True positives are more important than false positives
5%	0.053	Value of 1 TP is equal to 19 FP	
10%	0.11	Value of 1 TP is equal to 9 FP	
20%	0.25	Value of 1 TP is equal to 4 FP	
50%	1	Value of 1 TP is equal to 1 FP	Equal value of TP and FP
80%	4	Value of 4 TP is equal to 1 FP	False positives are more important than true positives
90%	9	Value of 9 TP is equal to 1 FP	
95%	19	Value of 19 TP is equal to 1 FP	
99%	99	Value of 99 TP is equal to 1 FP	

Using a low threshold probability means that true positives are more important than false positives. Cancer for example is a disease for which clinical guidelines recommend the use of low thresholds probabilities as it is important not to miss any cancer patients.

The other important definition is the net benefit which is the weighted-sum of true positive classifications which takes into account the weighted false positive classifications i.e. how many false positives for one true positive. The Net Benefit is given by the formula below

$$Net\ Benefit = \frac{TP - wFP}{N}$$

where TP corresponds to true positives, FP to false positives, N to the total number of patients and w is the weight as defined above. The maximum possible value for the net benefit is the prevalence of the disease and it can be equal to zero if the harm is equal to the benefit or it can be negative if the harm is more than the benefit.

Let's consider the example illustrated in table 1 where two hypothetical diagnostic blood tests are examined, with different sensitivities and specificities, and assume a threshold probability of

10% (i.e. weight 1/9). After estimating the net benefit according to the formulae illustrated, test 2 (T2) has a net benefit 0.8 higher per 100 people than test 1(T1). This can be interpreted as a net increase in the true positive classifications i.e. 8 more true positives are identified per 1000 population at the same number of false positive classifications, assuming a threshold probability of 10%. Thus based on this approach T2 is better than T1 when decisions are made based on a threshold probability of 10%. If the threshold changes to 40% then T2 has a lower benefit than T1 suggesting that T1 is better than T2 at the decision threshold of 40%.

The main limitation of using the net benefit method is the determination of the threshold and decision curve analysis was created to address this issue by examining the net benefit for a reasonable range for threshold probabilities.

The steps for DCA are the following

1. Determine a threshold probability p
2. Classify a patient as positive if the probability generated by the model is greater than the threshold probability. For a well–calibrated model the threshold probabilities should match the model probabilities
3. Based on the above classification the number of true and false positives is generated
4. The net benefit (NB) is calculated with the use of the following formula

$$NB = \frac{True\ positives}{N} - \left(\frac{False\ positives}{N} \times \frac{p}{1-p}\right)$$

5. The above process is repeated for a range of probability thresholds

The net benefit is then plotted against the range of threshold probabilities by the use of smoothing functions and the diagnostic (or prognostic) strategy with the highest net benefit is the best (strategy can be anything ranging from prediction rules to single tests or biomarkers). More about the interpretation of DCA along with its limitations will be shown in the results and discussion sections.

5.3.4 Combining estimates in multiple imputation

In the derivation of the clinical prediction rules (chapter 4) the final cox regression coefficients were derived by combining the coefficients from each imputed dataset using the Rubin's rule. As a reminder Rubin's rule combines the estimates by taking the average amongst all imputation datasets and estimates the standard error by taking into account the within and between variation of the imputation datasets. A similar procedure was followed in the validation for all the statistics of interest (sensitivity, specificity, likelihood ratios etc.). During the imputation procedure the statistic of interest is calculated in each imputation dataset and then the statistics are pooled in order to get an overall estimate of the statistic and an estimate of the variance (25). Estimates based on Rubin's rule can be more accurate when applied to parameters that follow a normal distribution. Parameters like model coefficients for example follow a normal distribution so combining estimates from a model is a straightforward application of Rubin's rule. If the parameter of interest is not following a normal distribution (like for example probabilities which are bounded between 0 and 1) then transformation is required before they are combined, followed by back transformation after combination. Another option would be to combine the estimates by taking the median rather than the mean which is more robust to violations of the normality assumption. Below table 3 illustrates the different measures that were estimated along with the methods used to combine them in addition to the sensitivity analysis conducted in order to check their robustness (26).

Table 3: Methods for combining parameter estimates in multiple imputation

Parameters	Main method of combining estimated	Sensitivity analysis used to confirm robustness of results
Regression coefficients	Rubin's rule	Not required
Linear predictor per patient	Rubin's rule	Not required
C-statistic	Rubin's rule	Median (range)
D statistic	Rubin's rule	Not required
R^2 statistic	Rubin's rule	Median (range)
Survival probabilities	Complementary log-log transformation, Rubin's rule and back transform to the original scale	Median
Sensitivity/specificity/Positive and negative predictive value	Rubin's rule	Median (range)
Likelihood ratios	Rubin's rule	Not required

The study was designed on the basis of the TRIPOD statement which is presented in appendix 16 for both derivation and validation chapters (27).

5.4 Results

The baseline characteristics of the validation dataset and its comparison with the derivation set can be seen in table 1 of chapter 4. Overall the characteristics of the derivation and the validation datasets are very similar making the two sets comparable. Briefly the validation dataset includes 446,986 patients from 228 practices, and 254 patients were diagnosed with myeloma in the two years after the index date. The distribution of the imputed variables in the validation dataset was very similar to the observed variables.

5.4.1 Test performance (individual and combinations)

Table 4 shows the diagnostic accuracy measures for individual symptoms, blood tests and NICE combinations. All symptoms had positive likelihood ratios lower than 5, with rib pain as the symptom with the highest likelihood ratio (3.2; 95% CI: 1.2-8.4) and the highest positive

predictive value (0.18%; 95% CI: 0.05-0.46). Back pain was the symptom with the highest sensitivity (21%, 95% CI: 16-26) followed by chest pain (9.5%, 95% CI: 6.2-13.7). No symptom had very good rule out properties with all of them having negative likelihood ratios above 0.8.

Of the full blood count, anaemia which is defined as a haemoglobin level lower than 13 g/dL for men and 12 g/dL for women, had the highest positive likelihood ratio (3.3, 95% CI: 2.8-3.9) followed by high mean corpuscular volume (MCV > 96 fL), white cell count lower than 4.5 x 10^9/L and platelets lower than 150 x 10^9/L. This was also the case when considering the rule out properties of the tests, with normal haemoglobin having the lowest negative likelihood ratio (0.54, 95% CI: 0.47-0.62) and highest sensitivity (54%; 95% CI: 48-61), followed by normal MCV (LR- 0.84, 95% CI: 0.78-0.91).

When considering the other tests, elevated ESR of at least 10 mm/hr for men and 15 mm/hr for women has very high sensitivity (80%, 95% CI: 70-90) resulting in very good rule out properties as indicated by the negative likelihood ratio (0.36, 95% CI: 0.28-0.47) while high calcium of more than 2.5 mmol/L has very high specificity (98%, 95% CI: 97-98) resulting in better rule in properties.

Amongst the combinations, back pain plus high calcium had the highest positive likelihood ratio (17, 95% CI: 6.5-27.6) followed by back pain plus low white cell counts (9.1, 95% CI: 7.0-11.1). Combining haemoglobin and MCV, because they are clinically interpreted together to define the type of anaemia, macrocytic anaemia (low haemoglobin <13 g/dL men and 12 g/dl for women, high MCV > 96 fL) had the highest positive likelihood ratio (7.7, 95% CI: 1.8-32.9) and the highest PPV (0.43%, 95% CI: 0.26-0.62). High ESR or low haemoglobin has a very high sensitivity (91%, 95% CI: 82-99) making it a very good rule out combination when both parameters are normal (LR -=0.19, 95% CI: 0.1-0.3). Very similar values were observed when examining high ESR or low haemoglobin or low calcium with a sensitivity of 91% (95% CI: 82% - 99%), specificity of 50% (95% CI: 49%-50%) and a LR- of 0.19 (0.1-0.3).

Table 4: Comparison of strategies for the diagnosis of myeloma

	Pr.[+]	Sensitivity (%)	Specificity (%)	LR+	LR-	PPV (%)	NPV (%)
Symptoms							
Back pain	NA	21 (16-26)	91.2 (91.1-91.3)	2.4 (1.9-3)	0.87 (0.82-0.93)	0.13 (0.1-0.18)	99.96 (99.94-99.96)
Rib pain	NA	1.6 (0.4-4.0)	99.5 (99.5-99.5)	3.2 (1.2-8.4)	0.99 (0.97-1)	0.18 (0.05-0.46)	99.94 (99.94-99.95)
Chest pain	NA	9.5 (6.2-13.7)	93.6 (93.5-93.7)	1.5 (1.0-2.2)	0.97 (0.93-1)	0.08 (0.05-0.13)	99.95 (99.94-99.95)
Nosebleeds	NA	1.2 (0.2-3.4)	99.2 (99.2-99.2)	1.5 (0.5-4.7)	0.99 (0.98-1.0)	0.09 (0.02-0.25)	99.94 (99.94-99.95)
FBC (index test)							
Anaemia (any type)	NA	54 (48-61)	84 (83-84)	3.3 (2.8-3.9)	0.54 (0.47-0.62)	0.19 (0.16-0.22)	99.96 (99.96-99.97)
Normocytic anaemia	NA	43 (37-49)	85 (84-85)	2.9 (2.3-3.6)	0.67 (0.59-0.75)	0.16 (0.13-0.20)	99.96 (99.95-99.96)
Macrocytic anaemia	NA	11 (8-15)	99 (98-99)	7.7 (1.8-32.9)	0.89 (0.86-0.94)	0.43 (0.26-0.62)	99.95 (99.94-99.95)
Low platelets	NA	8 (6-11)	97 (96-97)	2.4 (1.5-3.9)	0.95 (0.91-0.99)	0.14 (0.08-0.19)	99.95 (99.93-99.95)
Low WCC*	NA	7 (5-10)	97 (96-97)	2.3 (1.4-4)	0.96 (0.92-0.99)	0.15 (0.1-0.2)	99.94 (99.94-99.95)
High MCV	NA	23 (17-28)	92 (91-92)	2.7 (1.7-4.2)	0.84 (0.78-0.91)	0.13 (0.07-0.19)	99.95 (99.94-99.96)
Other tests							
Hypercalcemia*	NA	6 (2-9)	98 (97-98)	3.3 (0.5-21)	0.96 (0.92-1.0)	0.19 (0.04-0.34)	99.95 (99.94-99.95)
High ESR*	NA	80 (70-90)	55 (54-55)	1.8 (1.5-2.0)	0.36 (0.28-0.47)	0.1 (0.09-0.12)	99.97 (99.97-99.99)
NICE combinations**							
Back pain + high cal.	NA	2.6 (0.6-4.7)	99.8 (99.8-99.9)	17 (6.5-27.6)	0.98 (0.95-1.0)	1.0 (0.06-1.9)	99.95 (99.94-99.95)
Back pain + ESR	NA	16 (12-20)	96.3 (96.2-96.4)	4.3 (3.6-4.9)	0.87 (0.82-0.93)	0.24 (0.16-0.32)	99.95 (99.94-99.96)
Back pain + low WCC	NA	2 (0.07-3.4)	99.8 (99.7-99.8)	9.1 (7.0-11.1)	0.98 (0.96-0.99)	0.5 (0.07-1.0)	99.94 (99.93-99.95)
High ESR or anaemia[$]	NA	91 (82-99)	50 (49-50)	1.8 (1.7-1.9)	0.19 (0.1-0.3)	0.1 (0.09-0.12)	99.99 (99.98-99.99)
Model 1							
Prevalence	0.06	79 (74-84)	77.3 (77.2-77.4)	3.5 (3.3-3.7)	0.27 (0.21-0.34)	0.20 (0.17-0.23)	99.98 (99.98-99.99)
90th percentile	0.12	61 (54-67)	90.2 (90.1-90.3)	6.2 (5.6-6.8)	0.44 (0.38-0.51)	0.35 (0.30-0.41)	99.98 (99.97-99.98)
95th percentile	0.2	44 (38-50)	95.1 (95-95.1)	8.9 (7.7-10.2)	0.59 (0.53-0.66)	0.5 (0.41-0.61)	99.96 (99.96-99.97)
99th percentile	0.6	22 (18-28)	99 (99-99.1)	23.1 (18.4-29.1)	0.78 (0.73-0.84)	1.3 (0.99-1.7)	99.95 (99.95-99.96)
99.5th percentile	0.9	17 (12-22)	99.5 (99.5-99.5)	33.8 (25.6-44.7)	0.84 (0.79-0.89)	1.9 (1.4-2.5)	99.95 (99.95-99.96)
99.9th percentile	2.7	5.9 (3.3-9.6)	99.9 (99.9-99.9)	60.5 (36.7-99.7)	0.94 (0.91-0.97)	3.3 (1.9-5.4)	99.95 (99.94-99.95)
NICE threshold	3	5.9 (3.3-9.6)	99.9 (99.9-99.9)	71.7 (43.4-118)	0.94 (0.91-0.97)	3.9 (2.2-6.4)	99.91 (99.91-99.92)
Model 2							
Prevalence	0.06	72(66-78)	62.6 (62.4-62.7)	1.9 (1.8-2.1)	0.44 (0.36-0.54)	0.11 (0.09-0.13)	99.97 (99.97-99.98)
90th percentile	0.11	32 (26-38)	90.2 (90.1-90.2)	3.2 (2.7-3.8)	0.76 (0.70-0.83)	0.18 (0.14-0.23)	99.96 (99.95-99.96)
95th percentile	0.14	18 (13-23)	95.1 (95-95.1)	3.6 (2.8-4.7)	0.87 (0.82-0.92)	0.20 (0.15-0.27)	99.95 (99.95-99.96)
99th percentile	0.25	8.0 (5.0-12)	99.1 (99-99.1)	8.5 (5.6-12.9)	0.93 (0.90-0.96)	0.48 (0.29-0.74)	99.95 (99.94-99.95)
99.5th percentile	0.31	6.3 (3.6-10)	99.5 (99.5-99.5)	12.9 (8.0-20.7)	0.94 (0.91-0.97)	0.73 (0.42-1.18)	99.95 (99.94-99.95)
99.9th percentile	1.12	4.7 (2.5-8.1)	99.9 (99.9-99.9)	48.6 (27.8-85.2)	0.95 (0.93-0.98)	2.7 (1.4-4.7)	99.95 (99.94-99.95)
NICE threshold	3	1.6 (0.4-4.0)	99.9 (99.9-99.9)	38.7 (14.5-103)	0.99 (0.97 – 1)	2.2 (0.6-5.4)	99.94 (99.9-100)
Model 3							
Prevalence	0.06	82 (76-86)	83.9 (83.7-84)	5.1 (4.8-5.4)	0.22 (4.8-5.4)	0.29 (0.25-0.33)	99.98 (99.98-99.99)
90th percentile	0.09	72 (66-77)	90.3 (90.2-90.4)	7.4 (6.8-8.0)	0.31 (0.26-0.38)	0.42 (0.36-0.48)	99.98 (99.97-99.98)
95th percentile	0.15	61 (55-67)	95.1 (95-95.2)	12.5 (11.4-13.8)	0.41 (0.35-0.47)	0.71 (0.60-0.83)	99.98 (99.97-99.98)
99th percentile	0.46	36 (30-43)	99 (99-99)	36.5 (31-43.1)	0.64 (0.59-0.71)	2.0 (1.6-2.5)	99.96 (99.96-99.97)
99.5th percentile	0.74	25 (20-31)	99.5 (99.5-99.5)	51.6 (41.5-64)	0.75 (0.71-0.81)	2.9 (2.2-3.6)	99.96 (99.95-99.96)
99.9th percentile	2.23	9.5 (6-14)	99.9 (99.9-99.9)	98.6 (66.6-146)	0.91 (0.87-0.94)	5.3 (3.4-7.8)	99.95 (99.94-99.95)
NICE threshold	3	6.7 (4.0-11)	99.9 (99.9-99.9)	108 (67.2-173)	0.93 (0.90-0.97)	5.8 (3.4-9.1)	99.95 (99.94-99.95)

The results presented are based on multiple imputation as described in the methods section

* Leukopenia, abnormal inflammatory marker and high calcium are high risk features according to NICE

** There is a particular emphasis on back pain in NICE guidelines so combinations with back pain are examined

[+] Pr. is the corresponding probability (%) of the selected risk score percentile

$ Combination identified in chapter 3

5.4.2 Model validation - Discrimination and Calibration

Table 5 illustrates the discrimination statistics for each model. Overall model 1 and model 3 have the best discrimination with model 3 having a statistically significant higher AUC statistic than model 1 (0.88 compared to 0.84, $p<0.0001$).

The D statistic shows similar patterns as the AUC statistic with both model 1 and model 3 having values greater than 2.

Model 2 does not seem to perform very well. The AUC statistic for model 2 is approximately 0.7 and the R^2 is approximately 30% while the other two models have an R^2 over 50% with the best one being model 3.

Performance measures for apparent validation which is the performance of the clinical prediction rules on the dataset used to create the prediction rules are similar to that of the validation dataset although slightly better.

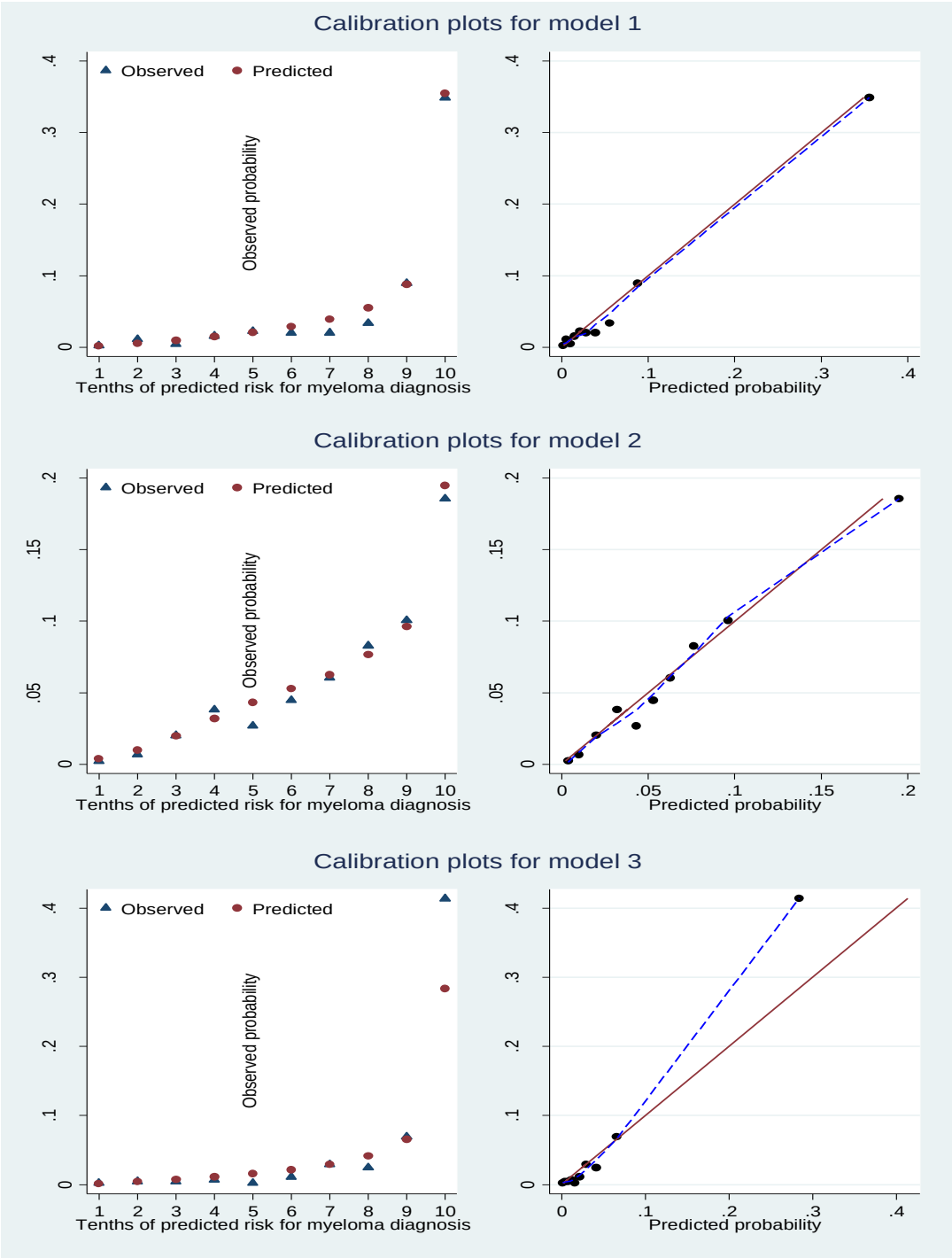
Table 5: Model performance measures

	AUC (95% CI)	R^2 statistic (95% CI)	D-statistic (95% CI)
Model 1	0.84 (0.82-0.87)	0.57 (0.53-0.62)	2.4 (2.2-2.6)
Model 2	0.73 (0.70-0.77)	0.32 (0.26-0.39)	1.4 (1.2-1.6)
Model 3	0.88 (0.85-0.91)	0.64 (0.59-0.68)	2.7 (2.5-3.0)

Figure 1 illustrates the calibration plots that were generated for the three models. Model 1 and model 2 show very good calibration. Some minor over- and under-prediction is occurring across the spectrum for model 1 and model 2. The calibration for model 3 is not very good though, especially in the last decile where substantial under-prediction is observed. The reasons for this will be further explained in the later sections of the chapter. A sensitivity analysis was performed for all three clinical prediction rules where the MGUS patients were removed in order to examine how they perform on patients who have not been identified with the

asymptotic form of the condition and both the validation statistics and the calibration plots were the same.

Figure 1: Calibration plots for models 1-3. The left panel illustrates a calibration plot where the observed and predicted probabilities are displayed for each decile (tenth) of risk. For good calibration the predicted and observed probabilities should be as close as possible. The right panel shows another calibration plot that is also encountered in the literature which is a scatter plot of observed vs predicted probabilities. The red line corresponds to the line of perfect calibration and the optimal scenario is for all points to lie on that line. A smoothed line (blue) is fitted through the points. If the blue line above the red then the model under-predicts myeloma risk and if the blue line is below the red then the model over-predicts the risk.



5.4.3 Comparison of different strategies for diagnosis of myeloma

For each model, the estimated probabilities were divided in percentiles, and the corresponding probability of myeloma for each percentile was calculated (table 4). The 90th percentile for model 1 corresponded to a myeloma probability of 0.12%. Patients with a risk of at least 0.12% are classified as patients with the highest 10% probability of being diagnosed with myeloma within the next 2 years. The sensitivity corresponding to this particular threshold is 61%, the likelihood ratio 6.2 (95% CI: 5.6-6.7) and the PPV 0.35% (95% CI: 0.30-0.41). It's important to note that the PPV in this case corresponds to the average myeloma probability in the group with a probability greater than 0.12%. Model 1 also has good rule out properties at this threshold with the negative likelihood ratio being 0.44 (95% CI: 0.38-0.51) which is better than the individual components of the full blood count but not better than ESR (as a rule out test). The top 99.9th percentile of this model corresponds to a myeloma probability of 2.7% and by using this threshold we will pick only patients that have a very high risk of myeloma, corresponding to very high specificity but low sensitivity: sensitivity is 6% (95% CI: 3.3-9.6), the specificity 99.9% (95% CI: 99.9-99.9) and the positive likelihood ratio 60.5 (95% CI: 36.7-99.7).

Model 2 performs less well than model 1. The probability corresponding to the 90th percentile is 0.11% and the sensitivity at this cut-off is only 32% which means that it picks only half of the patients that model 1 picks at the same percentile of risk. Additionally, the model does not have very good rule out properties as the negative likelihood ratio is 0.76 (95% CI: 070-0.83). This is in-line with the discrimination statistics which suggested that model 1 is better than model 2

Model 3 seems to be the best model from a statistical point of view. The probability corresponding to the 90th percentile is 0.09% and the sensitivity for this cut-off is 72% (95% CI: 66-77). The positive likelihood ratio is 7.4 (95% CI: 6.8-8) and the PPV 0.42% (95% CI: 0.36-0.48). This is also the model that has the best rule out properties with a negative likelihood ratio of 0.31 (95% CI: 0.26-0.38). The 99.9th percentile which corresponds to a myeloma probability of 2.2% is able to identify 9.5% of myeloma patients and the positive likelihood ratio is 98.6

(95% CI: 66.6-146).

5.4.4 Comparison of different strategies using net benefit approaches

Figure 2 displays decision curve analysis plots where the models are compared with different diagnostic strategies.

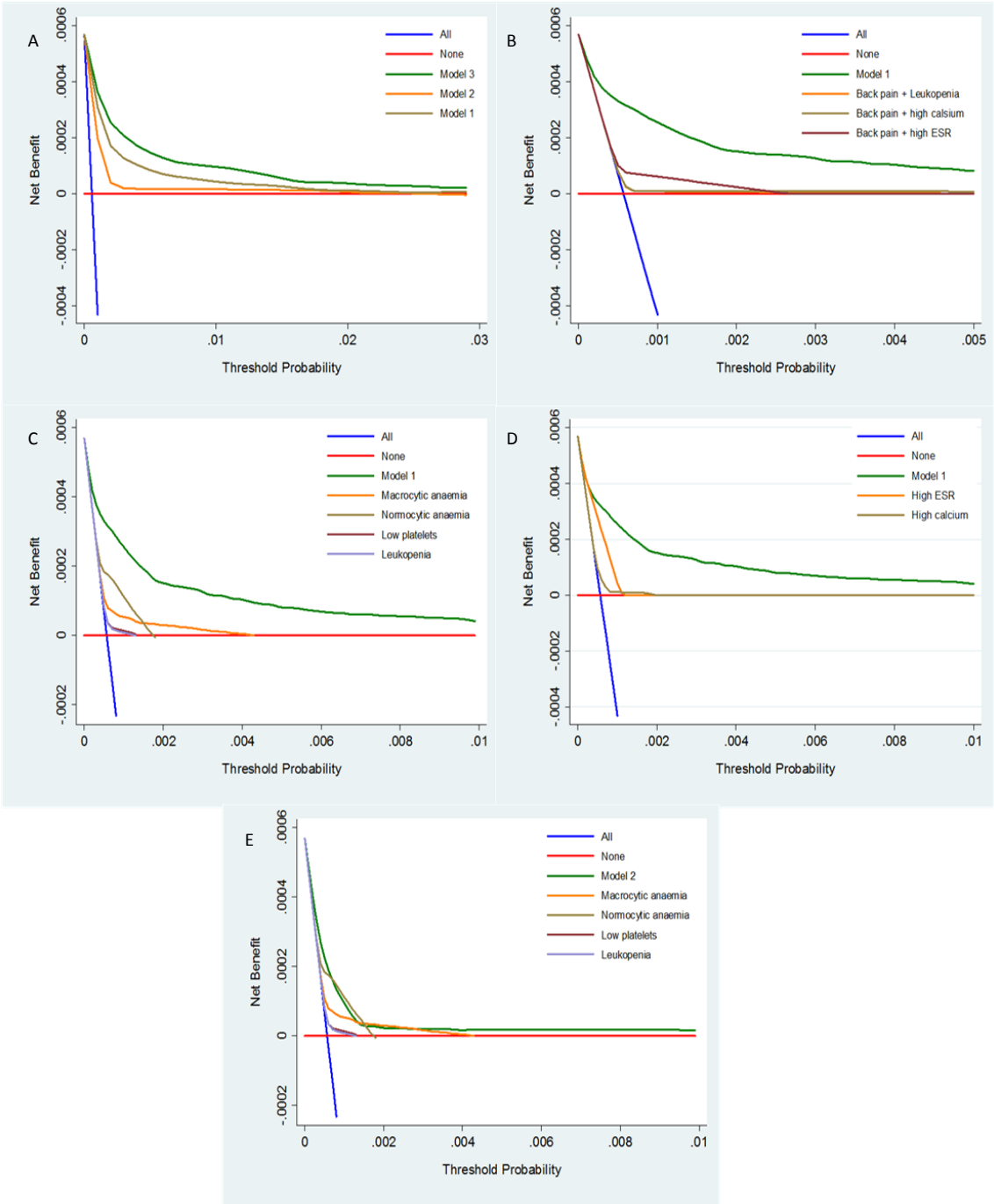
The x-axis corresponds to the decision/threshold probability, which is the probability at which we want to further test or refer a particular patient. For example, a GP might want to further test or refer patients suspected for cancer when they have a probability of more than 1%. By doing so, the GP should expect 99 false positives for every 1 true positive detected. This strategy may be appealing if the harm/cost of a false positive is trivial. The threshold probability also corresponds to the model probability so decision curve analysis can be affected by miscalibration. The y-axis corresponds to the net benefit which is estimated as shown in the methods section. I have selected to examine myeloma probabilities from 0 to 3% as current NICE guidelines recommend patients with cancer probability of more than 3% for cancer are referred to the 2-week wait pathway (20). As a result, examining higher probabilities will not be very useful. It shows the net benefit for referring all patients irrespective of myeloma probability, and for referring no-one, both of which are used as a point of reference.

Model 3 has the highest net benefit at all thresholds, followed by model 1 and then model 2. This means that model 3 will result in more true positives at the same number of false positives than model 1 and model 2 at the examined intervals. Similarly model 1 performs better than model 2 at all examined thresholds up until a threshold probability of 2%. Model 1 performs better than any individual test or simple combination (like back pain + high calcium) at all examined thresholds. In addition, single tests or combinations can be effective only at lower threshold probabilities while models can be effective even for high probability thresholds (>1%). Model 2 does not perform well according to net benefit approaches. It can perform better than other individual tests only for threshold probabilities greater than 0.35% (approximately). Macrocytic anaemia seems to be performing similarly with model 2 for

threshold probabilities from 0.2% to 0.35% and normocytic anaemia performs better than the model for threshold probabilities from approximately 0.1% to 0.2%.

I do not present the comparison of the net benefit between model 3 and the other tests because model 3 is better than model 1 and since model 1 is better than any other individual approach then it is the same for model 3 too.

Figure 2: Comparison of the different strategies for diagnosis of myeloma including the developed prediction rules, single symptoms, blood tests and NICE combinations using decision curve analysis. Y-axis corresponds to net benefit and x-axis corresponds to threshold probabilities that could potentially be used for the prediction models. All plots have lines corresponding to the net benefit of referring/treating all patients and referring/treating no one. The line that is the highest in each plot corresponds to the best diagnostic strategy at each particular threshold. The plots presented are the following: A) comparison of the 3 developed models, B) comparison of model 1 with combinations reported in NICE guidelines, C) comparison of model 1 with the other components of the full blood count, D) comparison of model 1 with ESR and calcium and E) comparison of model 2 with components of the full blood count. For certain plots the range was lower than the predefined 0-3% to improve visibility.



5.5 Discussion

5.5.1 Summary

In this chapter the clinical prediction rules developed in chapter 4 were validated in terms of discrimination, calibration and clinical utility. Model 3 was the best in terms of discrimination and net benefit but it was under-predicting in the last decile. Model 1 had very good discrimination and excellent calibration. Model 2 had good calibration but not very good discrimination and it was outperformed by single tests in terms of net benefit. A model based approach will result in better outcomes compared to decisions made based only on abnormalities on single blood tests or symptoms.

5.5.2 Comparison with existing literature and findings from previous chapters

Many results in this chapter confirm the findings of chapter 3 (case control study). For example in both occasions it has been shown that ESR has good rule out properties and calcium has good rule in properties while anaemia sits somewhere in the middle. Similarly the combination of high ESR or low haemoglobin has very good rule out properties as shown in chapter 3. Symptoms had similar diagnostic accuracy measures with other studies that examined their predictive ability (28).

To the best of my knowledge, there are no other clinical prediction rules for myeloma. A rule that was generated for ovarian cancer using the Qresearch database had very similar diagnostic performance measures as model 1 in terms of AUC, D statistic and R^2 with the 90^{th} percentile of the risk score identifying approximately 60% of cancer cases similarly to how model 1 performs (29). Very similar results were also obtained for a prediction rule generated for pancreatic cancer (30). In addition, model 3 has similar performance as clinical prediction rules that were generated for colorectal, lung and gastroesophageal cancers with all of them picking more than 70% of the cancer cases at the 90^{th} percentile of risk score (19,31,32).

There appears to be little to gain from measuring changes in full-blood count parameters only, as model 2 did not add any prognostic information over above model 1. It is also likely that having to do a second test will result in a delay in diagnosis compared to any rule based on one test.

5.5.3 Selection of final model based on clinical, statistical and implementation criteria

Based on the above results and other clinical and implementation criteria I have decided to select model 1 as my final model (model containing only the full blood count components, symptoms, demographic characteristics and risk factors). The reasons illustrated below explain my rationale for selecting model 1.

Model 3 had the best discrimination and clinical utility, followed by model 1 and model 2. Model 1 follows second and the last one is model 2. So the decision is between models 1 and 3. The problem with model 3 is calibration. It seems that model 3 under predicts in the last decile and this miscalibration can also affect the net benefit analysis (33). Miscalibration of model 3 could be attributed substantial of missing data for model 3 which would require more than 50 imputations in order for the models to stabilise, which was not feasible due to the size of the databases, the complexity of the imputation models, the computational power and the time required (34).

In addition, model 3 uses all tests that are currently being done when myeloma is suspected (full blood, calcium, inflammatory markers etc.). Model 3 might not be as useful for picking up patients suspected of myeloma due to its limited applicability but it can still be a useful approach for monitoring patients with MGUS since these patients are required to be tested with all tests at least once per year.

Based on the above it makes more sense to choose model 1 because it makes use of only the full blood count which is commonly requested even when myeloma is not necessarily suspected.

More on how this clinical prediction rule can be implemented in practice will be explored in more detail in the final conclusion chapter.

5.5.4 Threshold determination

The prediction rule with the single full blood works better than any other strategy at all examined thresholds. This means that a model based approach is better in practice than any other approach independently of selected threshold. I have decided to use 0.12% as the decision probability as it maximises sensitivity while at the same time keeps specificity over 90%. It also has a likelihood ratio over 5 which I have defined in the methods section as a useful rule in approach. Another option was to use a lower threshold such as the prevalence which maximizes sensitivity but this also produces a lot of false positives.

5.5.5 Strength and limitations

Strengths and limitations described in chapter 4 also apply to chapter 5. Below I describe additional strengths and limitations more specific to the validation chapter.
The reproducibility and geographical robustness of the clinical prediction rules was assessed using a large semi-independent dataset. The clinical prediction rules were examined in three different ways, i.e. discrimination, calibration and net benefit with the use of rigorous methodology. It has been shown that modelling approaches with use of methods like multiple imputation and fractional polynomials are reproducible and produce models that calibrated well in the other half of the data thus it is statistically valid.

The main limitation of this validation approach is that it's not a replacement of external validation. One of the most important criteria that an external validation approach satisfies and internal validation approaches do not is the transportability. Transportability refers to the accuracy of the prediction rules when they are applied to patients drawn from a different but related population and where the data are collected using different methods from the ones used in the development population.

Another limitation of the study design which could potentially affect the applicability of the clinical prediction rules are the constraints that I imposed on the inclusion criteria and more specifically selecting patients that had 2 full blood counts within a period of 1 year. Patients without a FBC on record or just a single FBC were excluded from the analysis. In a study that was conducted in Oxfordshire the investigators examined 151 473 full blood counts from 53 890 participants and identified that 56% of the population is investigated with at least 2 full blood counts (35). The same study also found that 30% of patients with a single recorded test had normocytic anaemia, a risk factor for myeloma. It is possible then that by restricting the clinical prediction rules to patients with two tests, a large number of patients with anaemia were missed. This has the potential to bias the prediction model if these patients differed systematically in their risk of myeloma, either by having a higher or lower risk than patients with two tests. One explanation for why some people get a single test whereas others get multiple tests is that in certain cases symptoms and abnormalities are very obvious (maybe due to more advanced disease) that the diagnosis happens very quickly after a single test. These patients are expected to have shorter diagnostic intervals. On the other hand patients that have more than one full blood count might be patients with mild disease so repeat examinations are conducted for monitoring purposes. The results of my study support this assumption as approximately 66% of the study population also had low haemoglobin at their repeat full blood count (90% for myeloma patients). These patients who get repeat tests are more likely to have longer diagnostic intervals due to having repeat tests but it is also possible that they have a slower progressing disease. These explanations are of course speculative as no data are available in order to evaluate the rationale behind why certain patients are examined with a single full blood count test and others with multiple full blood counts. Nevertheless the final prediction rule that was selected requires just one full blood count rather than two. There is potential for this clinical prediction rule to be applied on patients who are tested with a single full blood count if the population that has one full blood count is similar to the population in this study. This of course is quite a strong assumption so further validation of the selected

clinical prediction rules on different populations (i.e. patients who have a single full blood count) is required in order to confirm the clinical applicability of the developed prediction rules.

5.5.6 Simulation of different scenarios based on the selected clinical prediction rule and presentation of risk prediction tool for myeloma

I hereby present some simulated clinical examples on how model 1 would work. In order to make these examples as comparable as possible I assumed that all patients are male, at the age of 73 with a BMI of 26 and the only presenting symptom is back pain.

In table 6, patient 1 has a normal full blood count and the probability of myeloma within the next two years for this patient is 0.017% which is lower than the cut-off probability and the prediction rule says that no further action is needed. In contrast, patient 2 has a haemoglobin value of 10 (moderate to severe anaemia) and the probability of myeloma for this patient is almost 1%, and the rule suggests that serum and urine protein electrophoresis should be assessed. Equally for patient 3, who has signs of macrocytic anaemia (low haemoglobin and high MCV) and a myeloma probability of 1.6%. Patient 4 has microcytic anaemia (low haemoglobin but low MCV) and a probability of myeloma for this patient is 0.04% which is again lower than the prevalence and threshold used suggesting that anaemia is unlikely to be caused by myeloma and other causes should be looked at like for example colorectal cancer (36). Patient 5 has abnormalities in all almost all full blood count components resulting in a probability of 2.8% suggesting very urgent investigation is needed. Patient 6 was used as a proof of principle. This is a patient who has anaemia but had already been diagnosed with MGUS and his probability of myeloma is very high (21%). When a patient with MGUS starts having end organ damage (anaemia, kidney failure etc.), this is defined as progression to symptomatic myeloma thus treatment is required (37).

Table 6: Clinical examples

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
Gender	Male	Male	Male	Male	Male	Male
BMI	26	26	26	26	26	26
Age	73	73	73	73	73	73
Back pain	Yes	Yes	Yes	Yes	Yes	Yes
Rib pain	No	No	No	No	No	No
Chest pain	No	No	No	No	No	No
Nosebleeds	No	No	No	No	No	No
MGUS	No	No	No	No	No	Yes
Haemoglobin	15 (N)	10 (A)	11 (A)	11 (A)	11 (A)	10 (A)
White cell count	6 (N)	6 (N)	6 (N)	6 (N)	4 (A)	6 (N)
Platelets	300 (N)	300 (N)	300 (N)	300 (N)	300 (N)	300 (N)
MCV	85 (N)	85 (N)	100 (A)	70 (A)	100 (A)	85 (N)
Pr(myeloma)*	0.017%	0.89%	1.6%	0.04%	2.8%	21%

A-abnormal value
N-normal value
* Probability of myeloma within the next 2 years

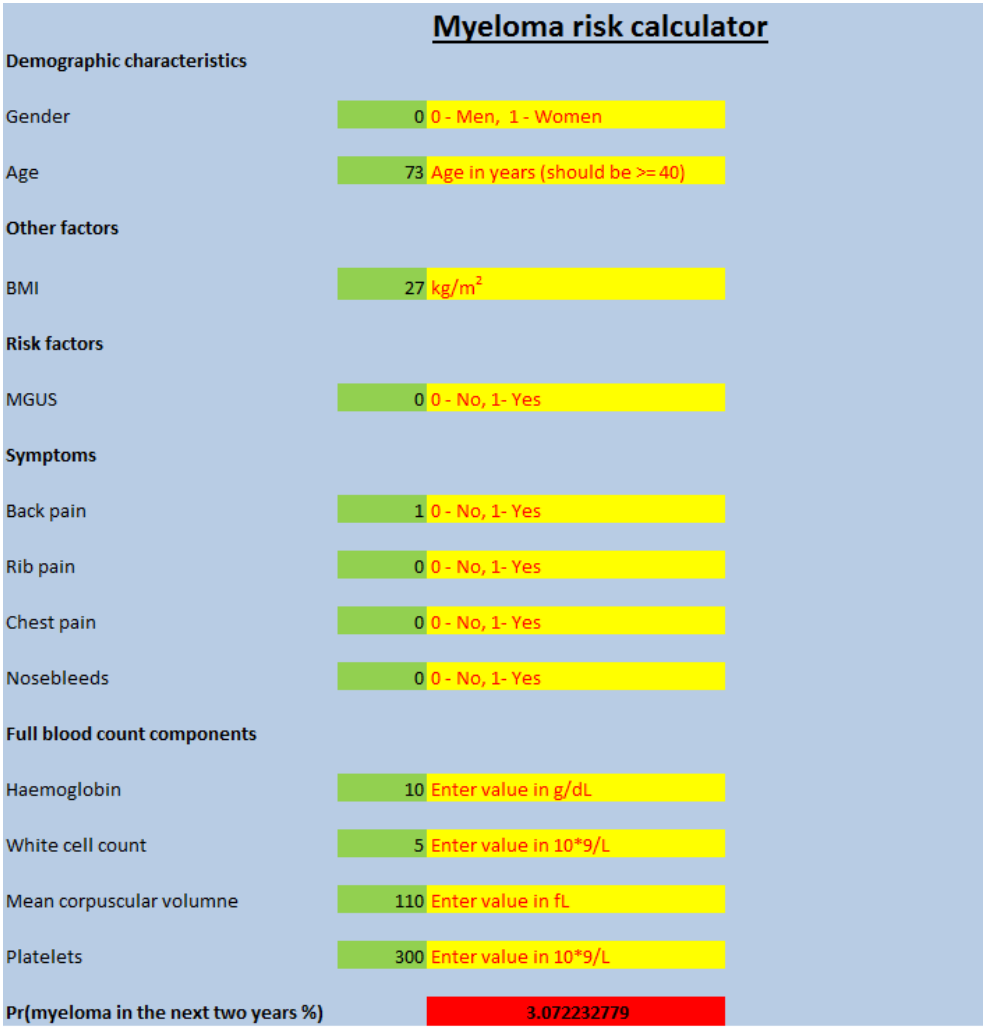
All patient examples were selected to be male, of age 73 and BMI 26 with a back pain symptom in order to make comparisons easier.

Patient 1: Normal full blood count
Patient 2: Normocytic anaemia
Patient 3: Macrocytic anaemia
Patient 4: Microcytic anaemia
Patient 5: Macrocytic anaemia and leukopenia
Patient 6: MGUS and normocytic anaemia

Boxes in red correspond to patients that have a risk over the decision threshold probability of 0.12%.

I have also created a calculator in excel which can used very easily to estimate the probability of a patient having myeloma or developing myeloma within the next two years (figure 3). This is just to demonstrate that the generated prediction rule can easily be used and it is also a tool that can also be used in future studies where the evaluation of the rules can be examined.

Figure 3: Risk score calculator



Myeloma risk calculator		
Demographic characteristics		
Gender	0	0 - Men, 1 - Women
Age	73	Age in years (should be >= 40)
Other factors		
BMI	27	kg/m^2
Risk factors		
MGUS	0	0 - No, 1- Yes
Symptoms		
Back pain	1	0 - No, 1- Yes
Rib pain	0	0 - No, 1- Yes
Chest pain	0	0 - No, 1- Yes
Nosebleeds	0	0 - No, 1- Yes
Full blood count components		
Haemoglobin	10	Enter value in g/dL
White cell count	5	Enter value in 10*9/L
Mean corpuscular volumne	110	Enter value in fL
Platelets	300	Enter value in 10*9/L
Pr(myeloma in the next two years %)		**3.072232779**

5.5.7 Implications for future research

Although the models were validated in a geographical way it is still important to do external validation. This could be on another UK database like the THIN database (38) . Other options could include other databases outside the UK. In addition, it should also be validated in a population where the selection criteria are based on a single full blood count rather than two in order to confirm if they are performing the same.

Furthermore, more studies should be conducted in order to properly examine and determine the appropriate threshold when classifying the low/moderate and high risk categories (by taking into account cancer outcomes, stage, quality of life, costs etc.). The decision curve analysis was enough to show that using a model based approach is better than any other symptom based or single test approaches but it should not be used to determine the threshold. This is more extensively discussed in the strengths and limitations section above. I presented some reasonable thresholds but these thresholds should be examined in more detail in an impact study (39).

5.5.8 Implications for future practice

Implementation strategies and implications for practice will be discussed extensively at the final conclusion chapter

References

1. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453–73.

2. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.

3. Subramanian J, Simon R. Overfitting in prediction models – Is it a problem only in high dimensions? Contemp Clin Trials. 2013;36(2):636–41.

4. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer. 1994;69(6):979–85.

5. Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JDF. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059–79.

6. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683–90.

7. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691–8.

8. Justice AC, Covinsky KE, Berlin JA. Assessing the Generalizability of Prognostic Information. Ann Intern Med. 1999;130(6):515.

9. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925–31.

10. Harrell FE Lee KL MDB. Tutorial in Biostatistics - Multivariable Prognostic Models:

Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. Stat Med. 1996;15(4):361–87.

11. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. 2004;23(5):723–48.

12. Royston P. Explained variation for survival models. Stata J. 2006;6(1):83–96.

13. Choodari-Oskooei B, Royston P, Parmar MKB. A simulation study of predictive ability measures in a survival model I: Explained variation measures. Stat Med. 2012;31(23):2627–43.

14. Jinks RC, Royston P, Parmar MK. Discrimination-based sample size calculations for multivariable prognostic models for time-to-event data. BMC Med Res Methodol. 2015;15(1):82.

15. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2018;

16. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models. Epidemiology. 2010;21(1):128–38.

17. Multiple myeloma - NICE CKS [Internet]. [cited 2017 Sep 1]. Available from: https://cks.nice.org.uk/multiple-myeloma#!scenario

18. Collins GS, Altman DG. Identifying patients with undetected pancreatic cancer in primary care: An independent and external validation of QCancer?? (Pancreas). Br J Gen Pract. 2013;63(614):636–42.

19. Hippisley-Cox J, Coupland C. Identifying patients with suspected colorectal cancer in primary care: Derivation and validation of an algorithm. Br J Gen Pract. 2012;62(594):29–37.

20. NICE Guideline. Suspected Cancer: Recognition and Referral. NICE Guidel. 2015;(June):1–378.

21. Whitehead SJ, Ali S. Health outcomes in economic evaluation: The QALY and utilities. Vol. 96, British Medical Bulletin. 2010; 96:5–21.

22. Vickers AJ, Cronin AM, Elkin EB, Gonen M. BMC Medical Informatics and Decision Extensions to decision curve analysis , a novel method for evaluating diagnostic tests , prediction models and molecular markers. 2008; p 8–53.

23. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6.

24. Fitzgerald M, Saville BR, Lewis RJ. Decision Curve Analysis. 2015;313(4):409–10.

25. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. J Clin Epidemiol. 2007;60(9):979.

26. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation : current practice and guidelines. BMC Med Res Methodol 2009;28:9–57.

27. Moons KGM, Altman DG, Reitsma JB, Ioannidis JP a, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162(1):W1–73.

28. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. Br J Gen Pract. 2015;65(631):e106-13.

29. Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. BMJ. 2011;344:d8009–d8009.

30. Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: Derivation and validation of an algorithm. Br J Gen Pract. 2012;62(594):38–45.

31. Hippisley-Cox J, Coupland C. Identifying patients with suspected gastro-oesophageal cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract. 2011;61(592):e707-14.

32. Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care : derivation and validation of an algorithm. Br J Gen Pract. 2011;61(592):e715–23.

33. Van Calster B, Vickers AJ. Calibration of Risk Prediction Models. Med Decis Mak. 2015;35(2):162–9.

34. Graham JW, Olchowski AE, Gilreath TD. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. Prev Sci. 2007;8(3):206–13.

35. McCartney D, Shine B, Hay D, Lasserson DS. The evaluation of anaemia in an older primary care population: retrospective population-based study. BJGP Open. 2018;1(4):bjgpopen17X101157.

36. Hamilton W, Lancashire R, Sharp D, Peters TJ, Cheng KK, Marshall T. The importance of anaemia in diagnosing colorectal cancer: a case-control study using electronic primary care records. Br J Cancer. 2008;98(2):323–7.

37. Röllig C, Knop S, Bornhäuser M. Multiple myeloma. Lancet. 2015;385(9983):2197–208.

38. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: Independent and external validation of QCancer?? (Ovarian) prediction model. Eur J Cancer Care (Engl). 2013;22(4):423–9.

39. Wallace E, Uijen MJM, Clyne B, Zarabzadeh A, Keogh C, Galvin R, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. BMJ Open. 2016;6(3):e009957.

6. Discussion

The work in my thesis was motivated by the finding that myeloma patients have more consultations in primary care than any other cancer (1). This implied that myeloma patients experience on average a longer diagnostic interval and that this could contribute to the relatively poor prognosis (2).

The aim of the thesis was the development of prediction rules that could potentially reduce diagnostic intervals for myeloma. I started this work with a systematic review, quantifying the diagnostic pathway and identifying where the longest delays are occurring. In a case-control study, I then looked at the progression of myeloma biomarkers (symptoms and blood tests) over time, identifying which are manifesting the earliest. Additionally, optimal test combinations were explored to help general practitioners rule in or rule out myeloma

This culminated in the generation of a clinical prediction rule in a cohort study, which can be used to flag patients at higher risk of myeloma even when myeloma is not suspected in the first place.

6.1 Summary of findings

In chapter 1 I reviewed the literature and identified key aspects of myeloma diagnosis both in primary and secondary care settings, and the prognosis of myeloma including survival, number of complications and their association with diagnostic delays. In addition, I identified what tools are currently available to doctors that can help them suspect whether a patient has myeloma or not.

The literature review highlighted both an association between diagnostic delays and adverse outcomes, and a lack of research when it comes to diagnosing myeloma in primary care, especially when compared to other cancers such as colorectal, lung, prostate and breast for which multiple tools are available (3–6).

In chapter 3 I conducted a systematic review in order to map the diagnostic pathway for multiple myeloma patients and quantify the length of each step on the pathway from symptom onset to confirmation of diagnosis. This helped me to identify how long it takes to diagnose myeloma and where the longest delays are observed. The key findings of this systematic review included the following:

1. Most of the delays in diagnosis are observed in the healthcare system. The median length of the patient interval is approximately one month, whereas the diagnostic intervals is longer than three months for 50% of patients and longer than eight months for 25%.
2. In primary care, the median interval is approximately three weeks which is longer than for other cancers.
3. Although no studies were identified that quantified the interval between referral to secondary care and diagnosis, inference based on the length of the other intervals showed that the secondary care interval could be longer than the primary care interval.

In chapter 3, I re-analysed a case-control study based on CPRD data to identify which abnormalities occur earlier in the patient interval. In addition, I was able to examine the diagnostic pathway in more detail and quantify the secondary care interval. The main findings from this work were:

1. Symptoms such as back, rib or chest pain, chest infections and nosebleeds are the earliest symptoms. Symptoms that manifest late in the disease progression include weight loss, nausea and fractures.
2. Average haemoglobin reaches anaemic levels at approximately two years before diagnosis but the decline starts as early as three years before diagnosis. Similar patterns were observed with the average erythrocyte sedimentation rate and plasma viscosity: the average value of these inflammatory markers crosses the upper threshold of the reference range at approximately two years before diagnosis. Creatinine and calcium increase, indicating myeloma has impacted renal function,

are present quite late in the disease progress, i.e. approximately three to six months before diagnosis.

3. Erythrocyte sedimentation rate and plasma viscosity are the most useful inflammatory markers both for ruling in and ruling out myeloma while C-reactive protein does not have any useful rule in/out properties.
4. A combination of a normal full blood count, a normal ESR or PV and in certain cases calcium is able to identify more than 90% of myeloma patients that are currently tested in primary care with these test combinations. Patients in whom all tests are normal have an extremely low probability of myeloma.
5. Only 13% of myeloma patients are sent to a haematologist as a first referral. The majority of myeloma patients are referred to other specialities and as a result their diagnosis is delayed, which confirms the results of the systematic review where I estimated the median diagnostic interval to be longer than three months.
6. For 62% of patients their first referral is routine and only 6% are referred through the two-week-wait cancer pathway. Myeloma patients referred through the two-week wait pathway had a shorter diagnostic interval. Patients who had a first referral to a haematologist had a shorter referral to diagnosis compared to patients who were referred via the 2ww route (median: 25 vs 40).

In chapter 4 I report the development of a myeloma clinical prediction rule using CPRD data in a retrospective cohort design. Some of the study design decisions were informed by the results of chapters 2 and 3, i.e. the length of follow up and which blood tests to focus on. The aim was to develop clinical predictions rules that are able to flag up patients with an elevated risk of myeloma that should be investigated further. The full blood count (FBC) which includes haemoglobin became the focus of this investigation as it is one of the earliest markers of myeloma and a test that is used routinely in primary care, allowing us to also include change in full blood count results. The main findings are the following:

1. Change in haemoglobin from the previous measurement was significant with one unit decrease corresponding to a risk increase of 14%. Abnormalities in single tests are more predictive than the change between two tests when the two tests are identified within a year. The other components of the full blood count (platelets, MCV, white cell counts) were not predictive over time.
2. ESR and PV were again identified as having better rule in and rule out properties than CRP reinforcing the results of chapter 3 even further.
3. Three clinical prediction rules were developed. All of them included patient demographics (age, gender, BMI), risk factors (MGUS) and symptoms (nosebleeds, back pain, chest pain, rib pain) and variations of different blood tests. The first model included a single full blood count, the second changes in full blood count parameters from the previous test and the third a single full blood count, calcium and ESR. Inclusion of both absolute values and change from previous test in a single model was resulting in either collinearity or to non-significance of change variables.

In chapter 5 I validated the models developed in chapter 4 in a geographically independent dataset using the same CPRD source data. The models were assessed for their discrimination, calibration and clinical utility. The main findings of this chapter include:

1. Model 1, which contains only the full blood count components, has very good discrimination and calibration.
2. Model 3, which includes all tests (full blood count, calcium, ESR) has better discrimination than model 1 but it does not calibrate very well especially at the higher risk percentile. Additionally it contains variables like calcium which is usually a later manifestation of myeloma and not that useful for early diagnosis.
3. Model 2, which uses the change from previous full blood count, did not have good discrimination (AUC of 70%).
4. Based on these, I selected model 1 for further assessment.
    a. The optimal decision threshold that I selected is 0.12% myeloma probability.

b. Decisions based on the clinical prediction rule result in better clinical utility than making decisions based on symptoms or single tests. Approaches using single tests only (i.e. test all patients with anaemia) result in fewer true positives and more false positives.

6.2 Strengths and limitations of the thesis

The individual strengths of each study are described in length in each chapter but briefly some of the main strengths of each individual chapter are the following.

The main strength of chapter 2 (systematic review) included the use of all available information when quantifying the length of each interval including the median and IQR, and not just focusing on the mean or the median. It is also the first systematic review which quantified in full the patient pathway from symptom onset to confirmation of diagnosis including the patient, primary and referral to diagnosis intervals.

For the chapters 3, 4 and 5 I used CPRD data, which allow the investigation of rare diseases like myeloma because of the size of the dataset, and the results of which are broadly generalizable to the primary care population (7). One of the main strengths and one of the most novel analyses done in chapter 3 was the mapping of the disease trajectories over time for both symptoms and blood tests parameters. This is another advantage of using CPRD (or other similar databases) as they allow for long periods of follow up. In addition this was the first study that examined how the different blood tests work in combination as most studies report them individually, usually not within a primary care setting and without quantifying their predictive ability by using statistical measures like sensitivity, specificity and likelihood ratios (8). In addition, the focus on the primary care setting is important because I have shown in chapter 2 (systematic review) and chapter 3 (case-control CPRD study), decisions taken in primary care, especially referral to which specialty and with which urgency, can affect the length of the diagnostic interval, thus improving decision making at the primary care level can have a substantial impact.

The models developed in chapters 4 and 5 should be applicable to most primary care populations since they include tests that are conducted routinely. In fact, the majority of myeloma patients (70%) included in our case-control study (chapter 3) had such tests at least once in the year before diagnosis. Additionally, I employed a cohort design which results in less biased estimates than a case-control design, and I constructed geographically independent derivation and validation datasets. Finally, advanced statistical methods were used, such as multiple imputation to deal with missing data and fractional polynomials to take into account the non-linear relationship between the continuous predictor variables and the response. Using the results of each chapter to inform the next is one of the main strengths of my thesis. For example, the systematic review helped identify where the longest delays are occurring and where new approaches should be implemented in order to reduce delays. Similarly, the case-control study informed the design decisions of the algorithm development.

As for the strengths, the limitations of each study are also presented in detail in each individual chapter. For the systematic review chapter multiple limitations were identified which included the lack of formal ways to perform meta-analysis on quartiles, the lack of standardised definitions for the different intervals and time points, the small number of included studies and the lack of studies quantifying the referral to diagnosis interval. Also it became clear that there was much heterogeneity in study design.

Using CPRD databases to answer diagnostic type questions is not without its limitations either. A main limitation includes the verification bias (9). Verification bias is a type of bias where the results of the index test will affect whether a patient will get the reference standard or not. This means that while myeloma cases would have been confirmed with the reference standard which usually involves bone marrow biopsy that is not the case for all the controls or the non-diseased patients. This type of bias can overestimate sensitivity and specificity measures (10). Verification bias is not necessarily limited to retrospective studies but can also be a problem in prospectively conducted cohort studies; if the reference standard is expensive or invasive then it is unlikely that all patients will get it. This bias to some extent is countered in the retrospective

cohort study due to having a long follow up which allows classifying whether a patient has myeloma or not with greater certainty. The case-control study in chapter 3 is more likely to suffer from this type of bias.

Another issue with using CPRD is the coding of the different symptoms. Symptoms are less likely to be accurate compared to diagnostic codes which have been shown to be fairly accurate (11). This could be due to patients not reporting all their symptoms or if they do they might not be recorded by the GP which can result in under-recording. This can affect analysis of association measures like the hazard/odds ratios and the diagnostic accuracy measures.

In general chapters 3, 4 and 5 highlighted one of the main limitations encountered in the CPRD database which is missing data especially in the case of blood tests. Two mechanisms of missing data were encountered. The first one is patients for which there was a record of them having a particular test but the value was not reported and the second one was when the blood test was not ordered in the first place. The first scenario can be dealt with using different imputation techniques as the data in the present study suggest an MAR assumption (the rationale of the assumption is explained in the corresponding chapter). For the second scenario where blood tests are not ordered in the first place the analysis is more complex as the data are likely to be Missing Not at Random (MNAR). For chapter 3 the analysis was restricted to patients who had the test hence the results are generalizable to those patients who are currently being tested. For chapters 4 and 5 multiple imputation procedure was used, taking into account symptoms, demographic characteristics and other blood tests. The results of the multiple imputation analysis and the full case analysis produced similar results in most cases suggesting the results are robust.

A further limitation of chapters 4 and 5 is the fact that I selected patients who had two full blood counts within a year. This poses a limitation as it imposes restrains on the population on which the clinical prediction rules can be applied. Patients that are not examined with a full blood count or are examined with a single full blood count are excluded. If patients with a

single full blood count systematically differ from patients with two full blood counts in their risk of myeloma then this can bias the clinical prediction rules. Extensive discussion on the limitation of the inclusion criteria are presented in chapter 5 section 5.5.5. The main reason I decided to examine this population is because I wanted to investigate whether a change from a previous full blood count can be predictive of the condition, similar to prior evidence on ovarian cancer which showed that earlier diagnosis is possible using a longitudinal screening algorithm that incorporates CA125 changes over time (12). However, although the change in haemoglobin was predictive, it was not as predictive as the absolute haemoglobin value. This was mainly due to the fact that in most cases both full blood count tests were abnormal even though the median interval between the two was two months for myeloma patients. The results indicated that this is a relatively short time period to examine meaningful changes between two full blood count parameters which could explain the small change between the tests or why both tests were showing abnormalities. An alternative approach which could be implemented in future studies is to pre-specify a minimum time period between the two tests i.e. 6 months. A different study that examined patterns of declining haemoglobin over time identified that a reduction of 0.3 g/dL (grams per decilitre) every six months can serve as a warning sign of the disease (13). This in combination with the findings of chapter 3 suggest that the optimal design is to have multiple full blood counts over long periods of time, preferably going back years before diagnosis in order to properly capture significant changes in haemoglobin. The aim of model 2 in this case was to examine whether a change from the previous test could provide enough information over and above the absolute value for predicting myeloma which does not seem to add much in this case.

The use of geographical validation resulted in very similar datasets for derivation and validation. This can be positive as it indicated that the regions in the UK are quite similar which confirmed the geographical robustness of the prediction rules. On the other hand, given the similarity of the two datasets, it cannot be said with certainty that the prediction rules will perform equally well in a totally different dataset. Other options of validation included splitting

by practice but it has been shown in other studies that this method also produces very similar datasets (14,15). Ultimately, the performance of the prediction rule needs to be evaluated further with external validation.

Another limitation which affected the feasibility of the fitted models and especially model 3 was the computational requirements of the analysis which was a consequence of the very large sample size. Due to using methods like multiple imputation and complex modelling approaches such as fractional polynomials each step of the analysis required a substantial amount of time. It was not feasible to do more than 10 imputations which is in-line with other studies that make use of such large datasets (16). This greatly affected model 3 which required more than 50 imputations for the coefficients to stabilize and this might also explain the bad calibration of that model. In addition other methods such as bootstrapping or k-fold cross-validation (which are classified as internal validation methods) were not used due to how computationally intensive they are especially in datasets of this size (17). K-fold cross-validation for example is a procedure where the dataset is split in smaller equally sized datasets (usually 10 thus 10 fold cross validation). The model is fitted in the 9 datasets (combined) and validated on the one left out. This procedure is repeated 10 times until all of the datasets go through validation and then the statistics of interest (C-statistic, D-statistic etc.) are combined. When used in conjunction with multiple imputation the cross-validation procedure needs to be conducted 10 times (one for each imputed dataset) which is infeasible both computationally and time wise. Similar problems are encountered with bootstrapping which is computationally much more demanding than cross-validation.

The ideal design for the development of the clinical prediction rules would be a prospective study where the necessary measures would be taken to overcome problems that are encountered when using CPRD type databases. In the studies presented in the thesis not all patients were having the tests of interests so either the analysis was restricted to the patients who had the tests or their missing data were imputed with the use of imputation procedures. In a prospective study all tests would be collected on all included patients. An example of a study like that

would be the design used for the Framingham risk score which is a prognostic model used to predict cardiovascular events (18). In this study, which started in 1948, 5,209 individuals between the ages of 30 and 62 were recruited and a lot of information were collected which included medical history, medical tests and lifestyle interviews. Similar cardiovascular risk scores were generated by using databases like the CPRD or QRESEARCH. On the most known cardiovascular risk scores is the Qrisk which was developed using the QRESEARCH database (19,20). Similar problems with missing data were also encountered there especially with variables like blood pressure, cholesterol, smoking status and BMI. But here we are restrained by what I believe is the main challenge I encountered in the thesis, and that is the rarity of myeloma. To study a rare disease and draw robust conclusions, extremely large sample sizes would be required.

The ideal prospective study for myeloma would recruit patients from the general practice population. Assuming an incidence of 0.06% would require a sample of approximately 1600 persons to get 1 case of myeloma. Recruiting 100 myeloma patients, which is a conservative sample size estimate for a clinical prediction rule, would require a sample of more than 160,000. The haematology department of the Churchill hospital in Oxford diagnoses approximately 10 myeloma patients per year. That means the recruitment period would be approximately 10 years if only 1 site was used. All patients would be getting all common tests that were identified to be useful like the full blood count, calcium, creatinine and the three inflammatory markers. Very conservatively the total cost of just the tests alone on a sample of 160,000 would be approximately 1.5 million pounds based on the average unit cost of haematology and clinical biochemistry tests provided by the NHS reference costs (21). This is without taking into account the extra testing required from the serum/urine electrophoresis and bone marrow biopsies. In addition to overcome problems with verification bias the reference standard tests which include serum/urine protein electrophoresis and bone marrow biopsy should be applied to everyone included in the study which is not something that can be done due to ethical reasons.

So based on the above, the ideal study design can overcome some of the limitations that CPRD type databases display like missing data but verification bias is likely to remain a problem. In addition, the cost of the study both financially and time wise will be enormous. Thus the CPRD database even with all the limitations is probably the only viable approach for studies of diseases as rare as myeloma which also require the use of blood tests.

6.3 Implications

6.3.1 Clinical Implications

When a patient presents in general practice with myeloma related symptoms which might include back pain, rib pain, infections etc. a GP will usually order certain tests which can include but are not limited to the combinations that were identified in the case-control study (chapter 3) like the full blood count, an inflammatory marker (ESR or PV) and calcium. If there are abnormalities in any of these tests myeloma could be suspected and then serum or urine protein electrophoresis could be ordered. In chapter 3 I have shown that the above combination can be quite effective for ruling out myeloma. The most likely scenario is that the GP will not suspect that the patient has myeloma and thus not order all tests. My work in the previous chapters has shown that it is uncommon for a doctor to order all blood tests but quite common to order one of them as an initial investigation, most likely as a triage or general testing strategy for a range of conditions and not particularly for myeloma. The most common test is the full blood count which was the central test used in the creation of the clinical prediction rules. The selected prediction rule (see chapter 5) can be implemented within the computer GP systems or within the laboratory management system. The prediction rule can be applied on patients that have a full blood count and flag patients with a high risk of myeloma.

An alternative and simpler strategy would be to test everyone with recent anaemia rather than implement the prediction rule. This approach as demonstrated in chapter 5 would result in fewer true positives and more false positives. To be more precise, sensitivity would be 7% higher and

specificity would be 6.2% higher by using the clinical prediction rule with a decision threshold of 0.12% compared to just using anaemia. For example, in the validation dataset of 446, 986 people, this increase in specificity resulted in approximately 26,000 fewer false positives. The net benefit analysis has shown that the prediction rule dominates the test all or test no-one strategies over all possible decision thresholds and also decisions made based on anaemia only. In addition, as shown in chapter 3 and in chapter 5, patients with myeloma are found to be anaemic on several occasions prior to diagnosis suggesting a lot of missed opportunities for earlier diagnosis, which means an approach that is able to flag the possibility of myeloma could be quite useful.

The full blood count prediction rule of course will still not pick up all patients or not as many if the patients were having all tests. The clinical prediction rule containing the ESR, full blood count and calcium had an estimated sensitivity of 72% compared to 61% of the prediction rule that uses only the full blood count at the selected threshold probability. The purpose of the algorithm though is to identify patients for whom not all the tests are ordered because myeloma is not yet suspected. In the analysis database only 10% of myeloma patients had all three tests so its use might be limited. On the other hand, a more comprehensive prediction rule including ESR and calcium can be a useful approach when monitoring MGUS patients since they are annually monitored with all the tests (22).

On a more practical level there are some difficulties that might prevent the implementation of this clinical prediction. Firstly, this is a complex scoring system so it requires a calculator similar to the one presented in chapter 5. Similar to Qcancer, which is a set of clinical prediction rules that can be used to identify cancer patients, this could be achieved through web based calculators (23). However, GP workload has been increasing over the years which might make it hard to make full use of these clinical prediction rules (24). One way to make the use of the clinical prediction rule as efficient as possible is for it to be incorporated in the laboratory information systems similar to the AKI algorithm (Acute Kidney injury) (25). The AKI algorithm is integrated into the laboratory systems and identifies potential cases of acute kidney

injury based on laboratory data and sends these results to patient management systems. Similar strategies could be used for the diagnosis of myeloma or any other cancer that is heavily reliant on blood tests that are commonly requested for a range of conditions. Abnormalities could result in a set of recommendations back to the GP. For patients who have a very high probability of myeloma, the recommendation could be to order further testing (SPE/UPE) or to refer straight to a haematology department given that current NICE guidelines recommend referral when the probability of cancer is more than 3% (26). One of the most important barriers that need to be explored is the implementation of the prediction rule. The AKI algorithm does not require the use of symptoms and BMI in contrast with the proposed algorithm which requires both. Enriching the laboratory databases with more information like symptoms accompanying the ordering of a specific test could be one solution. This of course is not an easy task but it can prove to be extremely useful not only for myeloma but for a range of conditions that require the use of blood tests.

An alternative approach not involving the laboratory would be to implement the prediction rule within the GP software systems like EMIS or VISION. For this approach software could be developed which can identify eligible patients based on the information entered into the system (over 40, with a full blood count and no previous diagnosis of myeloma).

Ideally these types of algorithmic approaches should be automated within the software systems in order to make the procedures as efficient as possible similar to how AKI works. It has been shown that in the cases of colorectal and prostate cancer an electronic trigger-based intervention is able to reduce delays by a significant amount (27). This in combination with the findings of chapter 5 regarding reduction of the length of the diagnostic interval present some evidence of the potential benefit that an algorithmic approach could have.

6.3.2 Future research

As mentioned in the previous section the best way to implement the prediction rules is through an automated system which could flag patients that have a high risk of myeloma. Before this

can be achieved though, more research should be done in order to establish the usefulness and the impact of the developed clinical prediction rules. Currently very few clinical prediction rules go through impact analysis which is a very crucial step during their evaluation and implementation process (28). Based on an existing framework the following steps should be followed when examining the impact of a developed clinical prediction rule (29):

Phase 1 – Exploratory phase: In this phase the derivation and validation studies take place. In order for this step to be completed external validation should be conducted. External validation studies should be done in order to confirm whether the developed clinical prediction rules perform similarly well in other databases (30). It is an integral component when developing a clinical prediction rule and currently many clinical prediction rules are developed but not many are externally validated (31,32). Before deciding whether the prediction rules developed in this study should be implemented external validation should be done. Successful external validation is required if the model is to be considered for implementation in practice.

Phase 2 – Preparation phase: If the clinical prediction rules have been successfully validated, then implementation strategies need to be considered. For example, if the clinical prediction rule is going to be implemented in the laboratory (similar to how the AKI algorithm mentioned above works) then computer software should be developed that can apply the prediction rule on the data available in the laboratory computer systems. Results from the prediction rules should then be automatically sent to the GP. This intervention should be piloted in a feasibility study in order to assess the practicality of the proposed intervention. This feasibility study could help identify the different barriers for the implementation of the clinical prediction rule at the proposed setting (some of them were mentioned in the previous section).

Phase 3 – Impact analysis: The aim of an impact analysis is to determine within a randomised clinical trial whether the use of the clinical prediction rule is cost-effective and whether it improves clinical outcomes. Initially, short term outcomes should be explored such as the effect

of the prediction rule on the clinical pathway while more long term outcomes should include impact on myeloma stage at diagnosis, disease free survival, overall survival etc.

Phase 4 - Long term implementation: If the analysis shows that the clinical prediction rule is an effective tool for reducing delays in myeloma and improving patient outcomes, then the focus should shift on their more widespread implementation within the health system.

In addition to the further evaluation of the clinical prediction rule, another avenue of research is a more detailed analysis of symptoms and blood tests over longer periods of time than what I used in chapter 4. A period of one year does not seem to be enough to capture the effect of changing blood parameters over time. Even though this can be quite useful currently there are many missed opportunities for an earlier diagnosis of myeloma even when blood tests show abnormalities as shown in chapters 3 and 5 which can be up to 2 years before diagnosis.

The thesis also revealed the need for more methodological research on missing laboratory data in CPRD. It is not clear which approach is currently the best and we cannot be certain about the missing data mechanisms in the data. For most tests it is highly likely that the recording of the tests and the test results are likely to be missing not at random. More research is required especially on common tests that are ordered for a variety of conditions like the full blood count and the different inflammatory markers.

6.3.3 Concluding remarks

In conclusion, I developed a clinical prediction rule that has the potential to reduce the time to diagnosis of myeloma patients. Future research should focus on establishing the performance of the algorithm in other databases and investigate potential implementation strategies. Currently the simple combinations of full blood count and inflammatory marker can help GPs pick up myeloma patients up until more complex interventions like the developed clinical prediction rules are implemented in practice.

References

1. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: Findings from the 2010 National Cancer Patient Experience Survey in England. Lancet Oncol. 2012;13(4):353–65.

2. Lyratzopoulos G, Abel G a, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. Br J Cancer. 2013;108(3):686-90.

3. McCowan C, Donnan PT, Dewar J, Thompson A, Fahey T. Identifying suspected breast cancer: Development and validation of a clinical prediction rule. Br J Gen Pract. 2011;61(586):205–14.

4. Shariat SF, Karakiewicz PI, Margulis V, Kattan MW. Inventory of prostate cancer predictive tools. Curr Opin Urol. 2008;18(3):279–96.

5. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk Prediction Models for Colorectal Cancer: A Systematic Review. Cancer Prev Res (Phila). 2016;9(1):13–26.

6. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132(2):365–77.

7. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T Van, et al. Data Resource Profile : Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.

8. Kyle RA, Gertz MA, Witzig TE, Lust JA, Lacy MQ, Dispenzieri A, et al. Review of 1027 patients with newly diagnosed multiple myeloma. Mayo Clin Proc. 2003;78(1):21-33

9. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, Van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. Cmaj. 2006;174(4):469–76.

10. Rosman AS, Korsten MA. Effect of Verification Bias on the Sensitivity of Fecal Occult Blood Testing: a Meta-Analysis. J Gen Intern Med. 2010;25(11):1211–21.

11. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract. 2010;60(572):e128-36.

12. Drescher CW, Shah C, Thorpe J, O'Briant K, Anderson GL, Berg CD, et al. Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule. J Clin Oncol. 2013;31(3):387–92.

13. Goldshtein I, Neeman U, Chodick G, Shalev V. Variations in hemoglobin before colorectal cancer diagnosis. Eur J Cancer Prev. 2010;19(5):342–4.

14. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017;357:j2099;

15. Hippisley-Cox J, Coupland C. Development and validation of QMortality risk prediction algorithm to estimate short term risk of death and assess frailty: cohort study. BMJ 2017;358:j4208.

16. Hippisley-cox J, Coupland C. Identifying patients with suspected gastro-oesophageal cancer in primary care : derivation and validation of an algorithm. Br J Gen Pract. 2011; 61(592): e707–e714.

17. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925–31.

18. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB.

Prediction of Coronary Heart Disease Using Risk Factor Categories. Circulation. 1998;97(18):1837–47.

19. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335:136

20. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2. BMJ 2008;336:1475

21. NHS reference costs 2015 to 2016 - GOV.UK [Internet]. [cited 2018 Aug 23]. Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016

22. van de Donk NWCJ, Palumbo A, Johnsen HE, Engelhardt M, Gay F, Gregersen H, et al. The clinical relevance and management of monoclonal gammopathy of undetermined significance and related disorders: Recommendations from the European Myeloma Network. Haematologica. 2014;99(6):984–96.

23. QCancer 2017 [Internet]. [cited 2018 Jul 27]. Available from: https://qcancer.org/

24. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. Lancet (London, England). 2016;387(10035):2323–30.

25. Sawhney S, Fluck N, Marks A, Prescott G, Simpson W, Tomlinson L, et al. Acute kidney injury-how does automated detection perform? Nephrol Dial Transplant. 2015;30(11):1853–61.

26. Suspected cancer: recognition and referral | Guidance and guidelines | NICE.

27. Murphy DR, Wu L, Thomas EJ, Forjuoh SN, Meyer AND, Singh H. Electronic Trigger-

Based Intervention to Reduce Delays in Diagnostic Evaluation for Cancer: A Cluster Randomized Controlled Trial. J Clin Oncol. 2015;33(31):3560–7.

28. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions. Ann Intern Med. 2006;144(3):201.

29. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). BMC Med Inform Decis Mak. 2011;11(1):62.

30. Usher-Smith JA, Harshfield A, Saunders CL, Sharp SJ, Emery J, Walter FM, et al. External validation of risk prediction models for incident colorectal cancer using UK Biobank. Br J Cancer. 2018;118(5):750–9.

31. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. BMC Med. 2010;8(1):21.

32. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. Macleod MR, editor. PLoS Med. 2012;9(5):e1001221.

APPENDIX 1 – Publication from thesis

Koshiaris C, Oke J, Abel L, Nicholson BD, Ramasamy K, Van den Bruel A. Quantifying intervals to diagnosis in myeloma: a systematic review and meta-analysis. BMJ Open. 2018 Jun 22;8(6):e019758.

BMJ Open

Quantifying intervals to diagnosis in myeloma: a systematic review and meta-analysis

Constantinos Koshiaris,[1] Jason Oke,[1] Lucy Abel,[1] Brian D Nicholson,[1] Karthik Ramasamy,[2,3] Ann Van den Bruel[1]

To cite: Koshiaris C, Oke J, Abel L, *et al*. Quantifying intervals to diagnosis in myeloma: a systematic review and meta-analysis. *BMJ Open* 2018;**8**:e019758. doi:10.1136/bmjopen-2017-019758

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-019758).

This study has been presented as an oral presentation at the SAPC Conference, Dublin, July 2016.

Received 25 September 2017
Revised 12 April 2018
Accepted 16 April 2018

Check for updates

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[2]Department of Haematology, Oxford University Hospitals NHS Trust, Oxford, UK
[3]National Institute for Health Research (NIHR) Biomedical Research Center Blood Theme, Oxford, UK

Correspondence to
Mr Constantinos Koshiaris; constantinos.koshiaris@phc.ox.ac.uk

ABSTRACT

Objectives To quantify the duration of each step of the diagnostic pathway for patients with multiple myeloma from symptom onset to confirmation of diagnosis.

Design Systematic review and meta-analysis.

Data sources and selection criteria The MEDLINE and Embase databases were searched up until January 2018 to identify articles that reported time intervals from onset of symptoms to diagnosis. Articles focusing on children or adolescents and on the asymptomatic form of the disease (monoclonal gammopathies and smouldering myeloma) were excluded.

Data collection and data analysis Data were extracted independently by two reviewers. Weighted estimates of the median and IQR were calculated. Risk of bias was assessed using the Aarhus checklist.

Main results Nine studies were included. The patient interval (first symptom to first presentation) had a median of 26.3 days (IQR: 1–98, n=465, two studies). Subsequently, the primary care interval (first presentation to first referral) was 21.6 days (IQR: 4.6–55.8, n=326, two studies), the diagnostic interval (first presentation to diagnosis) was 108.6 days (IQR: 33.3–241.7, n=5395, seven studies) and the time to diagnosis (first symptom to diagnosis) interval was 163 days (IQR: 84–306, n=341, one study). No studies reported data for the referral to diagnosis interval.

Conclusion The review demonstrates that there is scope for significant reductions in the time to myeloma diagnosis. At present, many patients experience a diagnostic interval longer than 3 months until diagnosis is confirmed.

Review registration Not available. Protocol available in the appendix.

Strengths and limitations of this study

- ► First systematic review to quantify the whole diagnostic pathway for patients with multiple myeloma including the different intervals in each step of the pathway.
- ► Use of all available information including the IQR rather than focusing on measures of central tendency like the mean and the median.
- ► No universally accepted methods for formal meta-analysis of median and IQR.
- ► Limited number of studies reporting the patient and primary care intervals, and no studies reported the referral to diagnosis interval, so any inferences regarding the referral to diagnosis interval should be interpreted with caution.

INTRODUCTION

Myeloma is a haematological malignancy characterised by uncontrolled plasma cell production in the bone marrow. It was the 17th most common cancer in the UK in 2013 accounting for 2% of all new cancer cases. Currently, there are more than 17 500 patients with myeloma in the UK with approximately 5500 cases being diagnosed every year.[1 2] It is a cancer that mainly affects the elderly population with 59% of the patients being diagnosed over the age of 70 years[2] and with a 5-year survival of 47%.[3]

It is considered one of the hardest cancers to suspect in primary care. Symptoms of myeloma are very common and non-specific, such as back pain, bone pain, fatigue and repeated infections.[4] This in combination with the fact that myeloma is a very rare condition in primary care results in very low predictive values for individual symptoms. For example, primary care patients with back pain, one of the most common myeloma symptoms, only have a 0.1% risk of myeloma.[5] By comparison, patients with rectal bleeding have a 2.4% risk of colorectal cancer.[6]

As a result, half of patients with symptomatic myeloma have three or more consultations in primary care before they are referred to specialist care, which is more than in any other cancer.[7] Attributing symptoms to comorbidities further prolongs the diagnostic process, which is particularly relevant in this older age group.[8 9]

Delays in diagnosing myeloma allow complications to develop (end organ damage), such as pathological fractures, irreversible renal failure and in some cases spinal cord compression.[10–12] These are considered medical emergencies in their own right and limit the opportunity for initiating effective

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

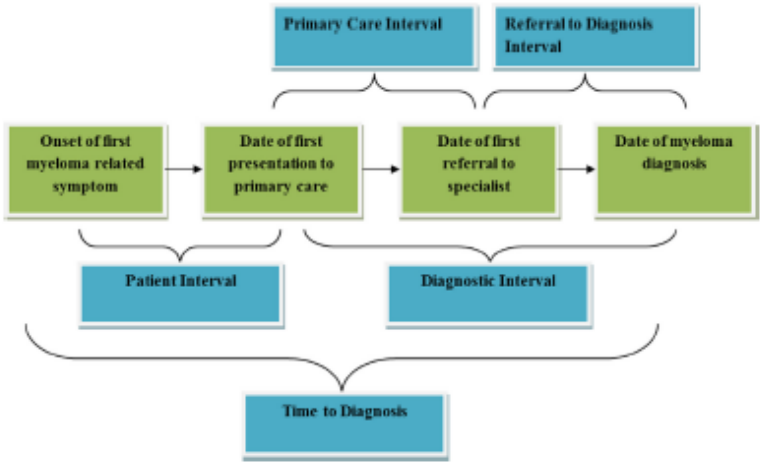


Figure 1 Outcome definition.

treatment.[13] A delayed diagnosis is also linked with higher cancer stage,[14 15] which is in turn associated with poorer survival.[16] Patients with longer diagnostic intervals also experience shorter disease free survival and more complications from treatment.[14]

Quantifying the time intervals leading up to diagnosis is important as it will inform future interventions that aim to shorten the diagnostic process. The aim of this systematic review was to quantify each step of the diagnostic pathway to myeloma diagnosis and identify where to focus efforts to reduce diagnostic delay.

METHODS

A protocol is available in the online supplementary appendix A1. A copy of the search strategy can be seen in the online supplementary appendix A2. We searched Embase and MEDLINE until January 2018 for studies that quantified any or all of the following five intervals[17]: the patient interval (from symptom onset to first consultation); the primary care interval (from first consultation for that symptom to referral to secondary care); the diagnostic interval (from first consultation with a myeloma-related symptom to diagnosis) and the time to diagnosis (from symptom onset to diagnosis). In addition, we looked for studies that estimated the referral to diagnosis interval (figure 1). Citation searching of key references like the Aarhus statement was conducted, and we also searched the reference list of systematic reviews with similar research questions.[18 19] We included any study designs that quantified at least one of the intervals mentioned above in days or months. Studies reporting the length of an interval only in number of consultations or referrals were excluded as were studies focusing on children or adolescents (<18 years) and on the asymptomatic forms of the disease (monoclonal gammopathies and smouldering myeloma). We included papers with an abstract in English but did not exclude full-text articles based on language (as long as there was an English abstract). Conference abstracts were excluded. Two reviewers (CK/LA) selected papers for inclusion using the criteria listed above, on title and abstracts first and on full text second. Disagreements were resolved through discussion with a third reviewer (JO/AVdB).

Data extraction

Two reviewers (CK/LA) independently extracted data from the included studies into a predefined spreadsheet. Study characteristics including author, year of publication, country of data collection, type of study, myeloma-related symptoms and sample size were extracted, as well as descriptive statistics including median, IQR, range, mean and the SD for each interval. Authors were contacted if data were not available or not in the appropriate format for extraction (ie, categorical rather than continuous).

Risk of bias assessment

The risk of bias was assessed by two independent researchers (CK/LA) using the Aarhus checklist.[18] The Aarhus checklist is a 20-item tool designed to help researchers design and evaluate studies on early diagnosis of cancer. It examines studies in terms of acknowledgement of the different biases influencing time point measurement and interval definition.

Analysis

In the context of illness duration, intervals are usually not normally distributed; therefore, we used the median and IQR to summarise the data. We present the 25th, 50th and 75th percentiles for all intervals. For intervals reported by more than one study, the pooled estimate was calculated by taking a weighted mean for each percentile. The

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

Table 1 Study characteristics

Study design	Study period	Population characteristics (age and gender)	Patients with myeloma (total)	Outcome measure (interval)
Friese *et al* (USA)[8]				
Retrospective analysis	1992–2002	Mean age 76.3 years 46% males	3831	Diagnostic
Howell *et al* (UK)[31]				
Survey	2004–2011	Median age 69.9 years 66.9% males	341	Patient diagnostic Time to diagnosis
Lyratzopoulos *et al* (UK)[26]				
Audit data	2009–2010	Not reported	176	Primary care
Varga *et al* (Hungary)[29]				
Retrospective analysis	Not reported	Median age 60 years 50% males	193	Diagnostic
Neal *et al* (UK)[25]				
Retrospective analysis	2001–2002	Mean age 72 years 53% males	221	Diagnostic
Din *et al* (UK)[24]				
Retrospective analysis	2007–2010	Median age 72 years 56% males	500	Diagnostic
Lyratzopoulos *et al* (UK)[27]				
Audit data	2009–2010	Not reported	124	Patient
Goldschmidt *et al* (Israel)[30]				
Retrospective analysis	2002–2011	Median age 63 years 53% males	107*	Diagnostic
Swann *et al* (UK)[28]				
Audit data	2014	Not reported	202	Primary diagnostic

*The total sample size for this study was 110 patients out of which seven were diagnosed with plasmacytoma. The analysis was conducted on 107 patients that had complete data.

weight was obtained by dividing the sample size in each study with the total numbers of patients. We also fitted a distribution through the three weighted percentile estimates where appropriate in order to generate the shape of the distribution of the interval under investigation. We chose the lognormal distribution as time intervals are usually skewed to the right.[4] A prespecified sensitivity analysis was conducted by excluding the study with the higher risk of bias. The sensitivity analysis was conducted only for the diagnostic interval as the rest of the outcomes were reported by only one or two papers, which can be seen in table 1.

Patient and public involvement

No patients or public were involved in this study.

RESULTS

We identified 3343 citations from the Embase and MEDLINE searches. After removal of conference abstracts and duplicates, we screened 1271 titles and abstracts, and 16 studies were candidate for inclusion. Nine studies were included in the final analysis (figure 2). Seven papers in total were excluded for the following reasons: reported the duration in numbers of consultations (n=2)[3 7]; none of the prespecified outcomes were reported (n=1)[20]; reported the same outcome based on the same database as one of the other included papers, so inclusion of this paper in the data synthesis would result in double counting the same patients (n=1)[21]; data were not in an appropriate format (n=2)[14 22]; and because the interval under investigation was reported only for patients that were referred to very specific departments making it a very selective population compared with the other studies (n=1).[25]

Study characteristics

A summary of all the included papers is provided in table 1. Studies were published between 2009 and 2018, and the sample size ranged from 107 to 3831 patients. Five studies reported intervals in various cancers,[24–28]

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

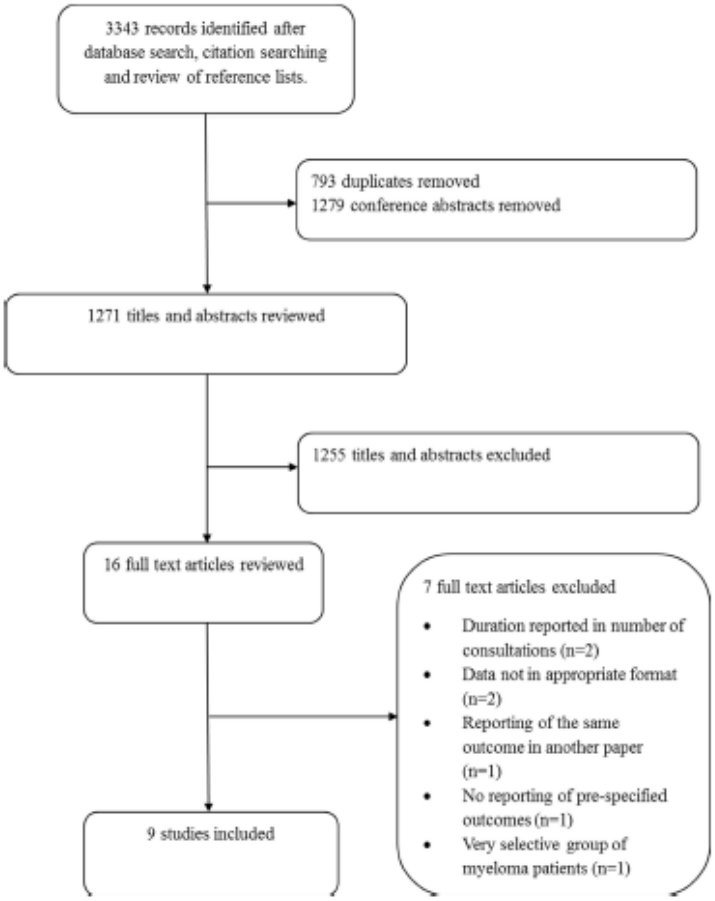


Figure 2 Study selection flow chart.

three reported only myeloma[8 29 30] and one for haematological malignancies.[31]

Six studies were conducted in the UK,[24–28 31] one in the USA,[8] one in Hungary[29] and one in Israel.[30] Two UK studies used data from two separate Clinical Practice Research Datalink (CPRD) cohorts,[24 25] a database of routinely collected electronic primary care records. Two other UK studies used data from the English National Audit of Cancer Diagnosis in Primary Care 2009–2010[32]; we extracted primary care interval data from the larger study and the patient interval from the smaller study.[26 27] Another UK study used data from the English National Audit of Cancer Diagnosis in Primary Care 2014, and the last UK study was a patient survey on patients diagnosed with haematological malignancies.[28 31] The study conducted in Hungary analysed data collected from patients treated in a haematology centre, and the study conducted in the USA analysed a retrospective database collected from the Surveillance, Epidemiology, and End Results programme (SEER).[8 29] The study conducted in Israel used data from the Israeli health maintenance organisation linked with the Israel National Cancer Registry.[30]

Definition of intervals to diagnosis

There was substantial heterogeneity in the symptoms and time points used to define each interval (table 2). In total, 19 different symptoms were used to define the start of myeloma but studies varied greatly regarding which symptom (or symptoms) were used, ranging from 3 to 12 symptoms. Three studies did not report the starting symptoms.[26–28] Some studies included multiple symptoms in more general categories.[24 25 31] For example, Howell *et al* used a general pain category that included

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

Table 2 Symptoms and date definitions

Symptoms used	Onset of first symptom	Date of first presentation in healthcare services	Date of first referral	Date of diagnosis
Friese *et al*[8] (USA)				
Anaemia Packed red blood cell transfusion Back pain	NA	1 year before diagnosis	NA	SEER cancer diagnosis date
Howell *et al*[21] (UK)				
Tiredness Pain Shortness of breath Infections Joint problems/fractures Stomach/bowel symptoms Other	Patient reported	Patient reported	NA	Date provided by the haematological malignancy Diagnostic service
Lyratzopoulos *et al*[26] (UK)				
Not reported	Estimated based on patient's clinical records	2 years before diagnosis	Date that the referral letter was sent	Clinical records and hospital correspondence
Varga *et al*[29] (Hungary)				
Bone symptoms Anaemia Renal failure General symptoms Other Tumour presence Metastatic bone disease	NA	3 years before diagnosis	NA	Tertiary haematology centre
Neal *et al*[25] (UK)				
Bleeding Bone pain Bruising Anaemia Fatigue Anorexia Weight loss	NA	1 year before diagnosis	NA	First occurrence of a myeloma Read Code in the patient's record in CPRD database
Din *et al*[24] (UK)				
Bleeding Bone pain Bruising Anaemia Fatigue Anorexia Weight loss	NA	1 year before diagnosis	NA	First occurrence of a myeloma Read Code in the patient's record in CPRD database
Lyratzopoulos *et al*[27] (UK)				
Not reported	NA	2 years before diagnosis	Date that the referral letter was sent	Clinical records and hospital correspondence
Goldschmidt *et al*[30] (Israel)*				
Pain (back, cervical spine, musculoskeletal, non-specific) Infection Weight loss Fatigue Peripheral oedema Constipation Presyncope Syncope Dizziness	NA	2 years before diagnosis	NA	Israel National Cancer Registry
Swann *et al*[28] (UK)				
Not reported	NA	2 years before diagnosis	Date that the referral letter was sent	Hospital Episode Statistics

NA-Not applicable

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

musculoskeletal, abdominal, chest and other type of pains, while the two CPRD studies included multiple musculoskeletal symptoms under a general bone pain category.[24 25] In addition, studies using CPRD or SEER data were using predefined symptoms to identify the onset of disease, while other studies like Howell *et al* documented the full range of symptoms reported by the patients during this time.

The start of the measuring period was defined as the date of onset of the first symptom or the date of first presentation for a myeloma-related symptom depending on whether the studies were investigating the patient interval, the diagnostic or both. Study authors used various prediagnostic time intervals to identify the first symptom (at 1, 2 or 3 years before diagnosis). Three identified the first symptom at 1 year before diagnosis, three at 2 years and one at 3 years. One study used patient-reported dates. Goldschmidt *et al*[30] did not use the first symptom as the start of the measurement period, but they used the first combination of symptom and laboratory result (ie, the earliest of blood test+pain complaint or two blood tests within a month or two pain complaints within 1–3 months).

Risk of bias

Most of the studies included in the analysis had a low risk of bias (online supplementary appendix A3 and A4). All studies clearly defined the start and end point of the intervals, and in most cases there was an adequate description of the databases along with the strengths, limitations and biases arising from the definitions of the different intervals and time point. Only one study did not mention the different limitations and biases arising from the study design and the choice of definitions for time points and intervals.[29] Most common sources of bias that were described included recall bias for studies that were using patient reported data and misclassification bias for studies that were using databases like CPRD. Most studies used a theoretical framework to define each interval usually the one reported by Olesen *et al*[17] or the Aarhus statement.[18] The category with the higher risk of bias was the use of a hierarchical rationale to determine the date of diagnosis, that is, date of first histological confirmation of the malignancy or date of admission to the hospital, for example. Most studies mentioned how the date of diagnosis was obtained, but there was no adequate description on how the choice of a particular definition can affect the diagnostic pathway.

Quantifying intervals

Seven papers reported the diagnostic interval,[8 24 25 28–31] two papers reported the patient interval,[27 31] two papers reported the primary care interval[26 28] and one paper reported the time to diagnosis interval.[31] No studies reported the referral to diagnosis interval. The length of the different intervals can be seen in table 3, and the fitted log normal distributions are shown in figure 3 along with the parameters used to fit them.

For the diagnostic interval, the pooled weighted mean of the 50th percentile is 108.6 days (n=5398) and the IQR is from 33.3 to 241.7 (n=5288). Removing the study with the largest risk of bias[29] based on the Aarhus statement checklist did not alter the results (107.9 days, IQR: 31.3–242.2). An additional sensitivity analysis was conducted by excluding the Goldschmidt *et al*[30] study as it was an outlier, but the results were not affected (pooled median of 103.8). While all the studies reported a median diagnostic interval less than 5 months, this study had an interval of 11.2 months. The IQR was not estimated for this sensitivity analysis as it was not reported by the authors.

The pooled estimate of the 50th percentile of the patient interval is 26.3 days (IQR: 0.7–97.7, n=465). The primary care interval was reported by two studies[26 28] with a median of 21.6 days (IQR: 4.6–55.8, n=326), and the time to diagnosis interval was also reported by one study[31] with a median of 163 days (IQR: 84–306, n=341).

No study reported the referral to diagnosis interval, but it can be inferred by subtracting the median length of the primary care and patient intervals from the diagnostic interval or the time to diagnosis interval. The median length of the referral to diagnosis interval can range from 60.7 to 115.1 days depending on whether we use the diagnostic or the time to diagnosis interval for the inference.

Table 3 Length of intervals

Percentile	N	25th	50th	75th
Patient interval				
Howell *et al*[31]	341	1	31	122
Lyratzopoulos *et al*[27]	124	0	13.5	31
Weighted estimate	465	0.7	26.3	97.7
Primary care interval				
Lyratzopoulos *et al*[26]	176	5	21	55
Swann *et al*[28]	150	4.2	23.5	56.8
Weighted estimate	326	4.6	21.6	55.8
Referral to diagnosis interval				
No papers reporting this interval				
Diagnostic interval				
Friese *et al*[8]	3831	27	99	252
Howell *et al*[31]	341	34	83	167
Varga *et al*[29]	193	88	125	230
Neal *et al*[25]	221	56	144	264
Din *et al*[24]	500	54	149	263
Goldschmidt *et al*[30]	107	NR	341	NR
Swann *et al*[28]	202	24	53.5	107.5
Weighted estimate	5395	33.3	108.6	241.7
Time to diagnosis				
Howell *et al*[31]	341	84	163	306

NR, not reported.

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

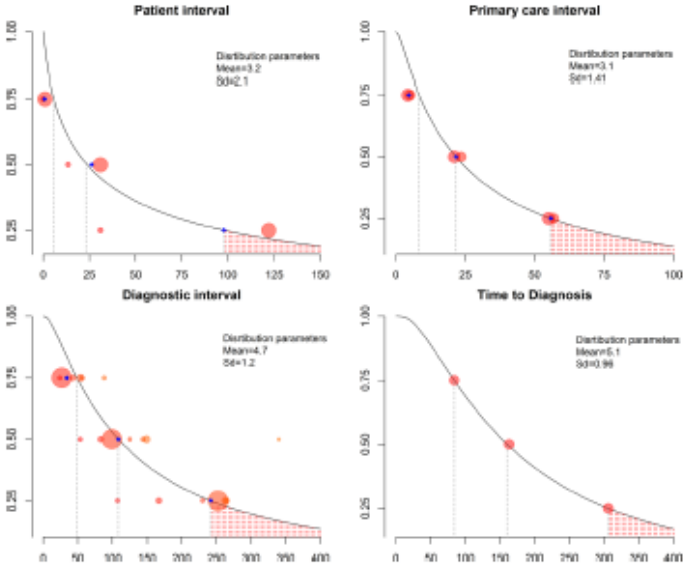


Figure 3 Distribution of the intervals. Legend: each circle corresponds to one study, and the size is proportional to the total sample size. The blue diamond corresponds to the weighted estimate. For intervals with only one study (time to diagnosis), no weighted estimates were calculated. Y-axis corresponds to 1−probability (interval>number of days), that is 0.25 corresponds to the 75th percentile and 0.75 to the 25th percentile. X-axis corresponds to number of days.

DISCUSSION

Our results show that patients with myeloma experience symptoms for a median of approximately 1 month before seeking help, and 25% of patients wait for more than 3 months (98 days). After attending primary care with symptoms, the median time to diagnosis is 108.6 days (IQR: 33.3–241.7) with 25% of patients waiting longer than 8 months. No studies report the referral to diagnosis interval

Strengths and weaknesses

This is the first systematic review that quantified the patient pathway of myeloma from onset of first symptom to diagnosis. There were no restrictions in the search strategy in terms of study design or healthcare systems. We focused our search on two medical databases that are more likely to contain papers on diagnostic pathways, but we acknowledge that this might have affected the identification of all literature. To counter this, we included additional strategies like citation searching of some key references and searching the reference lists of similar reviews.

We excluded conference abstracts, although there were several that addressed the review question. The reason for this is that the length of the different intervals reported is affected by design decisions like the choice of when to start the measurement period, initial symptoms, data collection methods and so on. Conference abstracts do not report this level of detail in their methods and therefore could not be included.

In addition to measures of central tendency like the median, we included the 25th and 75th percentiles, which are particularly important since time interval data are skewed to the right. Examining all three can provide a more complete idea of the delays that the patients experience especially at the tails of the distribution. Measures of central tendency like the mean might not be the most appropriate to describe the distribution as they tend to be overestimated when used on positively skewed distributions which also makes comparison with other cancer intervals more difficult as they are usually quantified using the median and IQR.

The main limitation of the analysis is that currently there are no universally accepted formal methods to perform meta-analysis of medians and IQRs. To overcome this, we combined estimates of the percentiles after weighting them based on the sample size: a method equivalent to a fixed effects meta-analysis. Our estimates of the diagnostic interval might therefore be an underestimation due to the fact that the biggest study reported one of the lowest diagnostic intervals.[8] This however does not change the interpretation of the results as these estimates still suggest very long diagnostic intervals for patients with myeloma. We were not able to produce a CI around the median and IQR as these are not usually measures that are reported by the included studies; thus, we present only the point estimate of each percentile. In addition, there are no formal ways of

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

estimating statistical heterogeneity when meta-analysing median and IQR.

No studies reported the referral to diagnosis interval, and it was inferred based on the results of the other interval so our results regarding this interval should be interpreted with caution.

Sources of heterogeneity

As mentioned in the strengths and limitations, no formal ways of estimating heterogeneity currently exist when performing meta-analysis of medians and IQR. In order to get an approximate measure of heterogeneity, we also performed a meta-analysis of the means for which we had CIs or we could approximate (online supplementary appendix A5), which resulted in an I^2 statistics of 98.6% (diagnostic interval). Although we expect high heterogeneity due to various design decisions that are described below, this statistic should be interpreted with caution as it might be an overestimation. We believe that to be the case because of the very small uncertainty for each within-study estimate. This results in very narrow CIs around each study that do not overlap and thus artificially inflate the I^2 statistic. In addition for three out of seven studies, either the means or the CIs had to be approximated, which could potentially be introducing more bias on the effect and heterogeneity estimates. Heterogeneity estimates might have been different if we were able to obtain CIs around median and IQR. We believe that clinical heterogeneity is more important in this case.

In order to compute intervals, the definition of the beginning and the end of the interval is crucial. There was variability in how studies defined starting points, especially for the first symptom and the first presentation to healthcare, using medical records or patient recall. Studies that use patient reported outcomes tend to suffer from recall bias, which might lead to overestimation or underestimation of the different intervals, while studies using medical records tend to suffer from loss to follow-up and misclassification.

For studies that used electronic health records, there was no agreement on when exactly myeloma starts to manifest. The time used to detect related symptoms prior to diagnosis spanned from 1 to 3 years. Although most studies used 1 year before diagnosis, there is some evidence to suggest that symptoms might be present for more than 1 year,[33] which may have led to an underestimation of intervals in these studies. However, the more you extend the symptom period, the more likely you are to detect symptoms that are unrelated to myeloma, which leads to the overestimation of the length of the intervals, especially with symptoms that are so non-specific such as back pain. In order to explore this, we conducted a sensitivity analysis where we estimated the length of the diagnostic interval by stratifying according to the time used to define the presenting symptoms with studies identifying the symptom at 1 year before diagnosis having a median of 105 days versus more than 1 year having a median of 142.3 day, which might explain some of the observed variability.

Even though there are various sources of heterogeneity, all the sensitivity analyses that were conducted did not change the result trend and their interpretation, as almost all studies reported diagnostic intervals longer than 3 months irrespectively of the way the study was conducted.

Findings compared with existing research

Our estimate for the patient interval is in-line with the findings of another study,[22] which reported that 15% of patients with myeloma wait more than 3 months before they go to the doctor. This study explained this delay in terms of patients' lack of understanding of the seriousness of their symptoms because of their non-specific nature.

Patients with myeloma experience the longest primary care interval out of all cancers with a median of 21.6 days. Other cancers that have been shown to have long primary care intervals include renal and lung with a median of 14 days.[28] The long primary care interval for patients with myeloma could be explained by the fact that symptoms on their own are not enough for referral, and multiple blood tests need to be conducted like a full blood count, calcium, creatinine and inflammatory markers. Conducting multiple tests has been shown to extend the primary care interval.[34]

The diagnostic interval, which takes place in healthcare and could potentially be amenable to improvement, is longer for myeloma than for many other cancers. It has been shown that only 17.2% of patients with myeloma are referred through the suspected cancer referral pathway ('2 week' wait), which is lower than other cancers such as breast cancer for example (43%).[20 35] This could be due to the non-specific nature of the symptoms that make it hard for both the general practitioner (GP) and the patient to suspect the presence of myeloma. This might also explain the difference in the length of the diagnostic intervals between these two cancers as the median diagnostic interval for breast cancer is approximately 14 days.[36] Other cancers with a similarly long median diagnostic interval also have non-specific clinical presentations, such as lung cancer, which has a diagnostic interval of 88 days.[36]

Implications for future research

We were not able to estimate the referral to diagnosis care interval directly, although it is reasonable to believe that it might be longer than the primary care. As in other cancers, referrals to different specialties or with an insufficient level of urgency or multiple referrals can prolong the referral to diagnosis interval.[37] The choice of referral route has also been shown to be a strong predictor of the length of the diagnostic interval, that is, patients that are diagnosed throught a referral pathway for cancer tend to have shorter intervals.[38] Future studies should estimate the duration of primary and referral to diagnosis intervals and investigate the impact of one setting on the other as in most cases the

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

type and severity of symptoms will determine the specialty and urgency of referral.

Also it is still not clear how long before diagnosis myeloma symptoms start to occur. In lung cancer, studies on symptom lead time (the time between symptoms attributable to cancer and diagnosis) show that symptom incidence increases considerably 6 months before diagnosis, but no such study has been conducted for myeloma.[39]

CONCLUSION

Myeloma is a complex disease to diagnose due to a combination of different factors. First, myeloma symptoms (like back pain and fatigue) are common and mostly caused by benign conditions, resulting in patients not visiting their doctor. In addition the rarity of the disease, makes it hard for GPs to suspect this cancer. There is no effective screening programme for myeloma as this might result in people having a lot of unnecessary tests and potentially over diagnosing monoclonal gammopathy of undetermined significance (MGUS) thus any benefits from the screening programme cannot outweigh the cost. The reasons described above can explain why certain patients with myeloma tend to experience long diagnostic intervals. Our results indicate that in some cases the length of the diagnostic interval can be over 8 months. There is potential for meaningful reductions in the time to diagnosis especially for the diagnostic interval, which could improve patient outcomes, but more research is required in order to do that. Further and more in-depth exploration of the diagnostic pathway is required especially for the intervals we were not able to explore in this study like the referral to diagnosis interval and its link with the primary care interval. Development of interventions that aim to reduce the length of the diagnostic interval are now needed.

Contributors Six authors contributed to this study: CK, JO, LA, BDN, KR, AWdB. CK designed the study with input from AWdB and JO. All authors were involved in the conduct of the study, interpreting the results and in revising and correcting the manuscript. CK and LA reviewed and extracted the data from articles. CK and JO planned and conducted the analysis. The manuscript drafting was led by CK with the contribution of all authors. All authors read and approved the final version of the manuscript.

Funding This manuscript presents work carried out as part of a DPhil scholarship awarded to CK funded by the Primary Care Research Trust, The University of Oxford and NIHR Oxford CLAHRC. AWdB is supported through the National Institute for Health Research (NIHR) Diagnostic Evidence Co-operative Oxford at Oxford Health Foundation Trust (IS_DEC_0812_100).

Disclaimer This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The dataset is available on request from the corresponding author.

Open access This is an open access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. England PH. National Cancer Intelligence Network Cancer statistics : availability and location About Public Health England. 2015.
2. Cancer Research UK. Myeloma statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma#heading-Six (accessed 26 Aug 2016).
3. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: a population-based study. *Lancet* 2015;385:1206–18.
4. Lyratzopoulos G, Wardle J, Rubin G. Rethinking diagnostic delay in cancer: how difficult is the diagnosis? *BMJ* 2014;349:g7400.
5. Shephard EA, Neal RD, Rose P, *et al*. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. *Br J Gen Pract* 2015;65:e106–13.
6. Hamilton W, Round A, Sharp D, *et al*. Clinical features of colorectal cancer before diagnosis: a population-based case–control study. *Br J Cancer* 2005;93:399–405.
7. Lyratzopoulos G, Neal RD, Barbiere JM, *et al*. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;13:353–65.
8. Friese CR, Abel GA, Magazu LS, *et al*. Diagnostic delay and complications for older adults with multiple myeloma. *Leuk Lymphoma* 2009;50:392–400.
9. Macdonald S, Macleod U, Campbell NC, *et al*. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *Br J Cancer* 2006;94:1272–80.
10. Tosi P. Diagnosis and treatment of bone disease in multiple myeloma: spotlight on spinal involvement. *Scientifica* 2013;2013:1–12.
11. Sonmez M, Akagun T, Topbas M, *et al*. Effect of pathologic fractures on survival in multiple myeloma patients: a case control study. *J Exp Clin Cancer Res* 2008;27:11–14.
12. Wirk B. Renal failure in multiple myeloma: a medical emergency. *Bone Marrow Transplant* 2011;46:771–83.
13. Smith D, Yong K. Multiple myeloma. *BMJ* 2013;346:f3863.
14. Kariyawasan CC, Hughes DA, Jayatillake MM, *et al*. Multiple myeloma: causes and consequences of delay in diagnosis. *QJM* 2007;100:635–40.
15. Ong F, Hermans J, Noordijk EM, *et al*. Presenting signs and symptoms in multiple myeloma: high percentages of stage III among patients without apparent myeloma-associated symptoms. *Ann Hematol* 1995;70:149–52.
16. Greipp PR, San Miguel J, Durie BG, *et al*. International staging system for multiple myeloma. *J Clin Oncol* 2005;23:3412–20.
17. Olesen F, Hansen RP, Vedsted P. Delay in diagnosis: the experience in Denmark. *Br J Cancer* 2009;101:S5–8.
18. Weller D, Vedsted P, Rubin G, *et al*. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer* 2012;106:1262–7.
19. Neal RD, Tharmanathan P, France B, *et al*. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 2015;112:S92–107.
20. Elliss-Brookes L, McPhail S, Ives A, *et al*. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer* 2012;107:1220–6.
21. Keeble S, Abel GA, Saunders CL, *et al*. Variation in promptness of presentation among 10,297 patients subsequently diagnosed with one of 18 cancers: evidence from a National Audit of Cancer Diagnosis in Primary Care. *Int J Cancer* 2014;135:1220–8.
22. Howell DA, Warburton F, Ramírez AJ, *et al*. Risk factors and time to symptomatic presentation in leukaemia, lymphoma and myeloma. *Br J Cancer* 2015;113:1114–20.
23. Li N, Lv Y, Zeng H, *et al*. Renal impairment in multiple myeloma: presenting features in different departments. *J Huazhong Univ Sci Technolog Med Sci* 2012;32:65–8.
24. Din NU, Ukoumunne OC, Rubin G, *et al*. Age and Gender Variations in Cancer Diagnostic Intervals in 15 Cancers: Analysis of Data from the UK Clinical Practice Research Datalink. *PLoS One* 2015;10:e0127717.

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

25. Neal RD, Din NU, Hamilton W, *et al*. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer* 2014;110:584–92.
26. Lyratzopoulos G, Abel GA, McPhail S, *et al*. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *Br J Cancer* 2013;108:686–90.
27. Lyratzopoulos G, Saunders CL, Abel GA, *et al*. The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *Br J Cancer* 2015;40:S35–S40.
28. Swann R, McPhail S, Witt J, *et al*. Diagnosing cancer in primary care: results from the National Cancer Diagnosis Audit. *Br J Gen Pract* 2018;68:e63–72.
29. Varga G, Mikala G, Andrikovics H, *et al*. [How long does a myeloma patient currently wait for the diagnosis in Hungary?]. *Orv Hetil* 2014;155:1538–43.
30. Goldschmidt N, Zamir L, Poperno A, *et al*. Presenting Signs of Multiple Myeloma and the Effect of Diagnostic Delay on the Prognosis. *J Am Board Fam Med* 2016;29:702–9.
31. Howell DA, Smith AG, Jack A, *et al*. Time-to-diagnosis and symptoms of myeloma, lymphomas and leukaemias: a report from the Haematological Malignancy Research Network. *BMC Hematol* 2013;13:9.
32. RCGP. *National Audit of Cancer Diagnosis in Primary Care*. London: Royal College of General Practitioners, 2011.
33. Corner J, Hopkinson J, Fitzsimmons D, *et al*. Is late diagnosis of lung cancer inevitable? Interview study of patients' recollections of symptoms before diagnosis. *Thorax* 2005;60:314–9.
34. Rubin GP, Saunders CL, Abel GA, *et al*. Impact of investigations in general practice on timeliness of referral for patients subsequently diagnosed with cancer: analysis of national primary care audit data. *Br J Cancer* 2015;112:676–87.
35. Howell D, Smith A, Appleton S, *et al*. Multiple myeloma: routes to diagnosis, clinical characteristics and survival - findings from a UK population-based study. *Br J Haematol* 2017;177:67–71.
36. Redaniel MT, Martin RM, Ridd MJ, *et al*. Diagnostic intervals and its association with breast, prostate, lung and colorectal cancer survival in England: historical cohort study using the Clinical Practice Research Datalink. *PLoS One* 2015;10:e0126608–17.
37. Barrett J, Hamilton W. Pathways to the diagnosis of lung cancer in the UK: a cohort study. *BMC Fam Pract* 2008;9:31.
38. Jensen H, Tørring ML, Olesen F, *et al*. Cancer suspicion in general practice, urgent referral and time to diagnosis: a population-based GP survey and registry study. *BMC Cancer* 2014;14:636.
39. Ades AE, Biswas M, Welton NJ, *et al*. Symptom lead time distribution in lung cancer: Natural history and prospects for early diagnosis. *Int J Epidemiol* 2014;43:1865–73.

BMJ Open: first published as 10.1136/bmjopen-2017-019758 on 22 June 2018. Downloaded from http://bmjopen.bmj.com/ on 12 July 2018 by guest. Protected by copyright.

APPENDIX 2 - Publication from thesis

Koshiaris C, Van den Bruel A, Oke JL, Nicholson BD, Shephard E, Braddick M, et al. Early detection of multiple myeloma in primary care using blood tests: a case-control study in primary care. Br J Gen Pract. 2018 Aug 13;bjgp18X698357

Constantinos Koshiaris, Ann Van den Bruel, Jason L Oke, Brian D Nicholson, Elizabeth Shephard, Mick Braddick and William Hamilton

Early detection of multiple myeloma in primary care using blood tests:

a case–control study in primary care

Abstract

Background
Multiple myeloma is a haematological cancer characterised by numerous non-specific symptoms leading to diagnostic delay in a large proportion of patients.

Aim
To identify which blood tests are useful in suggesting or excluding a diagnosis of myeloma.

Design and setting
A matched case–control study set in UK primary care using routinely collected data from the Clinical Practice Research Datalink.

Method
Symptom prevalence and blood tests were analysed up to 5 years before diagnosis in 2703 cases and 12 157 matched controls. Likelihood ratios (LR) were used to classify tests or their combinations as useful rule-in tests (LR+ ≥5), or rule-out tests (LR− ≤0.2).

Results
Raised plasma viscosity (PV) had an LR+ = 2.0, 95% confidence interval (CI) = 1.7 to 2.3; erythrocyte sedimentation rate (ESR) 1.9, 95% CI = 1.7 to 2.0; and C-reactive protein (CRP) 1.2, 95% CI = 1.1 to 1.4. A normal haemoglobin had an LR− = 0.42, 95% CI = 0.39 to 0.45; calcium LR− = 0.81, 95% CI = 0.78 to 0.83; and creatinine LR− = 0.80, 95% CI = 0.77 to 0.83. The test combination with the lowest LR− was all normal haemoglobin with calcium and PV, which had an LR− = 0.06, 95% CI = 0.02 to 0.18, though the LR− for normal haemoglobin and PV together was 0.12 (95% CI = 0.07 to 0.23).

Conclusion
Plasma viscosity and ESR are better for both ruling in and ruling out the disease compared with C-reactive protein. A combination of a normal ESR or PV and normal haemoglobin is a simple rule-out approach for patients currently being tested in primary care.

Keywords
blood; diagnosis; case–control studies; inflammatory; multiple myeloma; primary care.

INTRODUCTION

Multiple myeloma is a rare malignancy, characterised by clonal proliferation of plasma cells. These cells secrete immunoglobulins (paraproteins), which can lead to plasma hyperviscosity and renal damage. Proliferation of plasma cells can lead to bone marrow suppression, and may cause hypercalcaemia. These various features of myeloma give rise to different symptoms, such as bone pain from direct skeletal involvement, fatigue from anaemia, or headache from hyperviscosity. Presentation with complications from hypercalcaemia or renal failure is also common.

Diagnosis of myeloma is often difficult. Patients with myeloma have the longest intervals from initial symptom reporting to diagnosis of all common cancers, with the most consultations in primary care before referral.[1,2] Longer diagnostic intervals in myeloma are associated with more advanced disease stages and more complications at diagnosis.[3,4] Patients who are not referred to the appropriate department generally experience a longer diagnostic process.[5] A large proportion of patients are diagnosed through emergency presentations, with concomitant worse survival.[4,7] A recent study reported that 77% of all myeloma emergency presentations had at least one primary care consultation before the emergency and 56% of these had at least three.[8]

This prolonged diagnostic process probably represents the non-specific nature of myeloma symptoms, with positive predictive values for symptoms <1%, even in combination.[9] Guidance from the National Institute for Health and Care Excellence (NICE) uses an urgent cancer threshold for referral of 3%.[10] In myeloma, symptoms need to be combined with abnormal blood results such as full blood counts (FBC), calcium, and inflammatory markers to reach that threshold. The inflammatory markers C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and plasma viscosity (PV), when considered together, have been used for diagnosing myeloma but have not been reported individually.[11–13]

The aims of this study are to identify the best inflammatory marker for initial investigation of possible myeloma, useful blood tests for ruling out symptomatic myeloma, and how to distinguish early and late features of the disease. Symptoms can occur up to 2 years before diagnosis in other cancers, but little is known about the timing of symptoms and abnormal blood test results before diagnosis in myeloma.[14,15] The latter can explain why some features have better rule-out properties than others as features that manifest very late in the diagnosis can be useful for ruling in the disease but not as useful for ruling it out.

C Koshiaris, MSc, statistician, **A Van den Bruel**, MD, PhD, associate professor of general practice; **JL Oke**, DPhil, senior statistician; **BD Nicholson**, MSc, MRCGP, clinical researcher, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford. **E Shephard**, PhD, CPsychol, research fellow; **W Hamilton**, MD, FRCP, FRCGP, professor of primary care diagnostics, University of Exeter Medical School, Exeter. **M Braddick**, MRCP, GP, Chiddenbrook Surgery, Crediton, Devon.

Address for correspondence
Constantinos Koshiaris, Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK.

Email: constantinos.koshiaris@phc.ox.ac.uk

Submitted: 20 November 2017; **Editor's response:** 1 February 2018; **final acceptance:** 14 March 2018.

©British Journal of General Practice
This is the full-length article (published online 14 Aug 2018) of an abridged version published in print. Cite this version as: **Br J Gen Pract 2018; DOI: https://doi.org/10.3399/bjgp18X698357**

How this fits in

Patients with myeloma experience some of the longest diagnostic intervals, which could be due to the non-specific nature of their symptoms. The results of this study suggest that the best inflammatory markers to supplement symptoms of myeloma are plasma viscosity and erythrocyte sedimentation rate, while C-reactive protein is unhelpful. In addition, the combination of a normal haemoglobin and plasma viscosity can be used to rule out the disease on patients currently being tested in primary care. These results further demonstrate that the full blood count components and especially low haemoglobin are among the earliest presentations of the disease, which may allow the diagnosis to be expedited.

METHOD

Study design

This is an extension of a 2015 study that quantified the risk of multiple myeloma from symptoms reported in primary care using records from the Clinical Practice Research Datalink (CRPD).[9] Briefly, the study had a case-control design with 2703 cases and 12 157 matched controls based on age, sex, and general practice. Cases were selected if patients were >40 years of age with a myeloma diagnosis between January 2000 and December 2009. Date of diagnosis was defined as the first myeloma Read code, which was also used as the index date for the controls. Exclusion criteria included cases or controls with <1 year of records, cases without controls, controls with myeloma, and controls that did not seek medical care after registration. The study identified eleven symptoms (back pain, chest pain, chest infections, shortness of breath, nausea, fracture, joint pain, bone pain, weight loss, rib pain, and nosebleed) and five abnormal blood investigations associated with myeloma (cytopenia, raised inflammatory markers, raised creatinine, raised mean corpuscular volume, and hypercalcaemia).

Variable definitions

Symptoms and investigations were identified using libraries of Read codes. Abnormal blood test results were defined as results above or below the local laboratory's normal range. Implausible test values were excluded. Combinations of blood tests, such as FBC and calcium, were defined as positive if at least one of the test results was abnormal. For the combination to be defined as negative, all tests in the combination were required to be normal. For the FBC, which has multiple constituent parts, for example, white cell count (WCC), haemoglobin, and platelets, a result was defined abnormal if at least one of the parts indicated cytopenia (low haemoglobin, WCC, or platelets).

Analysis

Symptom prevalence and blood tests were analysed up to 5 years before diagnosis. Each inflammatory marker was examined separately using conditional logistic regression in order to examine its association with myeloma compared with matched controls, its predictive performance was quantified using positive likelihood ratios (LR+). To identify the best combinations for ruling out myeloma, the negative likelihood ratio (LR–) was estimated for each test individually and in combination. Positive likelihood tests indicate how many times more likely a positive test occurs in individuals with the disease compared with individuals without the disease, and negative likelihood tests represent how much less likely a negative result will occur in individuals with the disease compared with a healthy individual. A useful rule-in test was defined as one with an LR+ >5 and a useful rule-out test one with an LR– <0.2.[14,17]

Each combination included a different inflammatory marker. The constituent parts of the FBC were examined either separately or altogether. The analysis was restricted to patients who had been tested or had a combination of tests.

To examine when symptoms and investigations start to indicate symptomatic myeloma and to distinguish between the early and late features, the 5-year period before diagnosis was broken down into 90-day intervals, and symptom occurrence and blood test results in these periods were identified. The results for each 90-day interval were summarised using proportions for binary variables (symptoms) and means for continuous variables (blood test results).

RESULTS

Baseline characteristics of the population under investigation can be found in the original report,[9] but, briefly, total number of patients was the same as the 2015 study, *N* = 14 860, and the median age was 73 years with 53% being male.

Within the year before diagnosis 55% (8221/14 860) participants had at least one blood test (either FBC, calcium, creatinine, or one of the inflammatory markers) with testing being more prevalent in cases (78%,

2097/2703) compared to controls (50%, 6124/12 157). The most common tests were creatinine (50%, 7407/14 860) and FBC (40%, 5934/14 860); the least common test was calcium (19%, 2809/14 860). Among the inflammatory markers the most common tests included ESR, followed by CRP and PV. The most common test combination was FBC with creatinine and ESR (13%, (1864/14 860).

At 3 years before diagnosis, 84% (2270/2703) of patients with myeloma had at least one blood test compared with 66% (8058/12 157) in the controls, and at 5 years before diagnosis 85% (2307/2703) of the cases had a test compared with 70% (8524/12 157) of the controls.

Rule-in tests

Associations between raised inflammatory markers and a myeloma diagnosis are shown on Table 1. During the year before diagnosis 85% (844) of patients with myeloma had an abnormal ESR compared with 46% (546) of controls (odds ratio (OR) 5.7; 95% CI = 4.1 to 8.0; $P<0.001$; LR+ 1.9, 95% CI = 1.7 to 2.0), 81% (162) had an abnormal PV compared with 41% (95) of controls (OR 4.4; 95% CI = 2.2 to 8.8; $P<0.001$; LR+ 2.0, 95% CI = 1.7 to 2.3), and 46% (294) had an abnormal CRP compared with 37% (238) of controls (OR 1.5; 95% CI = 1.0 to 2.2; $P = 0.05$; LR+ 1.2, 95% CI = 1.1 to 1.4). The rule-in properties for the rest of the tests that are published in the original report included cytopenia (OR 5.4, 95% CI = 4.6 to 6.4), raised creatinine (OR 1.8, 95% CI = 1.5 to 2.2), raised mean corpuscular volume (MCV) (OR 3.1, 95% CI = 2.1 to 4.1), and hypercalcaemia (OR 11.4, 95% CI = 7.1 to 18).

Rule-out tests

No single blood test was able to rule out myeloma (Table 2), though confidence intervals for each marker approached the pre-specified cut-off of LR 0.2. A normal haemoglobin performed much better as a rule-out test (LR– 0.42, 95% CI = 0.39 to 0.45) compared with calcium (LR– 0.81, 95% CI = 0.78 to 0.83) and creatinine (LR– 0.80, 95% CI = 0.77 to 0.83). For inflammatory markers, the best ones for ruling out the disease were ESR (LR– 0.28, 95% CI = 0.24 to 0.33) and PV (LR– 0.32, 95% CI = 0.24 to 0.44). Ruling out is achieved by most of the test combinations, with the lowest LR– being when haemoglobin, calcium, and PV are all normal (LR– 0.06, 95% CI = 0.02 to 0.19). For two test combinations the lowest LR- was achieved by normal haemoglobin and normal PV (LR- 0.12, 95% CI = 0.07 to 0.23). Details of negative likelihood ratios for when whole FBC is used are available from authors upon request, but results were very similar to using just haemoglobin. A table illustrating how the post-test probability is affected depending on the background prevalence and the estimate of the likelihood ratio is also available from the authors.

Timing of symptoms and investigations

Figures 1 and 2 illustrate the mean values for blood tests for each 3-month period in the 5 years leading up to the date of myeloma diagnosis. In patients with myeloma, but not in control patients, haemoglobin values started to decrease approximately 3 years before diagnosis in both males and females, and the mean haemoglobin value met the WHO criteria for anaemia throughout the final 24 months in males (13 g/dL) and final 15 months for females (12 g/dL).[18] Mean corpuscular volume started to rise approximately 1.5 years before diagnosis, though the increase was minor. WCC and platelets remained stable and within the normal range at all time points. The mean calcium and creatinine values rose rapidly in the last 3–6 months before diagnosis. ESR and PV values started to increase at approximately 2 years before diagnosis. Importantly, CRP showed no difference between cases and controls at any point.

Table 1. Comparison of inflammatory markers in the year before diagnosis

	Blood test results							
		Cases		Controls				
Blood test	Number tested,[a] *n*	Abnormal (%)	Normal (%)	Abnormal (%)	Normal (%)	Odds ratio (95% CI)	*P*-value	LR+ (95% CI)
ESR	2192	844 (85)	154 (15)	546 (46)	648 (54)	5.7 (4.1 to 8.0)	<0.001	1.9 (1.7 to 2.0)
CRP	1274	294 (46)	341 (54)	238 (37)	401 (63)	1.5 (1.0 to 2.2)	0.05	1.2 (1.1 to 1.4)
PV	433	162 (81)	39 (19)	95 (41)	137 (59)	4.4 (2.2 to 8.8)	<0.001	2.0 (1.7 to 2.3)

[a]*Sample size corresponds to the number of patients who had the test or a particular combination of tests. CRP= C-reactive protein. ESR= erythrocyte sedimentation rate. LR+ = positive likelihood ratio. PV= plasma viscosity.*

Table 2. The rule-out value of individual and combinations of blood investigations for myeloma

Investigations	Blood test results: Number tested,[a] *n*	True positive	False positive	False negative	True negative	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR- (95% CI)
Individual tests								
ESR	2192	844	546	154	648	85 (82 to 87)	54 (51 to 57)	0.28 (0.24 to 0.33)
PV	433	162	95	39	137	81 (74 to 86)	59 (52 to 64)	0.32 (0.24 to 0.44)
Hb	5934	1243	912	596	3183	68 (65 to 70)	78 (76 to 79)	0.42 (0.39 to 0.45)
CAL	2809	246	43	893	1627	22 (19 to 24)	98 (97 to 0.98)	0.81 (0.78 to 0.83)[b]
CREAT	7407	648	1021	1202	4536	35 (33 to 37)	82 (81 to 83)	0.80 (0.77 to 0.83)[b]
CRP	1274	316	281	319	358	46 (42 to 50)	63 (59 to 67)	0.87 (0.78 to 0.94)[b]
Combinations (two tests)								
Hb and PV	396	175	119	10	92	95 (90 to 97)	44 (37 to 50)	0.12 (0.07 to 0.23)[c]
Hb and ESR	2071	884	603	72	512	93 (91 to 94)	46 (43 to 49)	0.16 (0.13 to 0.21)[c]
Hb and CRP	1193	491	311	111	280	84 (80 to 86)	47 (43 to 52)	0.39 (0.32 to 0.47)
Combinations (three tests)								
Hb, CREAT, and PV	356	165	111	6	74	97 (93 to 99)	40 (33 to 47)	0.09 (0.04 to 0.20)[c]
Hb, CAL, and PV	213	128	49	3	33	98 (94 to 100)	40 (30 to 52)	0.06 (0.02 to 0.18)[c]
Hb, CREAT, and ESR	1864	823	599	54	388	94 (92 to 95)	39 (36 to 42)	0.16 (0.12 to 0.21)[c]
Hb, CAL, and ESR	1146	590	286	45	225	93 (91 to 95)	44 (40 to 49)	0.16 (0.12 to 0.22)[c]
Hb, CREAT, and CRP	1115	505	341	70	199	88 (85 to 90)	37 (33 to 41)	0.33 (0.26 to 0.42)
Hb, CAL, and CRP	722	363	160	64	135	85 (81 to 88)	46 (40 to 52)	0.33 (0.25 to 0.42)
Combinations (four tests)								
Hb, CAL, CREAT, and PV	211	127	50	3	31	98 (93 to 100)	38 (28 to 50)	0.06 (0.02 to 0.19)[c]
Hb, CAL, CREAT, and ESR	1103	569	309	39	186	94 (91 to 95)	38 (33 to 42)	0.17 (0.12 to 0.24)[c]
Hb, CAL, CREAT, and CRP	706	373	189	44	100	89 (86 to 92)	35 (29 to 40)	0.31 (0.22 to 0.42)

[a]*Sample size corresponds to the number of patients who had the test or a particular combination of tests.* [b]*NRL = negative likelihood ratio (LR-) >0.8.* [c]*NRL ≤0.2. CAL = calcium. CREAT = creatinine. CRP = C-reactive protein. ESR = erythrocyte sedimentation rate. Hb = haemoglobin. LR- = negative likelihood ratio. PV = plasma viscosity.*

Some symptoms, back pain, rib pain, chest infections, chest pain, and nosebleed, become more common in patients with myeloma compared with controls at approximately 2 years before diagnosis (result figures for symptoms are available from the authors). Symptoms suggesting advanced disease, such as fractures, weight loss, and nausea, are more common up to 1 year before diagnosis, especially within the last 3 months.

DISCUSSION

Summary

Myeloma is a difficult disease to identify, especially in its early stages. The combined use of FBC, ESR, or PV and calcium in certain cases on patients currently being tested in primary care appears to be sufficient to select those who may have the disease. Those with test abnormalities can proceed to more specific primary care blood testing, and if these tests are abnormal they warrant haematology expertise. In contrast, if the initial tests are normal, myeloma, which was unlikely in the first place, as it is so rare, becomes extremely unlikely, and another cause should be sought for the patient's symptoms.

This study demonstrates that ESR and PV have better diagnostic performance for myeloma, both as rule-in tests when abnormal, and rule-out tests when normal in contrast to CRP, which has very little value in myeloma diagnosis. Both calcium and creatinine rise before a diagnosis of myeloma, but a negative calcium or creatinine result alone should not be used to rule it out because they only start to rise late in the disease process. Conversely, haemoglobin starts to decline much earlier, subsequently making the FBC a more useful rule-out test. In the year before diagnosis, 95% of patients with myeloma having PV and FBC tests have at least one abnormality (95% CI = 90 to 97%). In practical terms, the combination of normal ESR or PV and normal haemoglobin can be used to rule out myeloma in patients presenting with symptoms potentially associated with myeloma, such as back pain. Non-specific symptoms occur up to 2 years before diagnosis, with more specific symptoms occurring in the last year before diagnosis.

Strengths and limitations

Strengths include the large sample size

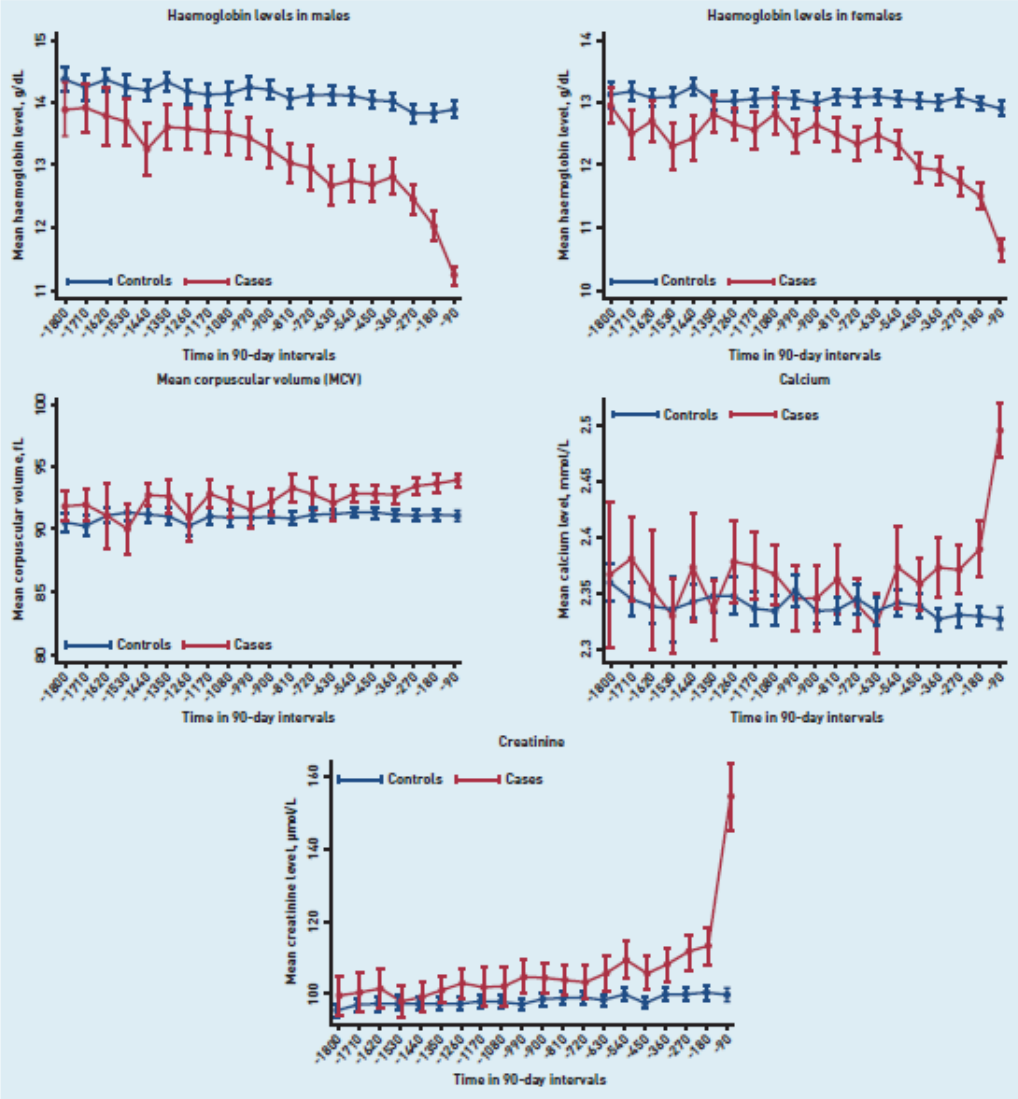


Figure 1. Haemoglobin, mean corpuscular volume, calcium, and creatinine trajectories up to 5 years before diagnosis. Time before diagnosis is split into 90-day intervals and the mean test value for each group is displayed with a 95% confidence interval.

representative of the UK and the primary care setting, both of which increase the generalisability of the results. Limitations include potential overestimation of diagnostic accuracy measures with the case-control design.[19] Another limitation is missing data, as not all patients have been tested. If patients who have been tested differ systematically from those who were not tested with respect to the test (one potential reason might be that sicker patients are more likely to be tested), this may lead to under- or overestimation of the sensitivity, specificity, and likelihood ratios reported here. The authors, therefore, suggest exercising caution in extrapolating their findings to patients not currently being tested in primary care. As with most clinical

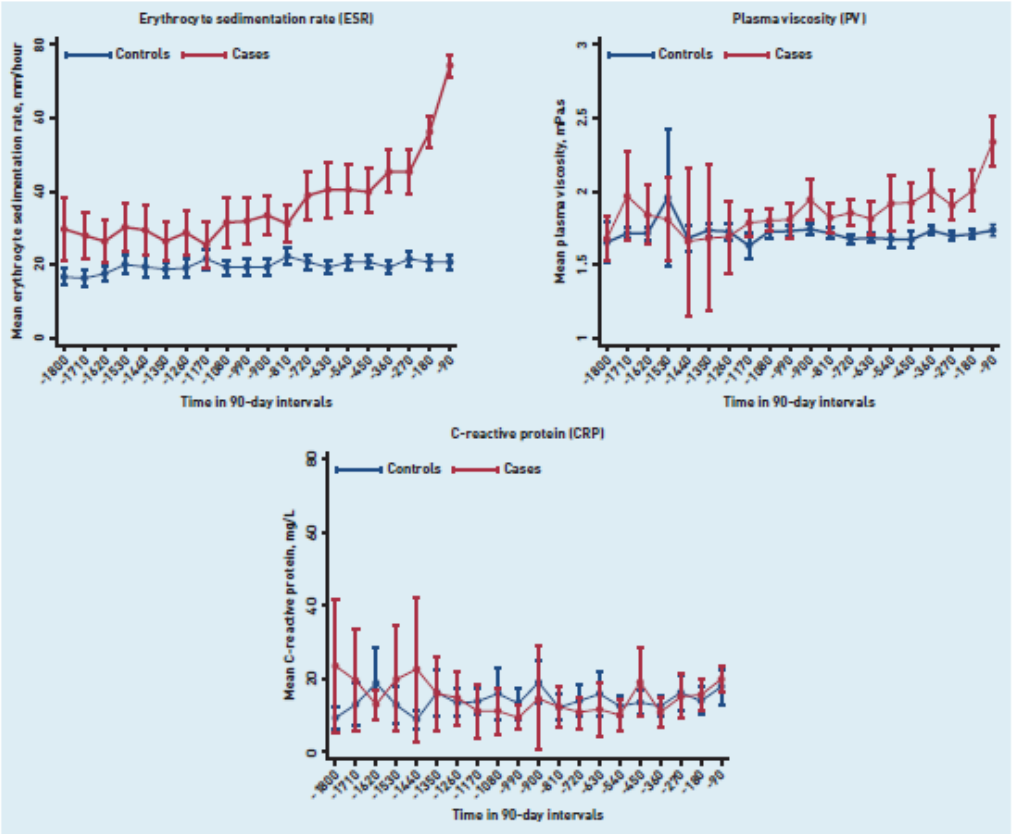


Figure 2. Inflammatory marker trajectories up to 5 years before diagnosis. Time before diagnosis is split into 90-day intervals and the mean test value is displayed for cases and controls with a 95% confidence interval.

databases, there will be some under-reporting or under-recording of symptoms, which can influence association measures.[20]

Comparison with existing literature

A Scandinavian study of 167 blood donors who developed myeloma has previously identified declining haemoglobin for 3 years before diagnosis, similar to the results presented here.[21] No similar reports could be found for the timing of hypercalcaemia or reduced kidney function, though hypercalcaemia is common in patients with advanced cancer stages.[22] Similarly, no reports in which myeloma symptoms were investigated over time could be found.

NICE guidelines currently recommend the use of PV or ESR when myeloma is suspected.[23] Results from this current study strongly support this recommendation. ESR has been reported as a good prognostic marker for myeloma with higher values of ESR associated with a more advanced cancer stage.[24] In contrast, CRP is not useful as an initial investigation if myeloma is suspected: previous reports have concentrated on its use in prognosis rather than diagnosis.[25,26]

Implications for practice

Myeloma diagnosis could potentially be expedited by the use of simple blood tests in patients with back pain, rib pain, chest pain, and recurrent chest infections, which have been identified as early presentations of myeloma. With a sensitivity value ≥93% of normal haemoglobin and an inflammatory marker (PV or ESR), the possibility of myeloma is extremely low. Adding CRP or creatinine adds little additional sensitivity. Conversely, in a patient reporting relevant symptoms, with abnormalities in one of these two tests, it is reasonable for the clinician

to request protein electrophoresis and urinary Bence Jones protein tests urgently. This should expedite myeloma diagnosis, without increasing haematology referrals. Hypercalcaemia is a high-risk marker of possible myeloma. Normal calcium levels do not make myeloma sufficiently less likely for calcium to be used as a safe 'rule-out' test alone. This represents the fact that hypercalcaemia and reduced renal function are late features of the disease. The researchers recommend the addition of calcium tests in the above combination in the presence of certain symptoms like back pain, rib pain, joint pain, and fracture, which have a relatively high rule-in potential (PPV >10%) when combined with abnormal calcium.[9]

Other symptoms appear to represent more advanced disease, and thus are only reported late in its course. Nausea is one of these, possibly arising from hypercalcaemia or kidney failure; similarly, breathlessness may relate to severe anaemia; and weight loss may suggest advanced multifocal disease.

It would be possible to integrate a system within the electronic health record that identifies relevant symptoms or changes in blood parameters over time to alert the clinician to the (small) possibility of myeloma.[27] Patients identified in this way could be tested with PV or ESR, FBC, and calcium in some cases; if these tests were normal, myeloma could be ruled out.

Funding

This manuscript presents work carried out as part of a DPhil scholarship awarded to Constantinos Koshiaris funded by the Primary Care Research Trust, the University of Oxford, and National Institute for Health Research (NIHR) Oxford CLAHRC. This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Ethical approval

This study was approved by the Independent Scientific Advisory Committee: protocol 09-110.

Provenance

Freely submitted; externally peer reviewed.

Open access

This article is Open Access: CC BY-NC 4.0 licence (http://creativecommons.org/licences/by-nc/4.0/).

Competing interests

The authors have declared no competing interests.

Discuss this article

Contribute and read comments about this article: **bjgp.org/letters**

REFERENCES

1. Lyratzopoulos G, Neal RD, Barbiere JM, *et al.* Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012; **13(4):** 353–365.
2. Lyratzopoulos G, Wardle J, Rubin G. Rethinking diagnostic delay in cancer: how difficult is the diagnosis? *BMJ* 2014; **349:** g7400.
3. Friese CR, Abel GA, Magazu LS, *et al.* Diagnostic delay and complications for older adults with multiple myeloma. *Leuk Lymphoma* 2009; **50(3):** 392–400.
4. Kariyawasan CC, Hughes DA, Jayatillake MM, Mehta AB. Multiple myeloma: causes and consequences of delay in diagnosis. *QJM* 2007; **100(10):** 635–640.
5. Barratt J, Hamilton W. Pathways to the diagnosis of lung cancer in the UK: a cohort study. *BMC Fam Pract* 2008; **9:** 31.
6. Elliss-Brookes L, McPhail S, Ives A, *et al.* Routes to diagnosis for cancer — determining the patient journey using multiple routine data sets. *Br J Cancer* 2012; **107(8):** 1220–1226.
7. Howell D, Smith A, Appleton S, *et al.* Multiple myeloma: routes to diagnosis, clinical characteristics and survival — findings from a UK population-based study. *Br J Haematol* 2017; **177(1):** 67–71.
8. Abel GA, Mendonca SC, McPhail S, *et al.* Emergency diagnosis of cancer and previous general practice consultations: insights from linked patient survey data. *Br J Gen Pract* 2017; DOI: https://doi.org/10.3399/bjgp17X690869.
9. Shephard EA, Neal RD, Rose P, *et al.* Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. *Br J Gen Pract* 2015; DOI: https://doi.org/10.3399/bjgp15X680545.
10. National Institute for Health and Care Excellence. *Suspected cancer: recognition and referral. NG12.* London: NICE, 2015. https://www.nice.org.uk/guidance/ng12 (accessed 27 Jul 2018).
11. Cals JW, Chappin FH, Hopstaken RM, *et al.* C-reactive protein point-of-care testing for lower respiratory tract infections: a qualitative evaluation of experiences by GPs. *Fam Pract* 2010; **27(2):** 212–218.
12. Toriola AT, Grankvist K, Agbonsangaya CB, *et al.* Changes in pre-diagnostic serum C-reactive protein concentrations and ovarian cancer risk: a longitudinal study. *Ann Oncol* 2011; **22(8):** 1916–1921.
13. Watson J, Round A, Hamilton W. Raised inflammatory markers. *BMJ* 2012; **344:** e454.
14. Iyen-Omofoman B, Tata LJ, Baldwin DR, *et al.* Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax* 2013; **68(5):** 451–459.
15. Lim AW, Mosher D, Gentry-Maharaj A, *et al.* Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst* 2012; **104(2):** 114–124.
16. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271(9):** 703–707.
17. Thompson M, Van den Bruel A, Verbakel J, *et al.* Systematic review and validation of prediction rules for identifying children with serious infections in emergency departments and urgent-access primary care. *Health Technol Assess* 2012; **16(15):** 1–100.
18. World Health Organization. *Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity.* Geneva: WHO, 2011. http://www.who.int/vmnis/indicators/haemoglobin/en/ (accessed 25 Jul 2018).
19. Rutjes AW, Reitsma JB, Di Nisio M, *et al.* Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; **174(4):** 469–476.
20. Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44(3):** 827–836.
21. Edgren G, Bagnardi V, Bellocco R, *et al.* Pattern of declining hemoglobin concentration before cancer diagnosis. *Int J Cancer* 2010; **127(6):** 1429–1436.
22. Mirrakhimov AE. Hypercalcemia of malignancy: an update on pathogenesis and management. *N Am J Med Sci* 2015; **7(11):** 483–493.
23. National Institute for Health and Care Excellence. *Clinical Knowledge Summaries: multiple myeloma.* Scenario: suspected multiple myeloma. May 2016. https://cks.nice.org.uk/multiple-myeloma#!scenario (accessed 25 Jul 2018).
24. Alexandrakis MG, Passam FH, Ganotakis ES, *et al.* The clinical and prognostic significance of erythrocyte sedimentation rate (ESR), serum interleukin-6 (IL-6) and acute phase protein levels in multiple myeloma. *Clin Lab Haematol* 2003; **25(1):** 41–46.
25. Brown RD, Snowdon L, Uhr E, *et al.* C-reactive protein (CRP) levels do not reflect disease status in patients with multiple myeloma. *Leuk Lymphoma* 1993; **9(6):** 509–512.
26. Hsu DC, Wilkenfeld P, Joshua DE. Multiple myeloma. *BMJ* 2012; **344:** d7953.
27. Kidney E, Berkman L, Macherianakis A, *et al.* Preliminary results of a feasibility study of the use of information technology for identification of suspected colorectal cancer in primary care: the CREDIBLE study. *Br J Cancer* 2015; **112(Suppl 1):** S70–S76.

APPENDIX 3 – GPUpdate (myeloma)

There is so much we don't know in medicine that could make a difference, and often we focus on the big things, and the little things get forgotten. To highlight some smaller but important issues, we've put together a series of pearls that the Red Whale found at the bottom of the ocean of knowledge!

Red Whale GP-Update

MACMILLAN CANCER SUPPORT

Myeloma

Mr Coombes was 68, a keen allotment gardener. He came to see me several times with low back pain and a lack of enthusiasm for digging the potatoes! Initially, it seemed like a musculoskeletal problem, but closer examination revealed tenderness over L5 – he wasn't one to complain and this raised alarm bells! Blood tests showed ESR 94, CRP 10, Hb 11.1 and eGFR 46 (no previous results to compare). I wondered about his prostate but PSA was normal, and his calcium came back raised – protein electrophoresis confirmed a single IgG band. He was subsequently diagnosed with multiple myeloma.

Myeloma is relatively rare and early diagnosis improves survival.

The bottom line is that individual symptoms of myeloma in primary care are generally low risk, and that is probably why there is diagnostic delay. We need a high degree of suspicion **and** take appropriate blood tests. This increases the probability of making the diagnosis.

What is myeloma?

Myeloma is a cancer of plasma cells – white blood cells that produce immunoglobulins. When large numbers of abnormal plasma cells are produced, they overtake the bone marrow and can affect the production of all other cell lines, potentially resulting in anaemia, leucopenia and thrombocytopenia.

The abnormal plasma cells may produce large amounts of abnormal immunoglobulin which is functionally impaired, and also inhibits production of normal immunoglobulin – therefore, the risk of infection may also be increased. This abnormal immunoglobulin is known as paraprotein.

Why is myeloma easily missed?

A BMJ Easily Missed article reminds us that of all of the cancers, myeloma has:

- The longest interval between first presentation of symptoms and diagnosis, with a third taking more than 12 months to diagnose.
- The greatest number of primary care consultations before diagnosis.

This is because it has a variable presentation, and symptoms can be attributed to many other more common conditions (BMJ 2012;344:d79530).

A large proportion are still diagnosed as emergencies, and 56% of these emergency presentations had had at least 3 primary care consultations before the emergency (BJGP 2018;DOI:https://doi.org/10.3399/bjgp18X698357).

Presentation

If the presentation is vague, how do we maximise our chances of making a diagnosis? This was the subject of a BMJ Easily Missed article which reminded us that (BMJ 2012;344:d79530):

- Incidence increases with age, with more than 70% occurring in over-65-year-olds.
- The most common presenting symptoms are:
  - Bone pain (especially back and ribs).
  - Fatigue.
  - Breathlessness.
  - Weight loss.

The problem is that **even in combination,** these symptoms have a PPV <1%.

Combining these symptoms with blood results can be helpful.

A BJGP case–control study used the GPRD to identify and quantify the risk of myeloma from specific clinical features as reported in primary care (BJGP 2015; 65(631):e101):

Combined symptoms and test results	PPV	Combined symptoms and test results	PPV
Hypercalcemia and		**Leucopenia and**	
Back pain (2nd episode)	>10	Nosebleeds	>10
Fracture	>10	Fractures	>10
Joint pain	>10	Combined bone pain	>5
Rib pain	>10	Back pain (2nd episode)	2.0
Back pain (1st episode)	4.0		
Chest infection	2.0	**Raised inflammatory markers and**	
Chest pain	1.9	Back pain (2nd episode)	1.1
SOB	1.5		
Combined bone pain	1.4		

Cancer_Summer 2017

The authors suggest that we investigate any patient aged >60y presenting with bone pain, weight loss or nosebleeds by taking an FBC and ESR. Back pain should be investigated on second presentation. This approach is supported in the new NICE referral guidelines.

So, we need to think about it and do the tests. But which tests?

Which blood tests?

The key thing to remember here is that the vague nature of the presenting symptoms means that we are likely to be considering a number of diagnoses.

The same authors have now extended the above study to look at which blood tests are most helpful in primary care (BJGP 2018;DOI:https://doi.org/10.3399/bjgp18X698357). Using the CPRD to do a case–control study in UK primary care, they found that:

- ESR and plasma viscosity are better for increasing the probability of an accurate myeloma diagnosis than CRP. If they are raised in a symptomatic patient, they make a diagnosis of myeloma more likely and we should do a calcium and serum protein electrophoresis.
- Haemoglobin is the most sensitive part of the FBC, and often started to fall 2–3y prior to diagnosis.
- Calcium and creatinine often do start to rise but a little later in the disease – if they are normal, this does not rule out myeloma.
- CRP is unhelpful – there was no difference between cases and controls at any point.
- No single blood test can rule out myeloma.
- Normal ESR or plasma viscosity AND normal full blood count perform sufficiently well to be used as a simple rule-out screen in primary care.

First-line blood tests

Pragmatically, if we are considering myeloma in the differential, we should request:

Blood test	What are we looking for?
FBC	Anaemia, leucopenia
ESR and/or plasma viscosity	Raised levels increase the likelihood of myeloma
CRP	May be normal in myeloma – if raised, consider other possibilities as well
Calcium	Often raised but less sensitive/specific than ESR/plasma viscosity and Hb
U&E/LFT	End organ damage
PSA	May be appropriate in male patients presenting with bone pain

Second-line blood tests

- Serum protein electrophoresis and urinary Bence Jones proteins should be requested if ESR/plasma viscosity, FBC or Ca are abnormal, or if creatinine is rising without explanation.

If monoclonal proteins are present, refer to haematology under 2ww – they will then do bone marrow biopsy, skeletal survey and 24h urine collection before formulating a treatment plan. They are looking for evidence of end organ damage – hypercalcemia from bone destruction, renal impairment, anaemia and lytic lesions in bone.

NICE guidance on myeloma

Referral should be based on the NICE suspected cancer guidelines (NG12 2015).

NICE Suspected cancer guideline: myeloma
Offer FBC, serum calcium and plasma viscosity or ESR to assess for myeloma in people aged ≥60y with persistent bone pain, particularly back pain or unexplained fracture. *N.B. CRP is often normal.* Offer protein electrophoresis and Bence Jones protein urine test within 48h to assess for myeloma in: • People aged ≥60y with hypercalcaemia or leucopenia, and presentation consistent with possible myeloma. • People with raised plasma viscosity or ESR, and presentation consistent with possible myeloma. Refer using cancer pathway if protein electrophoresis suggests myeloma.

Management

This is a secondary care issue and was covered recently by a new NICE guideline (NG35 2016) – simplistically, patients are divided into those who are eligible for autologous stem cell transplantation and those who are not.

Different chemotherapy regimens are then the mainstay of treatment, either in preparation for stem cell transplant (with the aim of cure), or to achieve a clinical remission where there is no evidence of end organ damage and no detectable paraprotein.

All patients with myeloma are offered bisphosphonates to reduce the risk of pathological fractures and skeletal damage.

All myeloma patients should be offered annual flu immunisation and be considered for pneumococcal vaccination, regardless of age.

Smouldering myeloma

There is a group of patients with 'smouldering myeloma' (presence of significant paraprotein but no evidence of end organ damage) who may be offered active monitoring – we may be involved in this in primary care.

NICE (CG35 2016) recommends 3-monthly monitoring for the first 5y after diagnosis, and then an individualised plan based on individual risk factors for progression.

Monitoring should include:

- Assessment of symptoms.
- FBC, renal function, bone profile, serum immunoglobulins and electrophoresis.
- Serum-free light chains (if appropriate).

Routine skeletal surveys are not recommended; instead, imaging should be requested according to new bone symptoms.

Complications of treatment

NICE (CG35 2016) highlights some of the common complications of treatment for myeloma. This is where we might be called by patients, and it is useful to have a simple overview:

- Peripheral neuropathy can be a side-effect of chemotherapy treatment.
- Thrombosis – patients may be offered thromboprophylaxis. If a thrombotic event is suspected, refer as normal.
- Fatigue – check for anaemia. Haematology may consider erythropoietin infusions to maintain haemoglobin between 110–120g/L.

Monoclonal gammopathy of undetermined significance (MGUS)

Increasing numbers of patients are being picked up while asymptomatic (perhaps through blood tests for other conditions). If their bone marrow biopsy has fewer than 10% plasma cells, or if their monoclonal proteinuria is less than 30g/L, they are given a diagnosis of MGUS.

An estimated 3% of >50-year-olds may have MGUS, with an estimated 1% progressing to multiple myeloma or other lymphoproliferative disease each year.

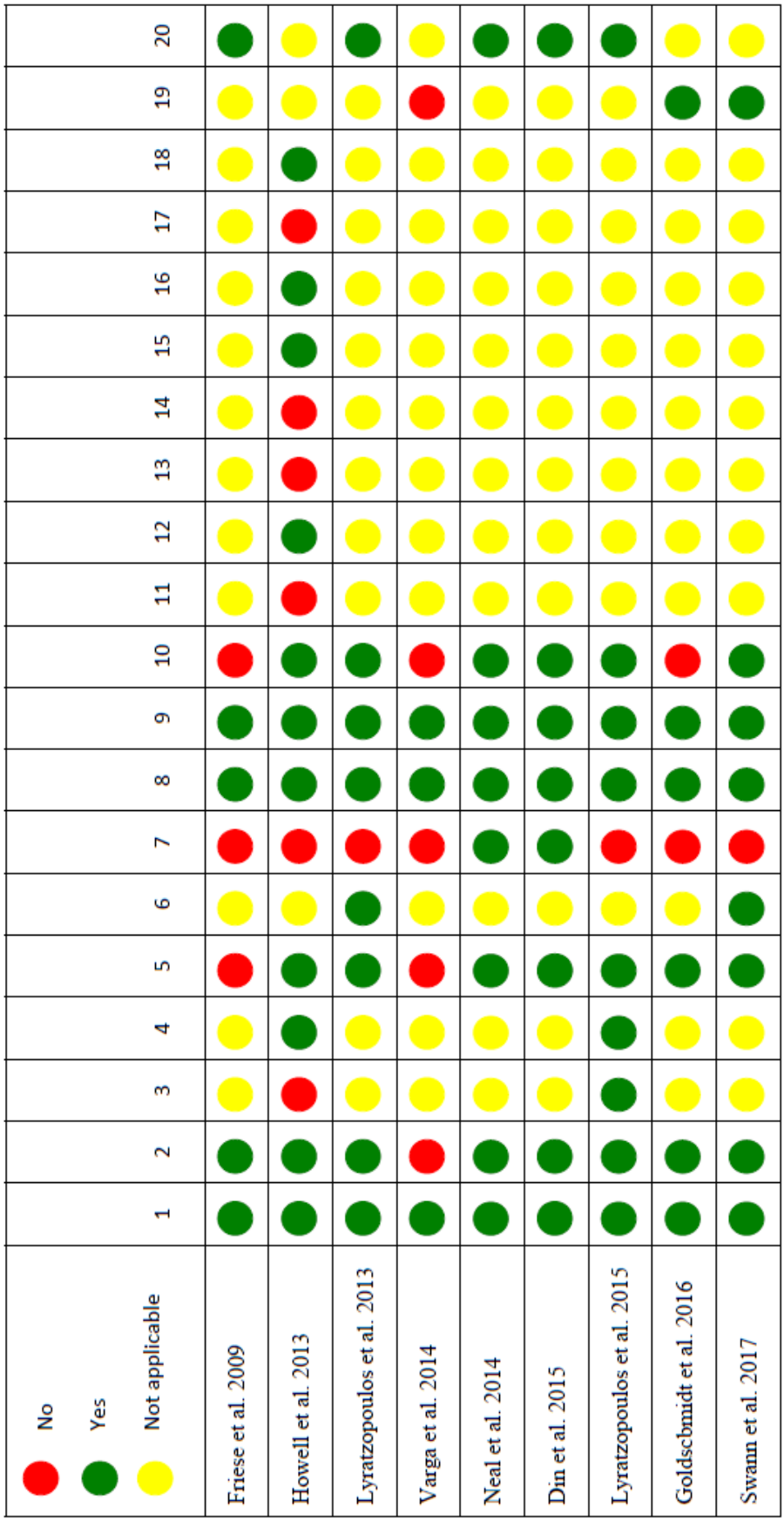
Active surveillance by blood test is the current recommended management of MGUS.

Take home messages	Multiple myeloma • Rare and commonly missed. • Bone pain, fatigue, breathlessness and weight loss are typical symptoms. • No individual or grouped symptoms have strong PPVs of >1%! We need to do blood tests. • Blood tests will characteristically identify the condition, particularly FBC, ESR and/or plasma viscosity. Calcium and creatinine rise later. Serum protein electrophoresis is second line if these are abnormal. • Offer an FBC and ESR to patients aged >60y presenting with bone pain, nosebleeds, weight loss or a second visit with back pain. • Look for leucopenia and hypercalcemia, and follow these up! • MGUS is increasingly being identified – it has a 1% per annum conversion to myeloma.
Other resources	Information for clinicians: If you are sharing care for a patient having chemotherapy for myeloma, the UKONS tool might be helpful: https://www.macmillan.org.uk/_images/oncology-treatment-toxicity-risk-assesment-tool_tcm9-317392.pdf Support for patients: Macmillan offers lots of resources to support patients with a diagnosis of myeloma: Understanding myeloma – free information booklet: https://be.macmillan.org.uk/be/p-254-understanding-myeloma.aspx Information relating specifically to myeloma: https://www.macmillan.org.uk/information-and-support/myeloma

APPENDIX 4 – Search strategy for systematic review

1. Multiple Myeloma/
2. Myeloma*.ti,ab.
3. 1 or 2
4. (time adj4 diagnos$).ti,ab.
5. (time adj4 consult$).ti,ab.
6. (time adj4 refer$).ti,ab.
7. (time adj4 present$).ti,ab.
8. 4 or 5 or 6 or 7
9. (delay$ adj4 diagnos$).ti,ab.
10. (delay$ adj4 consult$).ti,ab.
11. (delay$ adj4 refer$).ti,ab.
12. (delay$ adj4 present$).ti,ab
13. (delay$adj4 seek$).ti,ab
14. (delay$ adj4 detect*).ti,ab.
15. 9 or 10 or 11 or 12 or 13 or 14
16. (interval adj4 consult$).ti,ab.
17. (interval adj4 refer$).ti,ab.
18. (interval adj4 present$).ti,ab.
19. 16 or 17 or 18
20. (late adj4 diagnosis).ti,ab.
21. (late adj4 detect*).ti,ab.
22. (late adj4 present$).ti,ab.
23. 20 or 21 or 22
24. diagnos$ delay$.ti,ab.
25. early diagnos$.ti,ab.

APPENDIX 5 –Risk of bias summary

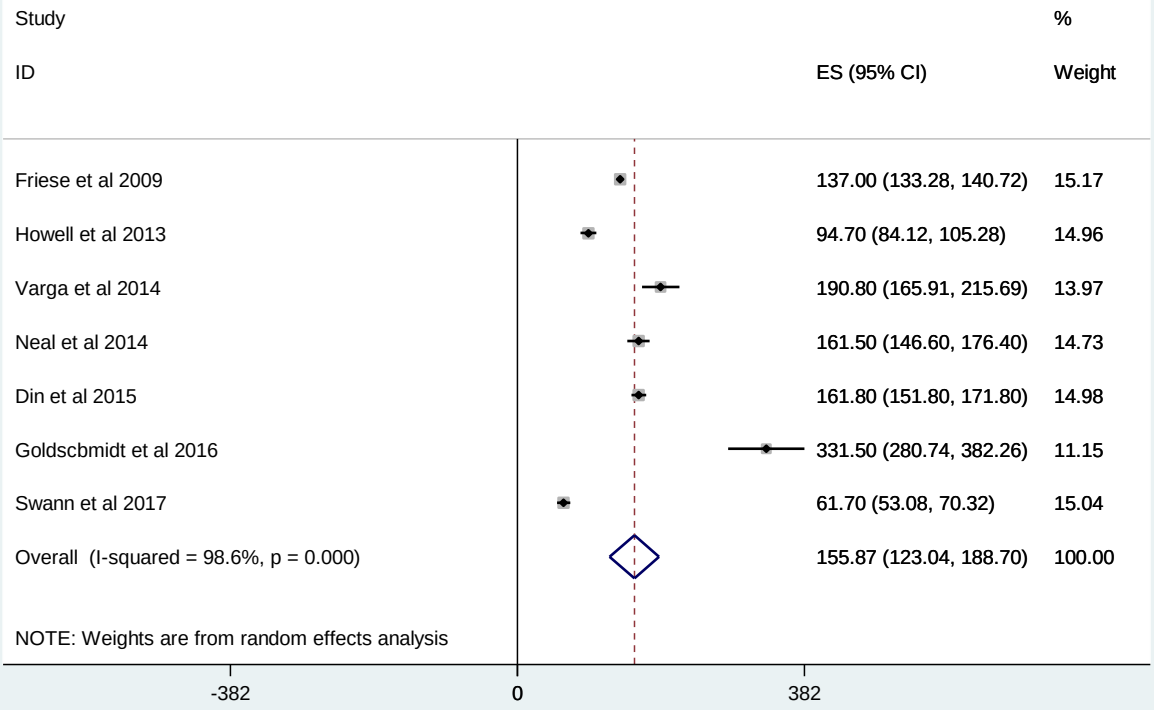


APPENDIX 6 – Risk of bias summary

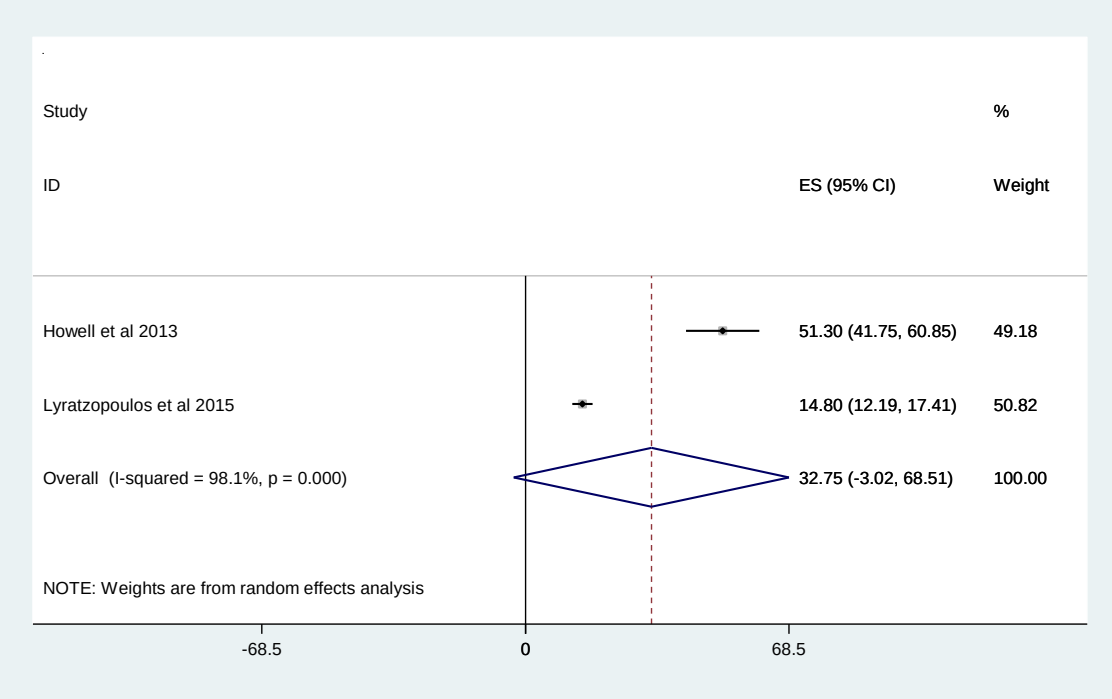


APPENDIX 7 – Sensitivity analysis using random effects meta-analysis

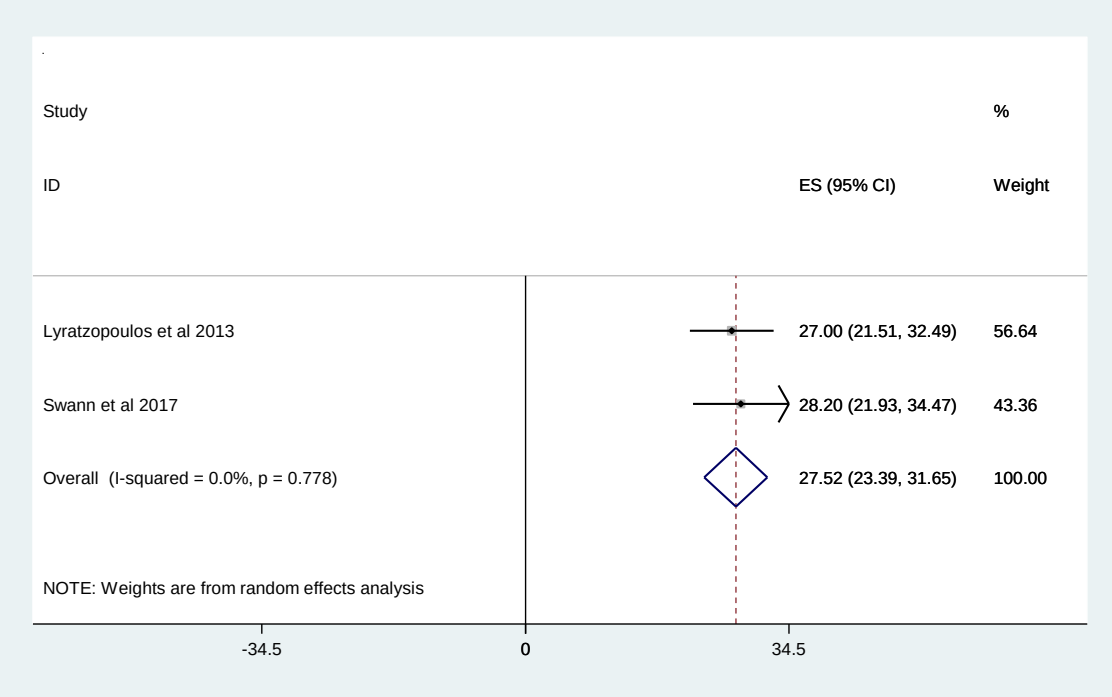
Diagnostic interval



Patient interval



Primary care interval



APPENDIX 8 –Read codes

Diagnostic codes for myeloma

Read code	Read term
B63z.00	IMMUNOPROLIFERATIVE NEOPLASM OR MYELOMA NOS
B630.12	MYELOMATOSIS
B630.00	MULTIPLE MYELOMA
B630300	LAMBDA LIGHT CHAIN MYELOMA
B631.00	PLASMA CELL LEUKEMIA
B630100	SOLITARY MYELOMA
B630200	PLASMACYTOMA NOS
BBn0.12	[M] MYELOMA NOS
BBn0.00	PLASMA CELL MYELOMA
BBn0.11	[M] MULTIPLE MYELOMA
B936.11	MYELOMA – SOLITARY
B63..00	MULTIPLE MYELOMA AND IMMUNOPROLIFERATIVE NEOPLASMS

Laboratory investigations codes

Full blood count

Read code	Read term
4243.00	FULL BLOOD COUNT ABNORMAL
424Z.00	FULL BLOOD COUNT NOS
424..00	FULL BLOOD COUNT - FBC
4241.00	FULL BLOOD COUNT NORMAL
4242.00	FULL BLOOD COUNT BORDERLINE
423..00	HAEMOGLOBIN ESTIMATION
4237	HAEMOGLOBIN NORMAL
4235	HAEMOGLOBIN LOW
423Z.00	HAEMOGLOBIN ESTIMATION NOS
4239	HAEMOGLOBIN HIGH
4236	HAEMOGLOBIN BORDERLINE LOW
4234	HAEMOGLOBIN VERY LOW
423B.00	HAEMOGLOBIN ABNORMAL
4238	HAEMOGLOBIN BORDERLINE HIGH
423A.00	HAEMOGLOBIN VERY HIGH
423..11	HB ESTIMATION
4258	HAEMATOCRIT
4255	HAEMATOCRIT - BORDERLINE LOW
4253	HAEMATOCRIT - PCV - HIGH
4251	HAEMATOCRIT - PCV - NORMAL
425..00	HAEMATOCRIT - PCV
425Z.00	HAEMATOCRIT - PCV - NOS
4252	HAEMATOCRIT - BORDERLINE HIGH
4256	HAEMATOCRIT - PCV - ABNORMAL
4254	HAEMATOCRIT - PCV - LOW

42H..00	TOTAL WHITE CELL COUNT
42H..11	WHITE BLOOD COUNT
42H..12	WHITE CELL COUNT
42H7.00	TOTAL WHITE BLOOD COUNT
42H1.00	WHITE CELL COUNT NORMAL
42HZ.00	TOTAL WHITE CELL COUNT NOS
42H5.00	WHITE CELL COUNT ABNORMAL
426..00	RED BLOOD CELL (RBC) COUNT
42C..00	RBC - RED BLOOD CELL SIZE
426Z.00	RBC COUNT NOS
4261	RBC COUNT NORMAL
42D..00	RBC - RED BLOOD CELL SHAPE
4263	RBC COUNT LOW
4264	RBC COUNT RAISED
42J..00	NEUTROPHIL COUNT
42b0.00	PERCENTAGE NEUTROPHILS
42JZ.00	NEUTROPHIL COUNT NOS
42J1.00	NEUTROPHIL COUNT NORMAL
42J4.00	NEUTROPHIL COUNT ABNORMAL
42P..00	PLATELET COUNT
42P1.00	PLATELET COUNT NORMAL
42PZ.00	PLATELET COUNT NOS
42P4.00	PLATELET COUNT ABNORMAL
42A..00	MEAN CORPUSCULAR VOLUME (MCV)
42AZ.00	MCV - NOS
42A1.00	MCV - NORMAL
42A3.00	MCV - RAISED
42A2.00	MCV - BORDERLINE RAISED
42A4.00	MCV - LOW
42A5.00	MCV - BORDERLINE LOW

Calcium

Read code	**Read term**
44I8.00	SERUM CALCIUM
44IC.00	CORRECTED SERUM CALCIUM LEVEL
44I8100	RAISED SERUM CALCIUM LEVEL
44I8000	NORMAL SERUM CALCIUM LEVEL
44h4.00	BLOOD CALCIUM LEVEL
44h7.00	PLASMA CALCIUM LEVEL
44h9.00	PLASMA CORRECTED CALCIUM LEVEL
4Q72.00	CALCIUM LEVEL

Creatinine

Read code	**Read term**
44I8.00	Serum calcium
44IC.00	Corrected serum calcium level
44ZR.00	Calcium profile
44IC000	Serum adjusted calcium concentration
44h9.00	Plasma corrected calcium level
44h7.00	Plasma calcium level

4Q72100	Calcium adjusted level
44h4.00	Blood calcium level
4Q72.00	Calcium level
44IC.11	Adjusted serum calcium level
44h9000	Plasma adjusted calcium concentration
44I8000	Normal serum calcium level

Sedimentation rate

Read code	Read term
R101.00	[D]ELEVATED SEDIMENTATION RATE
42B6.00	ERYTHROCYTE SEDIMENTATION RATE
42B6z00	ERYTHROCYTE SEDIMENT RATE NOS
42B6200	ESR NORMAL
42B6300	ESR RAISED
42B6000	ESR ABNORMAL
42B6100	ESR LOW

C-reactive protein

Read code	Read term
44CC000	C REACTIVE PROTEIN NORMAL
44CC100	C REACTIVE PROTEIN ABNORMAL
44CC.00	PLASMA C REACTIVE PROTEIN
44CS.00	SERUM C REACTIVE PROTEIN LEVEL

Plasma viscosity

Read code	Read term
42B..00	PLASMA VISCOSITY
R108.00	[D]ABNORMAL PLASMA VISCOSITY
42B4.00	PLASMA VISCOSITY LOW
42B3.00	PLASMA VISCOSITY RAISED
42B1.00	PLASMA VISCOSITY NORMAL
42B..11	PLASMA VISCOSITY - PV
42BZ.00	PLASMA VISCOSITY NOS

Symptom codes list

Back pain

Read code	Read term
16C2.00	BACKACHE
N142.11	LOW BACK PAIN
N145.00	BACKACHE, UNSPECIFIED
16C..00	BACKACHE SYMPTOM
16C6.00	BACK PAIN WITHOUT RADIATION NOS
16C5.00	C/O - LOW BACK PAIN
N23yE00	SPASM OF BACK MUSCLES

N142.13	ACUTE BACK PAIN - LUMBAR
SK1B.00	INJURY OF MUSCLE + TENDON OF ABDOMEN, LOWER BACK AND PELVIS
N12..13	ACUTE BACK PAIN - DISC
1D24.11	C/O - A BACK SYMPTOM
N143.11	ACUTE BACK PAIN WITH SCIATICA
1A53.12	C/O - LUMBAR PAIN
N141.11	ACUTE BACK PAIN - THORACIC
N145.11	ACUTE BACK PAIN - UNSPECIFIED
16C7.00	C/O - UPPER BACK ACHE
16CZ.00	BACKACHE SYMPTOM NOS
N145.12	BACK PAIN, UNSPECIFIED
16C3.00	BACKACHE WITH RADIATION
8HTH.00	REFERRAL TO BACK PAIN CLINIC
16C8.00	EXACERBATION OF BACKACHE
16C9.00	CHRONIC LOW BACK PAIN
16CA.00	MECHANICAL LOW BACK PAIN
16C4.00	BACK PAIN WORSE ON SNEEZING
	INJ/NERVES+LUMBAR SPINAL CORD/ABDO,LOWER
SJA..00	BACK+PELVIS LEVEL
Eu45412	[X]PSYCHOGENIC BACKACHE
E278200	PSYCHOGENIC BACKACHE

Chest pain

Read code	**Read term**
182..00	CHEST PAIN
R065000	[D]CHEST PAIN, UNSPECIFIED
1828	ATYPICAL CHEST PAIN
1825	PLEURITIC PAIN
R065800	[D]CHEST TIGHTNESS
R065.00	[D]CHEST PAIN
R065A00	[D]MUSCULOSKELETAL CHEST PAIN
R065z00	[D]CHEST PAIN NOS
1822	CENTRAL CHEST PAIN
R065B14	[D]NON-CARDIAC CHEST PAIN
R065600	[D]CHEST DISCOMFORT
182B000	COSTAL MARGIN CHEST PAIN
R065B00	[D]NON CARDIAC CHEST PAIN
1824	ANTERIOR CHEST WALL PAIN
182Z.00	CHEST PAIN NOS
R065200	[D]ANTERIOR CHEST WALL PAIN
R065700	[D]CHEST PRESSURE
182A.00	CHEST PAIN ON EXERTION
R065400	[D]PLEURITIC PAIN
R065900	[D]PARASTERNAL CHEST PAIN
R065C00	[D]RETROSTERNAL CHEST PAIN
1D22000	CHEST WALL TENDERNESS
182C.00	CHEST WALL PAIN
R065011	[D] RETROSTERNAL CHEST PAIN
R065D00	[D]CENTRAL CHEST PAIN
Ryu0400	[X]OTHER CHEST PAIN

Bone pain

Read code	Read term
N245.12	ARM PAIN
N33A.00	BONE PAIN
N142.00	PAIN IN LUMBAR SPINE
N245.18	THIGH PAIN
N141.00	PAIN IN THORACIC SPINE
N245300	PAIN IN ARM
N131.11	PAIN IN CERVICAL SPINE

Rib pain

Read code	Read term
182B.00	RIB PAIN
197..14	SUBCOSTAL PAIN
N241011	INTERCOSTAL MYALGIA
1973	LEFT SUBCOSTAL PAIN
1974	RIGHT SUBCOSTAL PAIN

Joint pain

Read code	Read term
N094611	KNEE JOINT PAIN
N094512	HIP JOINT PAIN
N094211	ELBOW JOINT PAIN
N094111	SHOULDER JOINT PAIN
N094311	WRIST JOINT PAIN
N094.00	PAIN IN JOINT - ARTHRALGIA
N094.11	ACHE IN JOINT
N096.12	MUSCULOSKELETAL PAIN - JOINTS
N094711	ANKLE JOINT PAIN
SP07C00	PAIN DUE TO SHOULDER JOINT PROSTHESIS
N094411	HAND JOINT PAIN
2H45.00	O/E - JOINT MOVEMENT PAINFUL
N094D11	ELBOW JOINT PAIN

Shortness of breath

Read code	Read term
R060D00	[D]BREATHLESSNESS
R060800	[D]SHORTNESS OF BREATH
173..00	BREATHLESSNESS
173C.00	SHORT OF BREATH ON EXERTION
1738	DIFFICULTY BREATHING
R060A00	[D]DYSPNOEA

1739	SHORTNESS OF BREATH
173..11	BREATHLESSNESS SYMPTOM
173..13	SHORTNESS OF BREATH SYMPTOM
173..12	DYSPNOEA - SYMPTOM
1732	BREATHLESS - MODERATE EXERTION
1736	PAROXYSMAL NOCTURNAL DYSPNOEA
2322	O/E - DYSPNOEA
1735	BREATHLESS - LYING FLAT
1733	BREATHLESS - MILD EXERTION
173C.12	SOBOE
173D.00	NOCTURNAL DYSPNOEA
173J.00	MRC BREATHLESSNESS SCALE: GRADE 3
173I.00	MRC BREATHLESSNESS SCALE: GRADE 2
173L.00	MRC BREATHLESSNESS SCALE: GRADE 5
173K.00	MRC BREATHLESSNESS SCALE: GRADE 4
173H.00	MRC BREATHLESSNESS SCALE: GRADE 1
173Z.00	BREATHLESSNESS NOS
173F.00	SHORT OF BREATH DRESSING/UNDRESSING
17ZZ.00	RESPIRATORY SYMPTOM NOS
173G.00	BREATHLESS - STRENUOUS EXERTION
1734	BREATHLESS - AT REST
173C.11	DYSPNOEA ON EXERTION
388H.00	CLASP SHORTNESS OF BREATH SCORE

Chest infections

Read code	Read term
H06z011	CHEST INFECTION
H06z000	CHEST INFECTION NOS
H06z200	RECURRENT CHEST INFECTION

Fatigue

Read code	Read term
R007500	[D]TIREDNESS
168..00	TIREDNESS SYMPTOM
1683	TIRED ALL THE TIME
1682	FATIGUE
E205.12	TIRED ALL THE TIME
168..11	FATIGUE - SYMPTOM
F286.00	CHRONIC FATIGUE SYNDROME
R007100	[D]FATIGUE
F286.11	CFS - CHRONIC FATIGUE SYNDROME
168Z.00	TIREDNESS SYMPTOM NOS
Eu46011	[X]FATIGUE SYNDROME
R007z00	[D]MALAISE AND FATIGUE NOS
R007.00	[D]MALAISE AND FATIGUE
F286100	MODERATE CHRONIC FATIGUE SYNDROME
F286000	MILD CHRONIC FATIGUE SYNDROME
F286200	SEVERE CHRONIC FATIGUE SYNDROME

Bleeding

Read code	Read term
R047.11	[D]NOSEBLEED
R09z000	[D]UMBILICAL BLEEDING
Kyu9D00	[X]OTHER SPECIFIED ABNORMAL UTERINE AND VAGINAL BLEEDING
K5Ez.00	ABNORMAL UTERINE AND VAGINAL BLEEDING, UNSPECIFIED
K5E0.00	ABNORMAL UTERINE BLEEDING UNRELATED TO MENSTRUAL CYCLE
K5E1.00	ABNORMAL UTERINE BLEEDING, UNSPECIFIED
K5E2.00	ABNORMAL VAGINAL BLEEDING, UNSPECIFIED
L11..12	ANTEPARTUM BLEEDING
K56y111	BLEEDING - VAGINAL NOS
J110111	BLEEDING ACUTE GASTRIC ULCER
J121111	BLEEDING CHRONIC DUODENAL ULCER
J111111	BLEEDING CHRONIC GASTRIC ULCER
D30..11	BLEEDING DISORDERS
J510900	BLEEDING DIVERTICULOSIS
SP03216	BLEEDING DUE TO INTRAUTERINE CONTRACEPTIVE DEVICE
K19y400	BLEEDING FROM URETHRA
1928.00	BLEEDING GUMS
G848000	BLEEDING HAEMORRHOIDS NOS
L10y.11	BLEEDING IN EARLY PREGNANCY
L36..11	BLEEDING POSTPARTUM
J573.11	BLEEDING PR
25T0.00	BLEEDING STOMA
42Q..11	BLEEDING TESTS
4762.11	BLOOD IN FAECES
19E6.00	BLOOD IN FAECES
19E6.11	BLOOD IN FAECES SYMPTOM
172..00	BLOOD IN SPUTUM - HAEMOPTYSIS
172..11	BLOOD IN SPUTUM - SYMPTOM
J681.11	BLOOD IN STOOL
J681.13	BLOOD IN STOOLS ALTERED
4B34.11	BLOOD IN SYNOVIAL FLUID
1A45.00	BLOOD IN URINE - HAEMATURIA
1A45.11	BLOOD IN URINE - SYMPTOM
1994.11	BLOOD IN VOMIT - SYMPTOM
4A23.11	BLOOD IN VOMIT O/E
19EG.00	BLOOD ON PANTS
19ED.00	BLOOD ON TOILET PAPER
K56y112	BPV - VAGINAL BLEEDING
K59z.11	BREAK - THROUGH BLEEDING
158..11	C/O P.V. BLEEDING
D3...00	CLOTTING AND BLEEDING DISORDERS
D3z..00	CLOTTING OR BLEEDING DISORDER NOS
42QZ.00	COAG./BLEEDING TEST NOS
42Q2.00	COAG./BLEEDING TESTS ABNORMAL
K587.00	CONTACT BLEEDING OF CERVIX
SP03217	CONTRACEPTION IUCD CAUSING BLEEDING
K59yx11	DYSFUNCTIONAL UTERINE BLEEDING

G845000	EXTERNAL BLEEDING HAEMORRHOIDS
J68z.11	GIB - GASTROINTESTINAL BLEEDING
158..00	H/O: ABNORMAL UTERINE BLEEDING
1455.11	H/O: BLEEDING DISORDER
14CA.11	H/O: GI BLEED
1582.00	H/O: INTER-MENSTRUAL BLEEDING
1581.00	H/O: POST-COITAL BLEEDING
1583.00	H/O: POST-MENOPAUSAL BLEEDING
14CD.11	H/O: UPPER GIT BLEED
158Z.00	H/O:ABNORMAL UTERINE BLEED NOS
1C62.00	HAS NOSE BLEEDS - EPISTAXIS
1584.00	HEAVY EPISODE OF VAGINAL BLEEDING
66UI.00	HORMONE REPLACEMENT THERAPY BLEED PATTERN - ABNORMAL
66UK.00	HORMONE REPLACEMENT THERAPY BLEED PATTERN - NO BLEEDING
K59y300	INTERMENSTRUAL BLEEDING
K596.11	INTERMENSTRUAL BLEEDING - IRREGULAR
K595.11	INTERMENSTRUAL BLEEDING - REGULAR
G842000	INTERNAL BLEEDING HAEMORRHOIDS
K599.00	MID-CYCLE BLEEDING
1C6..00	NOSE BLEED SYMPTOM
1C6Z.00	NOSE BLEED SYMPTOM NOS
2556.00	O/E - BLEEDING GUMS
2693.00	O/E-SPECULUM=BLEEDING THRO' OS
G850.00	OESOPHAGEAL VARICES WITH BLEEDING
G852000	OESOPHAGEAL VARICES WITH BLEEDING IN DISEASES EC
G851.00	OESOPHAGEAL VARICES WITHOUT BLEEDING
K5E..00	OTHER ABNORMAL UTERINE AND VAGINAL BLEEDING
196B.00	PAINFUL RECTAL BLEEDING
196C.00	PAINLESS RECTAL BLEEDING
Q41y111	PERINATAL TRANSIENT VAGINAL BLEEDING
K597.00	POSTCOITAL BLEEDING
K5A1.00	POSTMENOPAUSAL BLEEDING
K59B.00	POSTMENOPAUSAL POSTCOITAL BLEEDING
J573012	PRB - RECTAL BLEEDING
K59A.00	PREMENOPAUSAL POSTCOITAL BLEEDING
K593.11	PUBERTAL BLEEDING AND MENORRHAGIA
K593.00	PUBERTY BLEEDING
J573011	RECTAL BLEEDING
K19y411	URETHRAL BLEEDING
158..12	VAGINAL BLEEDING

Bruising

Read code	Read term
SE...00	CONTUSION (BRUISE) WITH INTACT SKIN
16B3.00	SPONTANEOUS BRUISING
SE4..11	LEG BRUISE
16B..00	BRUISING SYMPTOM
SE3..11	ARM BRUISE
R027.11	[D]SPONTANEOUS BRUISING
2I15.00	O/E - BRUISING
SE0..12	BRUISE OF HEAD
SE43.11	TOENAIL BRUISE

SE2..11	BRUISE, TRUNK
SE1..11	BRUISE OF EYE
SE30011	SHOULDER BRUISE
16B2.00	BRUISES EASILY
SE0..11	BRUISE OF FACE
16B4.00	POST-TRAUMATIC BRUISING
16BZ.00	BRUISING SYMPTOM NOS
SC13.11	LATE EFFECT OF BRUISING
SE42011	HEEL BRUISE
SE06.00	BRUISE OF MANDIBULAR JOINT AREA
SE24211	BRUISE OF SCROTUM
SE05.11	BRUISE OF EAR
SE11.11	BRUISE OF EYELIDS
SE11.12	BRUISE OF PERIOCULAR TISSUE
SE05.12	BRUISE OF AURICLE
L336.11	BRUISING OF CORD
Q416100	PERINATAL CUTANEOUS BRUISING

Fracture

Read code	Read term
S23x111	FRACTURE OF RADIUS NOS
S35..11	METATARSAL BONE FRACTURE
S23B.00	FRACTURE OF LOWER END OF RADIUS
S234.11	WRIST FRACTURE - CLOSED
S32..00	FRACTURE OF PATELLA
S120.00	CLOSED FRACTURE RIB
S34..00	FRACTURE OF ANKLE
S234100	CLOSED COLLES' FRACTURE
S3z..11	FRACTURE NOS
S020.00	CLOSED FRACTURE NOSE
S26..00	FRACTURE OF ONE OR MORE PHALANGES OF HAND
S3z0000	GREENSTICK FRACTURE
S26..12	THUMB FRACTURE EXCLUDING BASE
S20..00	FRACTURE OF CLAVICLE
S22..00	FRACTURE OF HUMERUS
S31z.00	FRACTURE OF FEMUR, NOS
S242000	FRACTURE OF SCAPHOID
S0...00	FRACTURE OF SKULL
S13..00	FRACTURE OR DISRUPTION OF PELVIS
S339.00	FRACTURE OF FIBULA ALONE
S10B200	FRACTURE OF COCCYX
S35..00	FRACTURE OF ONE OR MORE TARSAL AND METATARSAL BONES
S4...13	FRACTURE DISLOCATIONS AND FRACTURE SUBLUXATIONS
S23z.00	FRACTURE OF RADIUS AND ULNA, NOS
S3z1.00	OPEN FRACTURE OF BONES, UNSPECIFIED
S33x000	CLOSED FRACTURE OF TIBIA, UNSPECIFIED PART, NOS
S242200	FRACTURE OF OTHER METACARPAL BONE
S23x211	FRACTURE OF ULNA NOS
S21..00	FRACTURE OF SCAPULA
S28..11	ILL-DEFINED FRACTURE OF ARM
S224.11	ELBOW FRACTURE - CLOSED
S228.00	FRACTURE OF LOWER END OF HUMERUS

S130.00	CLOSED FRACTURE ACETABULUM
S352.11	MARCH FRACTURE
S234200	CLOSED FRACTURE OF THE DISTAL RADIUS, UNSPECIFIED
S354.00	FRACTURE OF CALCANEUS
S36..11	TOE FRACTURE
S30..11	HIP FRACTURE
S226.00	FRACTURE OF UPPER END OF HUMERUS
S362.00	FRACTURE OF GREAT TOE
S30..00	FRACTURE OF NECK OF FEMUR
S344.12	POTT'S FRACTURE - ANKLE
S024100	CLOSED FRACTURE ZYGOMA
S237.00	FRACTURE OF UPPER END OF RADIUS
S10B500	FRACTURE OF PUBIS
S355.00	FRACTURE OF TALUS
S01..00	FRACTURE OF BASE OF SKULL
S3z..00	FRACTURE OF UNSPECIFIED BONES
S3...11	LEG FRACTURE
S33..00	FRACTURE OF TIBIA AND FIBULA
S022.00	FRACTURE OF MANDIBLE, CLOSED
S25..11	HAND FRACTURE - METACARPAL BONE
S28z.00	ILL-DEFINED FRACTURES OF UPPER LIMB NOS
S230300	CLOSED MONTEGGIA'S FRACTURE
S36..00	FRACTURE OF ONE OR MORE PHALANGES OF FOOT
S35..12	TARSAL BONE FRACTURE
S234700	CLOSED SMITH'S FRACTURE
S25..00	FRACTURE OF METACARPAL BONE
S3z2.00	STRESS FRACTURE
S10A.00	FRACTURE OF NECK
S021.00	OPEN FRACTURE NOSE
S10x.00	CLOSED FRACTURE OF SPINE, UNSPECIFIED,
S10B100	FRACTURE OF SACRUM
S233.00	OPEN FRACTURE OF RADIUS AND ULNA, SHAFT
S00..11	FRONTAL BONE FRACTURE
S104.00	CLOSED FRACTURE LUMBAR VERTEBRA
S352700	CLOSED FRACTURE METATARSAL
S122.00	CLOSED FRACTURE STERNUM
S210300	CLOSED FRACTURE SCAPULA, GLENOID
S20..11	COLLAR BONE FRACTURE
S024000	CLOSED FRACTURE MAXILLA
S33x100	CLOSED FRACTURE OF FIBULA, UNSPECIFIED PART, NOS
S360.00	CLOSED FRACTURE OF ONE OR MORE PHALANGES OF FOOT
S352300	CLOSED FRACTURE CUBOID
S23x300	CLOSED FRACTURE OF THE RADIUS AND ULNA
S10..12	FRACTURE OF VERTEBRA WITHOUT SPINAL CORD LESION
S33x200	CLOSED FRACTURE OF TIBIA AND FIBULA, UNSPECIFIED PART
S26z.00	FRACTURE OF ONE OR MORE PHALANGES OF HAND NOS
S242300	MULTIPLE FRACTURES OF METACARPAL BONES
S34x.00	CLOSED FRACTURE ANKLE, UNSPECIFIED
N338000	MALUNION OF FRACTURE
S02x100	FRACTURE OF ORBIT NOS, CLOSED
S348.00	FRACTURE OF MEDIAL MALLEOLUS
S26..11	FINGER FRACTURE
S028000	FRACTURE OF NASAL BONES
S302.00	CLOSED FRACTURE OF PROXIMAL FEMUR, PERTROCHANTERIC
S132.00	CLOSED FRACTURE PUBIS

N338111	NONUNION OF FRACTURE
S312300	CLOSED FRACTURE DISTAL FEMUR, SUPRACONDYLAR
S210400	CLOSED FRACTURE SCAPULA, BLADE
S4A..00	FRACTURE-DISLOCATION OR SUBLUXATION SHOULDER
S240400	CLOSED FRACTURE PISIFORM
S15..00	FRACTURE OF THORACIC VERTEBRA
S100100	CLOSED FRACTURE ATLAS
S01..19	TEMPORAL BONE FRACTURE
S2...11	ARM FRACTURE
S356.00	FRACTURE OF METATARSAL BONE
S235100	OPEN COLLES' FRACTURE
S240100	CLOSED FRACTURE OF THE SCAPHOID
S2...00	FRACTURE OF UPPER LIMB
S23C.00	FRACTURE OF LOWER END OF BOTH ULNA AND RADIUS
S340.00	CLOSED FRACTURE ANKLE, MEDIAL MALLEOLUS
S263.00	FRACTURE OF OTHER FINGER
S312100	CLOSED FRACTURE OF FEMORAL CONDYLE, UNSPECIFIED
S235B00	OPEN FRACTURE RADIAL STYLOID
S250.00	CLOSED FRACTURE OF METACARPAL BONE(S)
SC3C000	SEQUELAE OF FRACTURE AT WRIST AND HAND LEVEL
S132100	CLOSED FRACTURE PELVIS, MULTIPLE PUBIC RAMI - STABLE
S349.00	FRACTURE OF LATERAL MALLEOLUS
S23..00	FRACTURE OF RADIUS AND ULNA
S339000	CLOSED FRACTURE OF DISTAL FIBULA
S310.00	CLOSED FRACTURE OF FEMUR, SHAFT OR UNSPECIFIED PART
S250z00	CLOSED FRACTURE OF METACARPAL BONE(S) NOS
S224100	CLOSED FRACTURE DISTAL HUMERUS, SUPRACONDYLAR
S234B00	CLOSED FRACTURE RADIAL STYLOID
S336.00	FRACTURE OF UPPER END OF TIBIA
S132000	CLOSED FRACTURE PELVIS, SINGLE PUBIC RAMUS
S230600	CLOSED FRACTURE RADIUS, HEAD
S342000	CLOSED FRACTURE ANKLE, LATERAL MALLEOLUS, LOW
S363.00	FRACTURE OF OTHER TOE
S344.00	CLOSED FRACTURE ANKLE, BIMALLEOLAR
S342.00	CLOSED FRACTURE ANKLE, LATERAL MALLEOLUS
S262.00	FRACTURE OF THUMB
S252.00	CLOSED FRACTURE SESAMOID BONE OF HAND
S242100	FRACTURE OF FIRST METACARPAL BONE
S231600	OPEN FRACTURE RADIAL HEAD
S230700	CLOSED FRACTURE RADIUS, NECK
S337.00	FRACTURE OF SHAFT OF TIBIA
S225.11	ELBOW FRACTURE - OPEN
S120000	CLOSED FRACTURE OF RIB, UNSPECIFIED
S239.00	FRACTURE OF SHAFT OF RADIUS
S31..00	OTHER FRACTURE OF FEMUR
S242.00	FRACTURE AT WRIST AND HAND LEVEL
S264.00	MULTIPLE FRACTURES OF FINGERS
S305.00	SUBTROCHANTERIC FRACTURE
S10..00	FRACTURE OF SPINE WITHOUT MENTION OF SPINAL CORD INJURY
S350.11	HEEL BONE FRACTURE
S104100	CLOSED FRACTURE LUMBAR VERTEBRA, WEDGE
S350.00	CLOSED FRACTURE OF CALCANEUS
S260.00	CLOSED FRACTURE OF ONE OR MORE PHALANGES OF HAND
S4A0.00	CLOSED FRACTURE-DISLOCATION SHOULDER

S238.00	FRACTURE OF SHAFT OF ULNA
S231B00	OPEN FRACTURE OLECRANON, INTRA-ARTICULAR
S334100	CLOSED FRACTURE DISTAL TIBIA, INTRA-ARTICULAR
ZV57700	[V]REHABILITATION FOLLOWING FRACTURE
S00..12	PARIETAL BONE FRACTURE
S315.00	FRACTURE OF LOWER END OF FEMUR
S10B600	MULTIPLE FRACTURES OF LUMBAR SPINE AND PELVIS
S314.00	FRACTURE OF SHAFT OF FEMUR
S302400	CLOSED FRACTURE OF FEMUR, INTERTROCHANTERIC
S224600	CLOSED FRACTURE DISTAL HUMERUS, LATERAL EPICONDYLE
S23y300	OPEN FRACTURE OF THE RADIUS AND ULNA
S3...00	FRACTURE OF LOWER LIMB
S293.00	MULTIPLE FRACTURES OF FOREARM
S120A00	COUGH FRACTURE
S10B400	FRACTURE OF ACETABULUM
S02..00	FRACTURE OF FACE BONES
S234300	CLOSED FRACTURE OF ULNA, STYLOID PROCESS
S3x4.00	MULTIPLE FRACTURES OF FOOT
S34z.00	FRACTURE OF ANKLE, NOS
S4C1000	OPEN FRACTURE-DISLOCATION, DISTAL RADIO-ULNAR JOINT
S3x3.00	MULTIPLE FRACTURES OF LOWER LEG
S221.00	OPEN FRACTURE OF THE PROXIMAL HUMERUS
S4B0000	CLOSED FRACTURE-DISLOCATION ELBOW JOINT
S230100	CLOSED FRACTURE OLECRANON, EXTRA-ARTICULAR
N338600	DELAYED UNION OF FRACTURE
S127.00	FRACTURE OF RIB
S01..15	OCCIPUT BONE FRACTURE
S020.11	CLOSED FRACTURE NASAL BONE
S347.00	OPEN FRACTURE ANKLE, TRIMALLEOLAR
S338.00	FRACTURE OF LOWER END OF TIBIA
N338.00	MALUNION AND NONUNION OF FRACTURE
S346.00	CLOSED FRACTURE ANKLE, TRIMALLEOLAR
S235.11	WRIST FRACTURE - OPEN
S234F00	CLOSED BARTON'S FRACTURE
S311100	OPEN FRACTURE SHAFT OF FEMUR
S23A.00	FRACTURE OF SHAFTS OF BOTH ULNA AND RADIUS
S24..11	HAND FRACTURE - CARPAL BONE
S4B..00	FRACTURE-DISLOCATION OR SUBLUXATION ELBOW
S231300	OPEN MONTEGGIA'S FRACTURE
S4D..00	FRACTURE-DISLOCATION/SUBLUXATION FINGER/THUMB
S1...00	FRACTURE OF NECK AND TRUNK
S240200	CLOSED FRACTURE LUNATE
S22z.00	FRACTURE OF HUMERUS NOS
S4D0500	CLOSED FRACTURE-DISLOCATION, INTERPHALANGEAL JOINT THUMB
S30y.11	HIP FRACTURE NOS
S221.11	SHOULDER FRACTURE - OPEN
S23..11	FOREARM FRACTURE
S127000	MULTIPLE FRACTURES OF RIBS
S21..11	SHOULDER BLADE FRACTURE
S10B000	FRACTURE OF LUMBAR VERTEBRA
S120900	CLOSED FRACTURE MULTIPLE RIBS
S220300	CLOSED FRACTURE PROXIMAL HUMERUS, GREATER TUBEROSITY
S234900	CLOSED VOLAR BARTON'S FRACTURE

S028300	FRACTURE OF MANDIBLE
S220.00	CLOSED FRACTURE OF THE PROXIMAL HUMERUS
S235300	OPEN FRACTURE OF ULNA, STYLOID PROCESS
S3X..00	FRACTURE OF LOWER LEG, PART UNSPECIFIED
S150.00	MULTIPLE FRACTURES OF THORACIC SPINE
S100.00	CLOSED FRACTURE OF CERVICAL SPINE
S220100	CLOSED FRACTURE PROXIMAL HUMERUS, NECK
S12z.11	RIB FRACTURE NOS
N331M00	FRAGILITY FRACTURE DUE TO UNSPECIFIED OSTEOPOROSIS
S352200	CLOSED FRACTURE NAVICULAR
S134z00	OTHER OR MULTIPLE CLOSED FRACTURE OF PELVIS NOS
S102y00	OTHER SPECIFIED CLOSED FRACTURE THORACIC VERTEBRA
S128.00	FRACTURE OF STERNUM
Q203.12	OTHER BIRTH FRACTURE
S230B00	CLOSED FRACTURE OLECRANON, INTRA-ARTICULAR
S022100	CLOSED FRACTURE OF MANDIBLE, CONDYLAR PROCESS
S339100	OPEN FRACTURE OF DISTAL FIBULA
S10B.00	FRACTURE OF LUMBAR SPINE AND PELVIS
S025100	OPEN FRACTURE ZYGOMA
S2B..00	FRACTURE OF BONE OF HAND
S250400	CLOSED FRACTURE FINGER METACARPAL NECK
S310011	THIGH FRACTURE NOS
N338100	PSEUDOARTHROSIS - FRACTURE NONUNION
S3xz.00	OTHER, MULTIPLE AND ILL-DEFINED FRACTURES OF LOWER LIMB NOS
S344.11	DUPUYTREN'S FRACTURE, FIBULA
S108.00	CLOSED FRACTURE PELVIS, COCCYX
S024.00	FRACTURE OF MALAR OR MAXILLARY BONES, CLOSED
S352100	CLOSED FRACTURE OF TALUS
S350.12	OS CALCIS FRACTURE
S00..00	FRACTURE OF VAULT OF SKULL
S224.00	CLOSED FRACTURE OF THE DISTAL HUMERUS
S4G..00	FRACTURE-DISLOCATION OR SUBLUXATION ANKLE
S100000	CLOSED FRACTURE OF UNSPECIFIED CERVICAL VERTEBRA
S240.00	CLOSED FRACTURE OF CARPAL BONE
S23x.00	CLOSED FRACTURE OF RADIUS AND ULNA, UNSPECIFIED PART
S106.00	CLOSED FRACTURE SACRUM
S352.00	CLOSED FRACTURE OF OTHER TARSAL AND METATARSAL BONES
S100200	CLOSED FRACTURE AXIS
S210100	CLOSED FRACTURE SCAPULA, ACROMION
S120100	CLOSED FRACTURE OF ONE RIB
TC7..00	FRACTURE, CAUSE UNSPECIFIED
S225700	OPEN FRACTURE DISTAL HUMERUS, MEDIAL EPICONDYLE
S022.12	FRACTURE OF LOWER JAW, CLOSED
N1y1.00	FATIGUE FRACTURE OF VERTEBRA
S292.00	MULTIPLE FRACTURES OF CLAVICLE, SCAPULA AND HUMERUS
S240A00	CLOSED FRACTURE SCAPHOID, PROXIMAL POLE
S29..12	MULTIPLE RIB FRACTURES
S4H..00	FRACTURE-DISLOCATION OR SUBLUXATION FOOT
S241100	OPEN FRACTURE OF THE SCAPHOID
S02z.11	JAW FRACTURE NOS
S240700	CLOSED FRACTURE CAPITATE
S230200	CLOSED FRACTURE OF ULNA, CORONOID
S4C2.00	CLOSED FRACTURE-SUBLUXATION OF THE WRIST

S4C0000	CLOSED FRACTURE-DISLOCATION DISTAL RADIO-ULNAR JOINT
S23x100	CLOSED FRACTURE OF RADIUS (ALONE), UNSPECIFIED
S201200	OPEN FRACTURE CLAVICLE, SHAFT
S30y.00	CLOSED FRACTURE OF NECK OF FEMUR NOS
S234.00	CLOSED FRACTURE OF RADIUS AND ULNA, LOWER END
S235F00	OPEN BARTON'S FRACTURE
S261.00	OPEN FRACTURE OF ONE OR MORE PHALANGES OF HAND
S260F00	CLOSED FRACTURE FINGER PROXIMAL PHALANX, SHAFT
S343.00	OPEN FRACTURE ANKLE, LATERAL MALLEOLUS
S234000	CLOSED FRACTURE OF FOREARM, LOWER END, UNSPECIFIED
S224200	CLOSED FRACTURE DISTAL HUMERUS, LATERAL CONDYLE
S345.00	OPEN FRACTURE ANKLE, BIMALLEOLAR
S4C..00	FRACTURE-DISLOCATION OR SUBLUXATION OF WRIST
S330300	CLOSED FRACTURE PROXIMAL TIBIA, MEDIAL CONDYLE (PLATEAU)
S4D1.00	OPEN FRACTURE-DISLOCATION DIGIT
S234D00	CLOSED FRACTURE DISTAL RADIUS, EXTRA-ARTICULAR, OTHER TYPE
S222000	CLOSED FRACTURE OF HUMERUS NOS
S10A200	MULTIPLE FRACTURES OF CERVICAL SPINE
SR1..00	FRACTURES INVOLVING MULTIPLE BODY REGIONS
S4D0400	CLOSED FRACTURE-DISLOCATION, PROXIMAL INTERPHALANGEAL JOINT
S302011	CLOSED FRACTURE OF FEMUR, GREATER TROCHANTER
S240B00	CLOSED FRACTURE SCAPHOID, WAIST, TRANSVERSE
S00z.00	FRACTURE OF VAULT OF SKULL NOS
S352000	CLOSED FRACTURE OF TARSAL BONE, UNSPECIFIED
S028100	FRACTURE OF ORBITAL FLOOR
S333200	OPEN FRACTURE OF TIBIA AND FIBULA, SHAFT
S310012	UPPER LEG FRACTURE NOS
S250x00	CLOSED FRACTURES OF MULTIPLE SITES OF UNSPECIFIED METACARPUS
S3x2.00	MULTIPLE FRACTURES OF FEMUR
S312200	CLOSED FRACTURE OF FEMUR, LOWER EPIPHYSIS
S312.11	CLOSED FRACTURE OF FEMUR, DISTAL END
S330400	CLOSED FRACTURE PROXIMAL TIBIA, LATERAL CONDYLE (PLATEAU)
S24..00	FRACTURE OF CARPAL BONE
S330012	CLOSED FRACTURE OF TIBIAL TUBEROSITY
S03z.11	DEPRESSED SKULL FRACTURE NOS
S301800	OPEN FRACTURE PROXIMAL FEMUR,SUBCAPITAL, GARDEN GRADE III
S234211	DUPUYTREN'S FRACTURE, RADIUS - CLOSED
S01..14	MIDDLE FOSSA FRACTURE
S30w.00	CLOSED FRACTURE OF UNSPECIFIED PROXIMAL FEMUR
S260D00	CLOSED FRACTURE FINGER PROXIMAL PHALANX
S4A2100	CLOSED FRACTURE-SUBLUXATION ACROMIO-CLAVICULAR JOINT
S4E..00	FRACTURE-DISLOCATION OR SUBLUXATION HIP
S250500	CLOSED FRACTURE FINGER METACARPAL HEAD
S4J1200	OPEN FRACTURE-DISLOCATION STERNO-CLAVICULAR JOINT, ANTERIOR
S352B00	CLOSED FRACTURE METATARSAL BASE
S23x200	CLOSED FRACTURE OF ULNA (ALONE), UNSPECIFIED
S101500	OPEN FRACTURE OF FIFTH CERVICAL VERTEBRA

S100700	CLOSED FRACTURE OF SEVENTH CERVICAL VERTEBRA
S310100	CLOSED FRACTURE SHAFT OF FEMUR
S026.00	CLOSED ORBITAL BLOW-OUT FRACTURE
S360000	CLOSED FRACTURE PROXIMAL PHALANX, TOE
S1z..00	FRACTURE OF NECK AND TRUNK NOS
S4D0100	CLOSED FRACTURE-DISLOCATION, METACARPOPHALANGEAL JOINT
S33z.00	FRACTURE OF TIBIA AND FIBULA, NOS
S250300	CLOSED FRACTURE FINGER METACARPAL SHAFT
S02z.00	FRACTURE OF FACIAL BONE NOS
S250800	CLOSED FRACTURE OF THUMB METACARPAL
S240500	CLOSED FRACTURE TRAPEZIUM
S232.00	CLOSED FRACTURE OF RADIUS AND ULNA, SHAFT
S02xz00	FRACTURE OF OTHER FACIAL BONES, CLOSED, NOS
S02x000	FRACTURE OF ALVEOLUS, CLOSED
S102.00	CLOSED FRACTURE THORACIC VERTEBRA
S031.00	CLOSED FRACTURE OF SKULL NOS WITH INTRACRANIAL INJURY
S352C00	CLOSED FRACTURE METATARSAL SHAFT
S100500	CLOSED FRACTURE OF FIFTH CERVICAL VERTEBRA
S235.00	OPEN FRACTURE OF RADIUS AND ULNA, LOWER END
S234z00	CLOSED FRACTURE OF FOREARM, LOWER END, NOS
S294.00	FRACTURES INVOLVING MULTIPLE REGIONS OF BOTH UPPER LIMBS
S210.00	CLOSED FRACTURE OF SCAPULA
S260S00	CLOSED FRACTURE FINGER DISTAL PHALANX, BASE
S100600	CLOSED FRACTURE OF SIXTH CERVICAL VERTEBRA
S001.00	CLOSED FRACTURE VAULT OF SKULL WITH INTRACRANIAL INJURY
S260E00	CLOSED FRACTURE FINGER PROXIMAL PHALANX, BASE
S4A0100	CLOSED FRACTURE-DISLOCATION ACROMIO-CLAVICULAR JOINT
S334.00	CLOSED FRACTURE DISTAL TIBIA
S335000	OPEN FRACTURE DISTAL TIBIA, EXTRA-ARTICULAR
S4C0300	CLOSED FRACTURE-DISLOCATION, CARPOMETACARPAL JOINT
S233z00	OPEN FRACTURE OF RADIUS AND ULNA, SHAFT, NOS
S12z.12	STERNUM FRACTURE NOS
S134600	CLOSED FRACTURE PELVIS, ILIAC WING
S250600	CLOSED FRACTURE FINGER METACARPAL
S222100	CLOSED FRACTURE OF HUMERUS, SHAFT
S10B300	FRACTURE OF ILIUM
SC3D400	SEQUELAE OF FRACTURE OF FEMUR
S335.00	OPEN FRACTURE DISTAL TIBIA
S20z.00	FRACTURE OF CLAVICLE NOS
S333.00	OPEN FRACTURE OF TIBIA/FIBULA, SHAFT
S4G3.00	OPEN FRACTURE-SUBLUXATION, ANKLE JOINT
S333000	OPEN FRACTURE SHAFT OF TIBIA
S10A000	FRACTURE OF FIRST CERVICAL VERTEBRA
S200z00	CLOSED FRACTURE OF CLAVICLE NOS
S261R00	OPEN FRACTURE FINGER DISTAL PHALANX
S333z00	OPEN FRACTURE OF TIBIA AND FIBULA, SHAFT, NOS
S33y.00	OPEN FRACTURE OF TIBIA AND FIBULA, UNSPECIFIED PART, NOS
S134400	CLOSED FRACTURE PELVIS, ANTERIOR SUPERIOR ILIAC SPINE
S120z00	CLOSED FRACTURE OF RIB(S) NOS

S260800	CLOSED FRACTURE THUMB DISTAL PHALANX
S360200	CLOSED FRACTURE DISTAL PHALANX, TOE
S321.00	OPEN FRACTURE OF THE PATELLA
S234E00	CLOSED FRACTURE DISTAL RADIUS, INTRA-ARTICULAR, OTHER TYPE
S200200	CLOSED FRACTURE CLAVICLE, SHAFT
S33y100	OPEN FRACTURE OF FIBULA, UNSPECIFIED PART, NOS
S352F00	CLOSED FRACTURE METATARSAL, MULTIPLE
S13y.00	CLOSED FRACTURE OF PELVIS NOS
S224800	CLOSED FRACTURE DISTAL HUMERUS, CAPITELLUM
S240300	CLOSED FRACTURE TRIQUETRAL
S240E00	CLOSED FRACTURE SCAPHOID, TUBEROSITY
S332100	CLOSED FRACTURE SHAFT OF FIBULA
S102100	CLOSED FRACTURE THORACIC VERTEBRA, WEDGE
S127100	COUGH FRACTURE OF RIBS
S330000	CLOSED FRACTURE OF THE PROXIMAL TIBIA
S362000	CLOSED FRACTURE OF GREAT TOE
S132z00	CLOSED FRACTURE PUBIS NOS
S234600	CLOSED FRACTURE RADIUS AND ULNA, DISTAL
S224700	CLOSED FRACTURE DISTAL HUMERUS, MEDIAL EPICONDYLE
S4F1.00	OPEN FRACTURE-DISLOCATION, KNEE JOINT
S220400	CLOSED FRACTURE PROXIMAL HUMERUS, HEAD
S23y200	OPEN FRACTURE OF ULNA (ALONE), UNSPECIFIED
S4H0400	CLOSED FRACTURE-DISLOCATION, IPJ, SINGLE TOE
S352J00	CLOSED FRACTURE OF BASE OF FIFTH METATARSAL
S022400	CLOSED FRACTURE OF MANDIBLE, RAMUS, UNSPECIFIED
S312.00	CLOSED FRACTURE DISTAL FEMUR
S304.00	PERTROCHANTERIC FRACTURE
S33y200	OPEN FRACTURE OF TIBIA AND FIBULA, UNSPECIFIED PART
S104400	CLOSED FRACTURE LUMBAR VERTEBRA, TRANSVERSE PROCESS
S022000	CLOSED FRACTURE MANDIBLE (SITE UNSPECIFIED)
S33x.00	CLOSED FRACTURE OF TIBIA AND FIBULA, UNSPECIFIED PART, NOS
S251.00	OPEN FRACTURE OF METACARPAL BONE(S)
S260R00	CLOSED FRACTURE FINGER DISTAL PHALANX
S02x.00	CLOSED FRACTURE OTHER FACIAL BONE
S332.00	CLOSED FRACTURE OF TIBIA/FIBULA, SHAFT
S220700	CLOSED FRACTURE PROXIMAL HUMERUS, FOUR PART
S302200	CLOSED FRACTURE PROXIMAL FEMUR, SUBTROCHANTERIC
S231100	OPEN FRACTURE OLECRANON, EXTRA-ARTICULAR
S3zz.00	FRACTURE OF BONES NOS
S33y000	OPEN FRACTURE OF TIBIA, UNSPECIFIED PART, NOS
S352D00	CLOSED FRACTURE METATARSAL NECK
S360100	CLOSED FRACTURE MIDDLE PHALANX, TOE
S200300	CLOSED FRACTURE CLAVICLE, LATERAL END
S4F..00	FRACTURE-DISLOCATION OR SUBLUXATION KNEE
S4F4.00	CLOSED FRACTURE-DISLOCATION, PATELLO-FEMORAL JOINT
S028200	FRACTURE OF MALAR AND MAXILLARY BONES
S10..11	FRACTURE OF TRANSVERSE PROCESS SPINE - NO SPINAL CORD LESION
S2z..00	FRACTURE OF UPPER LIMB NOS
S028.00	FRACTURE OF SKULL AND FACIAL BONES
Z6G1900	FRACTURE - TRACTION
S235800	OPEN GALEAZZI FRACTURE

S352111	CLOSED FRACTURE OF ASTRAGALUS
S227.00	FRACTURE OF SHAFT OF HUMERUS
S01..12	ETHMOID SINUS FRACTURE
S11x.00	CLOSED FRACTURE OF SPINE WITH SPINAL CORD LESION UNSPECIFIED
S02y100	FRACTURE OF ORBIT NOS, OPEN
S250000	CLOSED FRACTURE OF METACARPAL BONE (S), SITE UNSPECIFIED
S112700	CLS SPINAL FRACTURE WITH COMPLETE THORAC CORD LESION, T7-12
S29..11	MULTIPLE FRACTURES OF ARM
S225100	OPEN FRACTURE DISTAL HUMERUS, SUPRACONDYLAR
S04..11	MULTIPLE FACE FRACTURES
S362100	OPEN FRACTURE OF GREAT TOE
S240800	CLOSED FRACTURE HAMATE
SR10.00	FRACTURES INVOLVING HEAD WITH NECK
S025000	OPEN FRACTURE MAXILLA
S11..00	FRACTURE OF SPINE WITH SPINAL CORD LESION
S6z..00	INTRACRANIAL INJURY, EXCLUDING THOSE WITH SKULL FRACTURE NOS
S224400	CLOSED FRACTURE OF DISTAL HUMERUS, CONDYLE(S) UNSPECIFIED
S125200	CLOSED FRACTURE OF THYROID CARTILAGE
S225800	OPEN FRACTURE DISTAL HUMERUS, CAPITELLUM
S313.11	OPEN FRACTURE OF FEMUR, DISTAL END
S320.00	CLOSED FRACTURE OF THE PATELLA
S236.00	FRACTURE OF UPPER END OF ULNA
S331100	OPEN FRACTURE PROXIMAL FIBULA
S027.00	OPEN ORBITAL BLOW-OUT FRACTURE
S321200	OPEN FRACTURE PATELLA, DISTAL POLE
S220200	CLOSED FRACTURE OF PROXIMAL HUMERUS, ANATOMICAL NECK
S100611	C6 VERTEBRA CLOSED FRACTURE WITHOUT SPINAL CORD LESION
S044.00	MULTIPLE FRACTURES INVOLVING SKULL AND FACIAL BONES
S332200	CLOSED FRACTURE OF TIBIA AND FIBULA, SHAFT
S224z00	CLOSED FRACTURE OF DISTAL HUMERUS, NOT OTHERWISE SPECIFIED
S260300	CLOSED FRACTURE THUMB PROXIMAL PHALANX
S01z.00	FRACTURE OF BASE OF SKULL NOS
S260L00	CLOSED FRACTURE FINGER MIDDLE PHALANX, BASE
S260K00	CLOSED FRACTURE FINGER MIDDLE PHALANX
S260700	CLOSED FRACTURE THUMB PROXIMAL PHALANX, HEAD
S330100	CLOSED FRACTURE PROXIMAL FIBULA
S4F2.00	CLOSED FRACTURE-SUBLUXATION, KNEE JOINT
S260400	CLOSED FRACTURE THUMB PROXIMAL PHALANX, BASE
S260G00	CLOSED FRACTURE FINGER PROXIMAL PHALANX, NECK
S223.00	OPEN FRACTURE OF HUMERUS, SHAFT OR UNSPECIFIED PART
S261D00	OPEN FRACTURE FINGER PROXIMAL PHALANX
S4B0.00	CLOSED FRACTURE-DISLOCATION ELBOW
S0z..00	FRACTURE OF SKULL NOS
S260T00	CLOSED FRACTURE FINGER DISTAL PHALANX, SHAFT
S4A0000	CLOSED FRACTURE-DISLOCATION SHOULDER JOINT
S331000	OPEN FRACTURE OF THE PROXIMAL TIBIA
S224000	CLOSED FRACTURE OF ELBOW, UNSPECIFIED PART

S200.00	CLOSED FRACTURE OF CLAVICLE
S260M00	CLOSED FRACTURE FINGER MIDDLE PHALANX, SHAFT
S330600	CLOSED FRACTURE SPINE, TIBIA
S232z00	CLOSED FRACTURE OF RADIUS AND ULNA, SHAFT, NOS
S260U00	CLOSED FRACTURE FINGER DISTAL PHALANX, TUFT
S261M00	OPEN FRACTURE FINGER MIDDLE PHALANX, SHAFT
S21z.00	FRACTURE OF SCAPULA NOS
S2A..00	FRACTURE OF UPPER LIMB, LEVEL UNSPECIFIED
S230900	CLOSED FRACTURE OF THE PROXIMAL RADIUS
S37..00	FRACTURE OF LOWER LIMB, LEVEL UNSPECIFIED
S250200	CLOSED FRACTURE FINGER METACARPAL BASE
S27..00	MULTIPLE FRACTURES OF HAND BONES
SC0X.00	SEQUELAE OF OTHER FRACTURE OF THORAX AND PELVIS
S240z00	CLOSED FRACTURE OF CARPAL BONE NOS
S230000	CLOSED FRACTURE OF PROXIMAL FOREARM, UNSPECIFIED PART
S300700	CLOSED FRACTURE PROXIMAL FEMUR, SUBCAPITAL, GARDEN GRADE II
S134.00	OTHER OR MULTIPLE CLOSED FRACTURE OF PELVIS
S100211	C2 VERTEBRA CLOSED FRACTURE WITHOUT SPINAL CORD LESION
S341.00	OPEN FRACTURE ANKLE, MEDIAL MALLEOLUS
S4D0300	CLOSED FRACTURE-DISLOCATION, DISTAL INTERPHALANGEAL JOINT
S271.00	OPEN MULTIPLE FRACTURES OF HAND BONES
S332000	CLOSED FRACTURE SHAFT OF TIBIA
S260000	CLOSED FRACTURE OF PHALANX OR PHALANGES, UNSPECIFIED
S300900	CLOSED FRACTURE PROXIMAL FEMUR, SUBCAPITAL, GARDEN GRADE IV
S260900	CLOSED FRACTURE THUMB DISTAL PHALANX, BASE
S311000	OPEN FRACTURE OF FEMUR, UNSPECIFIED PART
S334000	CLOSED FRACTURE DISTAL TIBIA, EXTRA-ARTICULAR
S10z.00	FRACTURE OF SPINE WITHOUT MENTION OF SPINAL CORD LESION NOS
S225.00	OPEN FRACTURE OF THE DISTAL HUMERUS
S134800	CLOSED FRACTURE DISLOCATION OF SACRO-ILIAC JOINT
S120400	CLOSED FRACTURE OF FOUR RIBS
S4J2100	CLOSED FRACTURE-SUBLUXATION OF PELVIS
S4G0.00	CLOSED FRACTURE-DISLOCATION, ANKLE JOINT
S260P00	CLOSED FRACTURE FINGER MIDDLE PHALANX, HEAD
S300600	CLOSED FRACTURE PROXIMAL FEMUR, SUBCAPITAL, GARDEN GRADE I
S260N00	CLOSED FRACTURE FINGER MIDDLE PHALANX, NECK
S23y100	OPEN FRACTURE OF RADIUS (ALONE), UNSPECIFIED
S230800	CLOSED FRACTURE PROXIMAL RADIUS, COMMINUTED
S230400	CLOSED FRACTURE OF PROXIMAL ULNA, COMMINUTED
S10A100	FRACTURE OF SECOND CERVICAL VERTEBRA
S36z.00	FRACTURE OF ONE OR MORE PHALANGES OF FOOT NOS
S230500	CLOSED FRACTURE OF THE PROXIMAL ULNA
S4C0.00	CLOSED FRACTURE DISLOCATION OF WRIST
S133000	OPEN FRACTURE PELVIS, SINGLE PUBIC RAMUS
S134100	CLOSED FRACTURE PELVIS, ISCHIUM
S350100	CLOSED FRACTURE CALCANEUS, INTRA-ARTICULAR
S235700	OPEN SMITH'S FRACTURE
S231700	OPEN FRACTURE RADIAL NECK

S313500	OPEN FRACTURE DISTAL FEMUR, LATERAL CONDYLE
S100511	C5 VERTEBRA CLOSED FRACTURE WITHOUT SPINAL CORD LESION
S210200	CLOSED FRACTURE SCAPULA, CORACOID
S4J0100	CLOSED FRACTURE-DISLOCATION OF PELVIS
S352400	CLOSED FRACTURE MEDIAL CUNEIFORM
S4J0.00	OTHER CLOSED FRACTURE-DISLOCATION
S32z.00	FRACTURE OF PATELLA, NOS
S135z00	OTHER/MULTIPLE OPEN FRACTURE OF PELVIS NOS
S232300	CLOSED FRACTURE RADIUS AND ULNA, MIDDLE
S352E00	CLOSED FRACTURE METATARSAL HEAD
S104300	CLOSED FRACTURE LUMBAR VERTEBRA, SPINOUS PROCESS
S4D0.00	CLOSED FRACTURE-DISLOCATION DIGIT
S33x.11	LOWER LEG FRACTURE NOS
S150000	CLOSED MULTIPLE FRACTURES OF THORACIC SPINE
S02yz00	FRACTURE OF OTHER FACIAL BONES,OPEN, NOS
S211.00	OPEN FRACTURE OF SCAPULA
S6...00	INTRACRANIAL INJURY EXCLUDING THOSE WITH SKULL FRACTURE
S342100	CLOSED FRACTURE ANKLE, LATERAL MALLEOLUS, HIGH
S4A2.00	CLOSED FRACTURE-SUBLUXATION SHOULDER
S240600	CLOSED FRACTURE TRAPEZOID
S112.00	CLOSED FRACTURE OF THORACIC SPINE WITH SPINAL CORD LESION
S022z00	FRACTURE OF MANDIBLE, CLOSED, NOS
S352H00	CLOSED FRACTURE OF CUNEIFORMS
S23xz00	CLOSED FRACTURE OF RADIUS AND ULNA, NOS
S210600	CLOSED FRACTURE SCAPULA, NECK
S300400	CLOSED FRACTURE HEAD OF FEMUR
S024.11	FRACTURE OF UPPER JAW, CLOSED
S222.00	CLOSED FRACTURE OF HUMERUS, SHAFT OR UNSPECIFIED PART
S4D2100	CLOSED FRACTURE-SUBLUXATION, METACARPOPHALANGEAL JOINT
S261000	OPEN FRACTURE OF PHALANX OR PHALANGES, UNSPECIFIED
S300800	CLOSED FRACTURE PROXIMAL FEMUR, SUBCAPITAL, GARDEN GRADE III
S023.11	FRACTURE OF LOWER JAW, OPEN
S021.11	OPEN FRACTURE NASAL BONE
S3x0.00	OTHER, MULTIPLE AND ILL-DEFINED CLOSED FRACTURES LOWER LIMB
S3z0.00	CLOSED FRACTURE OF BONES, UNSPECIFIED
S352800	CLOSED FRACTURE TALUS, HEAD
S4D1400	OPEN FRACTURE-DISLOCATION, PROXIMAL INTERPHALANGEAL JOINT
S01..16	ORBITAL ROOF FRACTURE
S310000	CLOSED FRACTURE OF FEMUR, UNSPECIFIED PART
S01..18	SPHENOID BONE FRACTURE
S361.00	OPEN FRACTURE OF ONE OR MORE PHALANGES OF FOOT
S231A00	OPEN FRACTURE RADIUS AND ULNA, PROXIMAL
S022x00	CLOSED FRACTURE OF MANDIBLE, MULTIPLE SITES
S261800	OPEN FRACTURE THUMB DISTAL PHALANX
S210000	CLOSED FRACTURE OF SCAPULA, UNSPECIFIED PART
S023.00	FRACTURE OF MANDIBLE, OPEN
S100711	C7 VERTEBRA CLOSED FRACTURE WITHOUT SPINAL CORD LESION

S30z.00	OPEN FRACTURE OF NECK OF FEMUR NOS
S220z00	CLOSED FRACTURE OF PROXIMAL HUMERUS NOT OTHERWISE SPECIFIED
S312500	CLOSED FRACTURE DISTAL FEMUR, LATERAL CONDYLE
S235600	OPEN FRACTURE RADIUS AND ULNA, DISTAL
S4C0100	CLOSED FRACTURE-DISLOCATION RADIOCARPAL JOINT
S300.00	CLOSED FRACTURE PROXIMAL FEMUR, TRANSCERVICAL
S4J1.00	OTHER OPEN FRACTURE-DISLOCATION
S330700	CLOSED FRACTURE TUBERCLE, TIBIA
S34y.00	OPEN FRACTURE ANKLE, UNSPECIFIED
S301500	OPEN FRACTURE PROXIMAL FEMUR,SUBCAPITAL, GARDEN GRADE UNSPEC
S132y00	OTHER SPECIFIED CLOSED FRACTURE PUBIS
S4F7.00	OPEN FRACTURE-SUBLUXATION, PATELLO-FEMORAL JOINT
S303400	OPEN FRACTURE OF FEMUR, INTERTROCHANTERIC
S240900	CLOSED FRACTURE HAMATE, HOOK
S4D0200	CLOSED FRACTURE-DISLOCATION IPJ, UNSPECIFIED
S35z.00	FRACTURE OF TARSAL AND METATARSAL BONES NOS
S353100	OPEN FRACTURE OF TALUS
S102000	CLOSED FRACTURE THORACIC VERTEBRA, BURST
S04..12	MULTIPLE SKULL FRACTURES
S100A00	CLOSED FRACTURE AXIS, ODONTOID PROCESS
S4H1300	OPEN FRACTURE-DISLOCATION, METATARSOPHALANGEAL JOINT, SINGLE
S331.00	OPEN FRACTURE OF TIBIA AND FIBULA, PROXIMAL
S125100	CLOSED FRACTURE OF HYOID BONE
S330500	CLOSED FRACTURE PROXIMAL TIBIA, BICONDYLAR
S4E0.00	CLOSED FRACTURE-DISLOCATION, HIP JOINT
S234800	CLOSED GALEAZZI FRACTURE
S261K00	OPEN FRACTURE FINGER MIDDLE PHALANX
S220600	CLOSED FRACTURE PROXIMAL HUMERUS, THREE PART
S223100	OPEN FRACTURE OF HUMERUS, SHAFT
S251600	OPEN FRACTURE FINGER METACARPAL
S224300	CLOSED FRACTURE DISTAL HUMERUS, MEDIAL CONDYLE
S370.00	CLOSED FRACTURE OF LOWER LIMB, LEVEL UNSPECIFIED
S125000	CLOSED FRACTURE LARYNX
S4H1000	OPEN FRACTURE-DISLOCATION, SUBTALAR JOINT
S234500	CLOSED FRACTURE DISTAL ULNA, UNSPECIFIED
S120200	CLOSED FRACTURE OF TWO RIBS
S260z00	CLOSED FRACTURE OF ONE OR MORE PHALANGES OF HAND NOS
S134500	CLOSED FRACTURE PELVIS, ANTERIOR INFERIOR ILIAC SPINE
S134000	CLOSED FRACTURE OF ILIUM, UNSPECIFIED
S4F6.00	CLOSED FRACTURE-SUBLUXATION, PATELLO-FEMORAL JOINT
S4F0.00	CLOSED FRACTURE-DISLOCATION, KNEE JOINT
S353H00	OPEN FRACTURE CUNEIFORMS
S4J..00	OTHER FRACTURE-DISLOCATION OR SUBLUXATION
S4D2400	CLOSED FRACTURE-SUBLUXATION, PROXIMAL INTERPHALANGEAL JOINT
S292000	CLOSED MULTIPLE FRACTURES OF CLAVICLE, SCAPULA AND HUMERUS
S353200	OPEN FRACTURE NAVICULAR
S102z00	CLOSED FRACTURE THORACIC VERTEBRA NOT OTHERWISE SPECIFIED
S320400	CLOSED FRACTURE PATELLA, COMMINUTED (STELLATE)

S100400	CLOSED FRACTURE OF FOURTH CERVICAL VERTEBRA
S4D2000	CLOSED FRACTURE-SUBLUXATION DIGIT, UNSPECIFIED
S01..13	FRONTAL SINUS FRACTURE
S134300	CLOSED FRACTURE PELVIS, ISCHIAL TUBEROSITY
S022500	CLOSED FRACTURE OF MANDIBLE, ANGLE OF JAW
S022200	CLOSED FRACTURE OF MANDIBLE, SUBCONDYLAR
S100z00	CLOSED FRACTURE OF CERVICAL SPINE NOT OTHERWISE SPECIFIED
S33xz00	CLOSED FRACTURE OF TIBIA AND FIBULA, UNSPECIFIED PART, NOS
N338300	ATROPHIC NON-UNION OF FRACTURE
S234400	CLOSED FRACTURE OF ULNA, LOWER EPIPHYSIS
S260B00	CLOSED FRACTURE THUMB DISTAL PHALANX, TUFT
S100C00	CLOSED FRACTURE AXIS, SPINOUS PROCESS
S100111	C1 VERTEBRA CLOSED FRACTURE - NO SPINAL CORD LESION
S105.00	OPEN FRACTURE LUMBAR VERTEBRA
S313300	OPEN FRACTURE DISTAL FEMUR, SUPRACONDYLAR
S4C1.00	OPEN FRACTURE DISLOCATION WRIST
S232100	CLOSED FRACTURE OF THE RADIAL SHAFT
S4H1400	OPEN FRACTURE-DISLOCATION, IPJ, SINGLE TOE
S230z00	CLOSED FRACTURE OF PROXIMAL FOREARM NOT OTHERWISE SPECIFIED
S104000	CLOSED FRACTURE LUMBAR VERTEBRA, BURST
S344000	CLOSED FRACTURE ANKLE, BIMALLEOLAR, LOW FIBULAR FRACTURE
S311.00	OPEN FRACTURE OF FEMUR, SHAFT OR UNSPECIFIED PART
S330z00	CLOSED FRACTURE OF TIBIA AND FIBULA, PROXIMAL NOS
S4D0000	CLOSED FRACTURE-DISLOCATION DIGIT, UNSPECIFIED
S112600	CLS SPINAL FRACTURE WITH UNSPEC THORACIC CORD LESION, T7-12
S352z00	CLOSED FRACTURE OF ONE OR MORE TARSAL + METATARSAL BONES NOS
S351.00	OPEN FRACTURE OF CALCANEUS
S4D1300	OPEN FRACTURE-DISLOCATION, DISTAL INTERPHALANGEAL JOINT
S135400	OPEN FRACTURE PELVIS, ANTERIOR SUPERIOR ILIAC SPINE
S343000	OPEN FRACTURE ANKLE, LATERAL MALLEOLUS, LOW
S230.00	CLOSED FRACTURE OF PROXIMAL RADIUS AND ULNA
S260500	CLOSED FRACTURE THUMB PROXIMAL PHALANX, SHAFT
S11..12	FRACTURE OF VERTEBRA WITH SPINAL CORD LESION
S24z.00	FRACTURE OF CARPAL BONE NOS
S4B1.00	OPEN FRACTURE-DISLOCATION ELBOW
S114.00	CLOSED FRACTURE OF LUMBAR SPINE WITH SPINAL CORD LESION
S225000	OPEN FRACTURE OF ELBOW, UNSPECIFIED PART
S241400	OPEN FRACTURE PISIFORM
S371.00	OPEN FRACTURE OF LOWER LIMB, LEVEL UNSPECIFIED
S331300	OPEN FRACTURE PROXIMAL TIBIA, MEDIAL CONDYLE (PLATEAU)
S320200	CLOSED FRACTURE PATELLA, DISTAL POLE
S02A.00	LE FORT I FRACTURE MAXILLA
S270.00	CLOSED MULTIPLE FRACTURES OF HAND BONES
S230A00	CLOSED FRACTURE RADIUS AND ULNA, PROXIMAL
S4H0.00	CLOSED FRACTURE-DISLOCATION FOOT
S4C2000	CLOSED FRACTURE-SUBLUXATION, DISTAL RADIO-ULNAR JT

S4H2.00	CLOSED FRACTURE-SUBLUXATION, FOOT
S261E00	OPEN FRACTURE FINGER PROXIMAL PHALANX, BASE
S200100	CLOSED FRACTURE CLAVICLE, MEDIAL END
S240D00	CLOSED FRACTURE SCAPHOID, WAIST, COMMINUTED
S200000	CLOSED FRACTURE OF CLAVICLE, UNSPECIFIED PART
S220000	CLOSED FRACTURE OF PROXIMAL HUMERUS, UNSPECIFIED PART
S260J00	CLOSED FRACTURE FINGER PROXIMAL PHALANX, MULTIPLE
S331400	OPEN FRACTURE PROXIMAL TIBIA, LATERAL CONDYLE (PLATEAU)
S232200	CLOSED FRACTURE OF THE ULNAR SHAFT
S121200	OPEN FRACTURE OF TWO RIBS
S330.00	CLOSED FRACTURE OF TIBIA AND FIBULA, PROXIMAL
S234C00	CLOSED FRACTURE DISTAL RADIUS, INTRA-ARTICULAR, DIE-PUNCH
SC0z.11	DELAYED UNION OF FRACTURE
S235200	OPEN FRACTURE OF THE DISTAL RADIUS, UNSPECIFIED
S260H00	CLOSED FRACTURE FINGER PROXIMAL PHALANX, HEAD
S02B.00	LE FORT II FRACTURE MAXILLA
S4J1100	OPEN FRACTURE-DISLOCATION OF PELVIS
S260W00	CLOSED FRACTURE FINGER DISTAL PHALANX, MULTIPLE
S302100	CLOSED FRACTURE PROXIMAL FEMUR, INTERTROCHANTERIC, TWO PART
S221z00	OPEN FRACTURE OF PROXIMAL HUMERUS NOT OTHERWISE SPECIFIED
S260x00	CLOSED FRACTURES OF PHALANX OR PHALANGES, MULTIPLE SITES
S3x..00	OTHER, MULTIPLE AND ILL-DEFINED FRACTURES OF LOWER LIMB
S130z00	CLOSED FRACTURE ACETABULUM NOS
S313000	OPEN FRACTURE DISTAL FEMUR, UNSPECIFIED
S312400	CLOSED FRACTURE DISTAL FEMUR, MEDIAL CONDYLE
S352A00	CLOSED FRACTURE TALUS, BODY
S4D1500	OPEN FRACTURE-DISLOCATION, INTERPHALANGEAL JOINT THUMB
S231.00	OPEN FRACTURE OF PROXIMAL RADIUS AND ULNA
S12z.00	FRACTURE OF RIB(S), STERNUM, LARYNX OR TRACHEA NOS
S300A00	CLOSED FRACTURE OF FEMUR, UPPER EPIPHYSIS
SR11.00	FRACTURES INVOLVING THORAX WITH LOWER BACK AND PELVIS
S04..00	MULTIPLE FRACTURES INVOLVING SKULL OR FACE WITH OTHER BONES
S132200	CLOSED FRACTURE PELVIS, MULTIPLE PUBIC RAMI - UNSTABLE
S4C3100	OPEN FRACTURE-SUBLUXATION RADIOCARPAL JOINT
S352900	CLOSED FRACTURE TALUS, NECK
S280.00	CLOSED ILL-DEFINED FRACTURES OF UPPER LIMB
Q203.11	OTHER FRACTURES DUE TO BIRTH TRAUMA
S347000	OPEN FRACTURE ANKLE, TRIMALLEOLAR, LOW FIBULAR FRACTURE
S240C00	CLOSED FRACTURE SCAPHOID, WAIST, OBLIQUE
S225z00	OPEN FRACTURE OF DISTAL HUMERUS, NOT OTHERWISE SPECIFIED
S000.00	CLOSED FRACTURE VAULT OF SKULL WITHOUT INTRACRANIAL INJURY
S251z00	OPEN FRACTURE OF METACARPAL BONE(S) NOS

S313200	OPEN FRACTURE OF FEMUR, LOWER EPIPHYSIS
S225600	OPEN FRACTURE DISTAL HUMERUS, LATERAL EPICONDYLE
S121.00	OPEN FRACTURE RIB
S221300	OPEN FRACTURE PROXIMAL HUMERUS, GREATER TUBEROSITY
S233300	OPEN FRACTURE RADIUS AND ULNA, MIDDLE
S302012	CLOSED FRACTURE OF FEMUR, LESSER TROCHANTER
S025.00	FRACTURE OF MALAR OR MAXILLARY BONES, OPEN
S261S00	OPEN FRACTURE FINGER DISTAL PHALANX, BASE
S210500	CLOSED FRACTURE SCAPULA, SPINE
S4D2.00	CLOSED FRACTURE-SUBLUXATION DIGIT
S102400	CLOSED FRACTURE THORACIC VERTEBRA, TRANSVERSE PROCESS
S353z00	OPEN FRACTURE OF TARSAL AND METATARSAL BONES NOS
S112100	CLS SPINAL FRACTURE WTH COMPLETE THORACIC CORD LESION,T1-6
S223000	OPEN FRACTURE OF HUMERUS NOS
S300y00	CLOSED FRACTURE PROXIMAL FEMUR, OTHER TRANSCERVICAL
S4C1100	OPEN FRACTURE-DISLOCATION RADIOCARPAL JOINT
S28..00	ILL-DEFINED FRACTURES OF UPPER LIMB
S320000	CLOSED FRACTURE PATELLA, TRANSVERSE
S114000	CLOSED SPINAL FRACTURE WITH UNSPECIFIED LUMBAR CORD LESION
S241E00	OPEN FRACTURE SCAPHOID, TUBEROSITY
S313100	OPEN FRACTURE OF FEMORAL CONDYLE, UNSPECIFIED
S250B00	CLOSED FRACTURE THUMB METACARPAL NECK
S024z00	FRACTURE OF MALAR OR MAXILLARY BONES, CLOSED, NOS
S235500	OPEN FRACTURE DISTAL ULNA - OTHER
S331700	OPEN FRACTURE TUBERCLE, TIBIA
S331012	OPEN FRACTURE OF TIBIAL TUBEROSITY
S353300	OPEN FRACTURE CUBOID
S02y000	FRACTURE OF ALVEOLUS, OPEN
S352600	CLOSED FRACTURE LATERAL CUNEIFORM
S234A00	CLOSD DORSAL BARTON'S FRACTURE
S4C2100	CLOSED FRACTURE-SUBLUXATION RADIOCARPAL JOINT
S231000	OPEN FRACTURE OF PROXIMAL FOREARM, UNSPECIFIED
S321000	OPEN FRACTURE PATELLA, TRANSVERSE
S04z.00	MULTIPLE FRACTURES INVOLVING SKULL/FACE WITH OTHER BONES NOS
N338200	HYPERTROPHIC NON-UNION OF FRACTURE
S321400	OPEN FRACTURE PATELLA, COMMINUTED (STELLATE)
SR1z.00	MULTIPLE FRACTURES, UNSPECIFIED
N338z00	FRACTURE MALUNION OR NONUNION NOS
S361000	OPEN FRACTURE PROXIMAL PHALANX, TOE
S320100	CLOSED FRACTURE PATELLA, PROXIMAL POLE
S4A1100	OPEN FRACTURE-DISLOCATION ACROMIO-CLAVICULAR JOINT
S250A00	CLOSED FRACTURE THUMB METACARPAL SHAFT
S23x000	CLOSED FRACTURE OF FOREARM, UNSPECIFIED
S133.00	OPEN FRACTURE OF PUBIS
S261z00	OPEN FRACTURE OF ONE OR MORE PHALANGES OF HAND NOS
S118.00	CLOSED FRACTURE OF COCCYX WITH SPINAL CORD LESION
S133100	OPEN FRACTURE PELVIS, MULTIPLE PUBIC RAMI - STABLE
S313.00	OPEN FRACTURE DISTAL FEMUR
S361200	OPEN FRACTURE DISTAL PHALANX, TOE
S003.00	OPEN FRACTURE VAULT OF SKULL WITH INTRACRANIAL INJURY

S232000	CLOSED FRACTURE OF RADIUS, SHAFT, UNSPECIFIED
S241300	OPEN FRACTURE TRIQUETRAL
S261U00	OPEN FRACTURE FINGER DISTAL PHALANX, TUFT
S300311	CLOSED FRACTURE, BASE OF NECK OF FEMUR
S333100	OPEN FRACTURE SHAFT OF FIBULA
S4D3300	OPEN FRACTURE-SUBLUXATION, DISTAL INTERPHALANGEAL JOINT
S301900	OPEN FRACTURE PROXIMAL FEMUR,SUBCAPITAL, GARDEN GRADE IV
S253.00	OPEN FRACTURE SESAMOID BONE OF HAND
S224500	CLOSED FRACTURE OF DISTAL HUMERUS, TROCHLEA
S300300	CLOSED FRACTURE PROXIMAL FEMUR, BASICERVICAL
S110.00	CLOSED FRACTURE OF CERVICAL SPINE WITH CORD LESION
S251300	OPEN FRACTURE FINGER METACARPAL SHAFT
S310z00	CLOSED FRACTURE OF SHAFT OR UNSPECIFIED PART, NOS
S330900	CLOSED FRACTURE FIBULA, NECK
S261T00	OPEN FRACTURE FINGER DISTAL PHALANX, SHAFT
S4H1.00	OPEN FRACTURE-DISLOCATION, FOOT
S346100	CLOSED FRACTURE ANKLE, TRIMALLEOLAR, HIGH FIBULAR FRACTURE
S344100	CLOSED FRACTURE ANKLE, BIMALLEOLAR, HIGH FIBULAR FRACTURE
S234111	SMITH'S FRACTURE - CLOSED
S260V00	CLOSED FRACTURE FINGER DISTAL PHALANX, MALLET
S220500	CLOSED FRACTURE OF HUMERUS, UPPER EPIPHYSIS
Syu4300	[X]FRACTURE OF OTHER PARTS OF SHOULDER AND UPPER ARM
S134700	CLOSED VERTICAL FRACTURE OF ILIUM
S330800	CLOSED FRACTURE FIBULA, HEAD
S251400	OPEN FRACTURE FINGER METACARPAL NECK
S231200	OPEN FRACTURE OF ULNA, CORONOID
S100311	C3 VERTEBRA CLOSED FRACTURE WITHOUT SPINAL CORD LESION
S250700	CLOSED FRACTURE FINGER METACARPAL, MULTIPLE
Syu5400	[X]FRACTURE OF FOREARM, UNSPECIFIED
Zw01.00	[Q] FRACTURES INVOLVING THE EPIPHYSEAL PLATE
Syu6500	[X]FRACTURE OF OTHER & UNSPECIFIED PARTS OF WRIST AND HAND
S312000	CLOSED FRACTURE OF DISTAL FEMUR, UNSPECIFIED
S100H00	CLOSED FRACTURE CERVICAL VERTEBRA, WEDGE
SR12.00	FRACTURES INVOLVING MULTIPLE REGIONS OF ONE UPPER LIMB
S131.00	OPEN FRACTURE ACETABULUM
S261B00	OPEN FRACTURE THUMB DISTAL PHALANX, TUFT
S221100	OPEN FRACTURE PROXIMAL HUMERUS, NECK
S250C00	CLOSED FRACTURE THUMB METACARPAL HEAD
S224900	CLOSED FRACTURE DISTAL HUMERUS, BICONDYLAR (T-Y FRACTURE)
S221000	OPEN FRACTURE OF PROXIMAL HUMERUS, UNSPECIFIED PART
S234911	CLOSED VOLAR BARTON'S FRACTURE-DISLOCATION
S251500	OPEN FRACTURE FINGER METACARPAL HEAD
S235D00	OPEN FRACTURE DISTAL RADIUS, EXTRA-ARTICULAR OTHER TYPE
Syu4400	[X]FRACTURE OF SHOULDER AND UPPER ARM, UNSPECIFIED
S360300	CLOSED FRACTURE MULTIPLE PHALANGES, TOE
S27z.00	MULTIPLE FRACTURES OF HAND BONES NOS

S330011	CLOSED FRACTURE OF TIBIAL CONDYLES
S101200	OPEN FRACTURE AXIS
S261W00	OPEN FRACTURE FINGER DISTAL PHALANX, MULTIPLE
S331200	OPEN FRACTURE OF TIBIA AND FIBULA, PROXIMAL
S4B1000	OPEN FRACTURE-DISLOCATION ELBOW JOINT
S312600	CLOSED FRACTURE DISTAL FEMUR, BICONDYLAR (T-Y FRACTURE)
S330200	CLOSED FRACTURE OF TIBIA AND FIBULA, PROXIMAL
S100K00	CLOSED FRACTURE CERVICAL VERTEBRA, SPINOUS PROCESS
S12X.00	FRACTURE OF BONY THORAX, PART UNSPECIFIED
S023x00	OPEN FRACTURE OF MANDIBLE, MULTIPLE SITES
S320300	CLOSED FRACTURE PATELLA, VERTICAL
S23y.00	OPEN FRACTURE OF RADIUS AND ULNA, UNSPECIFIED PART
N331700	FRACTURE OF BONE IN NEOPLASTIC DISEASE
S125300	CLOSED FRACTURE OF TRACHEA
S11z.00	FRACTURE OF SPINE WITH SPINAL CORD LESION NOS
S231z00	OPEN FRACTURE OF FOREARM, UPPER END, NOS
S4C0600	CLOSED FRACTURE-DISLOCATION PERI-LUNATE TRANS-SCAPHOID
S109.00	OPEN FRACTURE PELVIS, COCCYX
S312x00	CLOSED FRACTURE DISTAL FEMUR, COMMINUTED/INTRA-ARTICULAR
S100x00	MULTIPLE CLOSED FRACTURES OF CERVICAL VERTEBRAE
S4D2300	CLOSED FRACTURE-SUBLUXATION, DISTAL INTERPHALANGEAL JOINT
S4C2400	CLOSED FRACTURE-SUBLUXATION LUNATE (VOLAR)
S120500	CLOSED FRACTURE OF FIVE RIBS
S332z00	CLOSED FRACTURE OF TIBIA AND FIBULA, SHAFT, NOS
S02x200	FRACTURE OF PALATE, CLOSED
S101.00	OPEN FRACTURE OF CERVICAL SPINE
S211100	OPEN FRACTURE SCAPULA, ACROMION
S241A00	OPEN FRACTURE SCAPHOID, PROXIMAL POLE
S353700	OPEN FRACTURE METATARSAL
S023500	OPEN FRACTURE OF MANDIBLE, ANGLE OF JAW
S241500	OPEN FRACTURE TRAPEZIUM
S120300	CLOSED FRACTURE OF THREE RIBS
S346000	CLOSED FRACTURE ANKLE, TRIMALLEOLAR, LOW FIBULAR FRACTURE
S4H0000	CLOSED FRACTURE-DISLOCATION, SUBTALAR JOINT
S240y00	CLOSED FRACTURE OF OTHER CARPAL BONE
S4G1.00	OPEN FRACTURE-DISLOCATION, ANKLE JOINT
S12..00	FRACTURE OF RIB(S), STERNUM, LARYNX AND TRACHEA
S022700	CLOSED FRACTURE OF MANDIBLE, ALVEOLAR BORDER OF BODY
S4F3.00	OPEN FRACTURE-SUBLUXATION, KNEE JOINT
SR16000	CLOSED FRACTURE INV THORAX WTH LOW BACK AND PELVIS AND LIMBS
S002.00	OPEN FRACTURE VAULT OF SKULL WITHOUT INTRACRANIAL INJURY
S03z.00	SKULL FRACTURE NOS
N338400	ANGULAR MAL-UNION OF FRACTURE
S331z00	OPEN FRACTURE OF TIBIA AND FIBULA, PROXIMAL NOS
S116.00	CLOSED FRACTURE OF SACRUM WITH SPINAL CORD LESION
S210z00	CLOSED FRACTURE OF SCAPULA NOS
S030.00	CLOSED FRACTURE OF SKULL NOS WITHOUT INTRACRANIAL

	INJURY
S234A11	CLOSED DORSAL BARTON'S FRACTURE-DISLOCATION
S130300	CLOSED FRACTURE ACETABULUM, POSTERIOR COLUMN
S350000	CLOSED FRACTURE CALCANEUS, EXTRA-ARTICULAR
S240000	CLOSED FRACTURE OF CARPAL BONE, UNSPECIFIED
S14z.00	FRACTURE OF ILL-DEFINED BONE OF TRUNK NOS
S352500	CLOSED FRACTURE INTERMEDIATE CUNEIFORM
S12X000	CLOSED FRACTURE OF BONY THORAX PART UNSPECIFIED
S30x.00	OPEN FRACTURE OF UNSPECIFIED PROXIMAL FEMUR
S4E1.00	OPEN FRACTURE-DISLOCATION, HIP JOINT
S4B2.00	CLOSED FRACTURE-SUBLUXATION ELBOW
S022800	CLOSED FRACTURE OF MANDIBLE, BODY, OTHER AND UNSPECIFIED
S4C3.00	OPEN FRACTURE-SUBLUXATION OF THE WRIST
S02y.00	OPEN FRACTURE OTHER FACIAL BONE
S022300	CLOSED FRACTURE OF MANDIBLE, CORONOID PROCESS
S4G2.00	CLOSED FRACTURE-SUBLUXATION, ANKLE JOINT
S01..11	ANTERIOR FOSSA FRACTURE
S130y00	OTHER SPECIFIED CLOSED FRACTURE ACETABULUM
S221500	OPEN FRACTURE OF HUMERUS, UPPER EPIPHYSIS
S241C00	OPEN FRACTURE SCAPHOID, WAIST, OBLIQUE
S101711	C7 VERTEBRA OPEN FRACTURE WITHOUT SPINAL CORD LESION
S211300	OPEN FRACTURE SCAPULA, GLENOID
S225200	OPEN FRACTURE DISTAL HUMERUS, LATERAL CONDYLE
S023z00	FRACTURE OF MANDIBLE, OPEN, NOS
S023800	OPEN FRACTURE OF MANDIBLE, BODY, OTHER AND UNSPECIFIED
S4C3000	OPEN FRACTURE-SUBLUXATION, DISTAL RADIO-ULNAR JOINT
S101311	C3 VERTEBRA OPEN FRACTURE WITHOUT SPINAL CORD LESION
S4D2200	CLOSED FRACTURE-SUBLUXATION IPJ, UNSPECIFIED
S233000	OPEN FRACTURE OF RADIUS, SHAFT, UNSPECIFIED
S4A2000	CLOSED FRACTURE-SUBLUXATION SHOULDER JOINT
S100300	CLOSED FRACTURE OF THIRD CERVICAL VERTEBRA
S140.00	CLOSED FRACTURE OF ILL-DEFINED BONE OF TRUNK
S113.00	OPEN FRACTURE OF THORACIC SPINE WITH SPINAL CORD LESION
S235000	OPEN FRACTURE OF FOREARM, LOWER END, UNSPECIFIED
S023100	OPEN FRACTURE OF MANDIBLE, CONDYLAR PROCESS
S4F5.00	OPEN FRACTURE-DISLOCATION, PATELLO-FEMORAL JOINT
S251800	OPEN FRACTURE OF THUMB METACARPAL
S4A3100	OPEN FRACTURE-SUBLUXATION ACROMIO-CLAVICULAR JOINT
S301600	OPEN FRACTURE PROXIMAL FEMUR,SUBCAPITAL, GARDEN GRADE I
S104200	CLOSED FRACTURE LUMBAR VERTEBRA, SPONDYLOLYSIS
S260C00	CLOSED FRACTURE THUMB DISTAL PHALANX, MALLET
S261300	OPEN FRACTURE THUMB PROXIMAL PHALANX
S231500	OPEN FRACTURE OF THE PROXIMAL ULNA
S222z00	CLOSED FRACTURE OF HUMERUS, SHAFT OR UNSPECIFIED PART NOS
S03..00	OTHER AND UNQUALIFIED SKULL FRACTURES
S261H00	OPEN FRACTURE FINGER PROXIMAL PHALANX, HEAD
S361300	OPEN FRACTURE MULTIPLE PHALANGES, TOE
S235900	OPEN VOLAR BARTON'S FRACTURE
S303.00	OPEN FRACTURE OF PROXIMAL FEMUR, PERTROCHANTERIC
S01..17	POSTERIOR FOSSA FRACTURE

S312z00	CLOSED FRACTURE OF DISTAL FEMUR NOT OTHERWISE SPECIFIED
S201.00	OPEN FRACTURE OF CLAVICLE
S4D1100	OPEN FRACTURE-DISLOCATION, METACARPOPHALANGEAL JOINT
S103100	OPEN FRACTURE THORACIC VERTEBRA, WEDGE
S261F00	OPEN FRACTURE FINGER PROXIMAL PHALANX, SHAFT
S110000	CLS SPINAL FRACTURE WITH UNSPEC CERVICAL CORD LESION, C1-4
S131y00	OTHER SPECIFIED OPEN FRACTURE ACETABULUM
S4J2000	CLOSED FRACTURE-SUBLUXATION OF STERNUM
SP04C00	FRACTURE OF BONE ALLOGRAFT
S022.11	FRACTURE OF INFERIOR MAXILLA, CLOSED
S101611	C6 VERTEBRA OPEN FRACTURE WITHOUT SPINAL CORD LESION
S33yz00	OPEN FRACTURE OF TIBIA AND FIBULA, UNSPECIFIED PART, NOS
S260A00	CLOSED FRACTURE THUMB DISTAL PHALANX, SHAFT
S224x00	CLOSED FRACTURE OF DISTAL HUMERUS, MULTIPLE
S251000	OPEN FRACTURE OF METACARPAL BONE(S), SITE UNSPECIFIED
S4B0100	CLOSED FRACTURE-DISLOCATION SUPERIOR RADIO-ULNAR JOINT
S300z00	CLOSED FRACTURE PROXIMAL FEMUR, TRANSCERVICAL, NOS
S010.00	CLOSED FRACTURE BASE OF SKULL WITHOUT INTRACRANIAL INJURY
S4D0600	CLOSED FRACTURE-DISLOCATION MULTIPLE DIGITS
S4D2500	CLOSED FRACTURE-SUBLUXATION, INTERPHALANGEAL JOINT THUMB
S14..00	FRACTURE OF ILL-DEFINED BONES OF TRUNK
S103.00	OPEN FRACTURE THORACIC VERTEBRA
S4C1600	OPEN FRACTURE-DISLOCATION PERI-LUNATE TRANS-SCAPHOID
S235E00	OPEN FRACTURE DISTAL RADIUS, INTRA-ARTICULAR OTHER TYPE
S331600	OPEN FRACTURE SPINE, TIBIA
S011.00	CLOSED FRACTURE BASE OF SKULL WITH INTRACRANIAL INJURY
S4C2600	CLOSED FRACTURE-SUBLUXATION PERI-LUNATE TRANS-SCAPHOID
S225500	OPEN FRACTURE OF DISTAL HUMERUS, TROCHLEA
S231900	OPEN FRACTURE OF THE PROXIMAL RADIUS
S123.00	OPEN FRACTURE STERNUM
S211200	OPEN FRACTURE SCAPULA, CORACOID
S260600	CLOSED FRACTURE THUMB PROXIMAL PHALANX, NECK
S261G00	OPEN FRACTURE FINGER PROXIMAL PHALANX, NECK
S13z.00	OPEN FRACTURE OF PELVIS NOS
S4A1.00	OPEN FRACTURE-DISLOCATION SHOULDER
S100L00	CLOSED FRACTURE CERVICAL VERTEBRA, TRANSVERSE PROCESS
S352G00	CLOSED TARSAL FRACTURES, MULTIPLE
S4C0400	CLOSED FRACTURE-DISLOCATION LUNATE (VOLAR)
S353C00	OPEN FRACTURE METATARSAL SHAFT
S241D00	OPEN FRACTURE SCAPHOID, WAIST, COMMINUTED
S241700	OPEN FRACTURE CAPITATE
S131z00	OPEN FRACTURE ACETABULUM NOS
S102300	CLOSED FRACTURE THORACIC VERTEBRA, SPINOUS PROCESS
S4H2100	CLOSED FRACTURE-SUBLUXATION, MIDTARSAL JOINT

S135.00	OTHER OR MULTIPLE OPEN FRACTURE OF PELVIS
S241.00	OPEN FRACTURE OF CARPAL BONE
S4J0000	CLOSED FRACTURE-DISLOCATION OF STERNUM
S335100	OPEN FRACTURE DISTAL TIBIA, INTRA-ARTICULAR
S4J3100	OPEN FRACTURE-SUBLUXATION OF PELVIS
S101600	OPEN FRACTURE OF SIXTH CERVICAL VERTEBRA
S23yz00	OPEN FRACTURE OF RADIUS AND ULNA, NOS
S105100	OPEN FRACTURE LUMBAR VERTEBRA, WEDGE
N338500	ROTATIONAL MAL-UNION OF FRACTURE
S120600	CLOSED FRACTURE OF SIX RIBS
S4D3400	OPEN FRACTURE-SUBLUXATION, PROXIMAL INTERPHALANGEAL JOINT
S234912	CLOSED VOLAR BARTON FRACTURE-SUBLUXATION
S300200	CLOSED FRACTURE PROXIMAL FEMUR, MIDCERVICAL SECTION
S261A00	OPEN FRACTURE THUMB DISTAL PHALANX, SHAFT
S261900	OPEN FRACTURE THUMB DISTAL PHALANX, BASE
S4D1000	OPEN FRACTURE-DISLOCATION DIGIT, UNSPECIFIED
S4H3300	OPEN FRACTURE-SUBLUXATION, METATARSOPHALANGEAL JOINT, SINGLE
S353F00	OPEN FRACTURE METATARSAL, MULTIPLE
S10y.00	OPEN FRACTURE OF SPINE, UNSPECIFIED,
S353400	OPEN FRACTURE MEDIAL CUNEIFORM
S233100	OPEN FRACTURE OF THE RADIAL SHAFT
S292100	OPEN MULTIPLE FRACTURES OF CLAVICLE, SCAPULA AND HUMERUS
S150100	OPEN MULTIPLE FRACTURE OF THORACIC SPINE
S107.00	OPEN FRACTURE SACRUM
S4D3100	OPEN FRACTURE-SUBLUXATION, METACARPOPHALANGEAL JOINT
S235z00	OPEN FRACTURE OF FOREARM, LOWER END, NOS
S345000	OPEN FRACTURE ANKLE, BIMALLEOLAR, LOW FIBULAR FRACTURE
S281.00	OPEN ILL-DEFINED FRACTURES OF UPPER LIMB
S353.00	OPEN FRACTURE OF OTHER TARSAL AND METATARSAL BONES
S261x00	OPEN FRACTURE OF PHALANX OR PHALANGES, MULTIPLE SITES
S4B2000	CLOSED FRACTURE-SUBLUXATION ELBOW JOINT
S260Q00	CLOSED FRACTURE FINGER MIDDLE PHALANX, MULTIPLE
S353J00	OPEN FRACTURE OF BASE OF FIFTH METATARSAL
S313400	OPEN FRACTURE DISTAL FEMUR, MEDIAL CONDYLE
S100411	C4 VERTEBRA CLOSED FRACTURE WITHOUT SPINAL CORD LESION
S301700	OPEN FRACTURE PROXIMAL FEMUR,SUBCAPITAL, GARDEN GRADE II
S4C0200	CLOSED FRACTURE-DISLOCATION MID CARPAL
S033.00	OPEN FRACTURE OF SKULL NOS WITH INTRACRANIAL INJURY
S135600	OPEN FRACTURE PELVIS, ILIAC WING
S261L00	OPEN FRACTURE FINGER MIDDLE PHALANX, BASE
S4H0200	CLOSED FRACTURE-DISLOCATION, TARSOMETATARSAL JOINT
S100G00	CLOSED FRACTURE CERVICAL VERTEBRA, BURST
S241000	OPEN FRACTURE OF CARPAL BONE, UNSPECIFIED
S300y11	CLOSED FRACTURE OF FEMUR, SUBCAPITAL
S4C3600	OPEN FRACTURE-SUBLUXATION PERI-LUNATE TRANS-SCAPHOID
S4H3400	OPEN FRACTURE-SUBLUXATION, IPJ, SINGLE TOE
S201300	OPEN FRACTURE CLAVICLE, LATERAL END

S4C2300	CLOSED FRACTURE-SUBLUXATION, CARPOMETACARPAL JOINT
S120800	CLOSED FRACTURE OF EIGHT OR MORE RIBS
S023400	OPEN FRACTURE OF MANDIBLE, RAMUS, UNSPECIFIED
S301y00	OPEN FRACTURE PROXIMAL FEMUR, OTHER TRANSCERVICAL
S135300	OPEN FRACTURE PELVIS, ISCHIAL TUBEROSITY
S241200	OPEN FRACTURE LUNATE
S29..13	MULTIPLE FRACTURES OF STERNUM
S025z00	FRACTURE OF MALAR OR MAXILLARY BONES, OPEN, NOS
S02C.00	LE FORT III FRACTURE MAXILLA
S101111	C1 VERTEBRA OPEN FRACTURE WITHOUT SPINAL CORD LESION
S241800	OPEN FRACTURE HAMATE
S225300	OPEN FRACTURE DISTAL HUMERUS, MEDIAL CONDYLE
S261N00	OPEN FRACTURE FINGER MIDDLE PHALANX, NECK
S130400	CLOSED FRACTURE ACETABULUM, FLOOR
S111.00	OPEN FRACTURE OF CERVICAL SPINE WITH SPINAL CORD LESION
SD92000	FRACTURE BLISTER
SP04A00	FRACTURE OF BONE AUTOGRAFT
S101100	OPEN FRACTURE ATLAS
S361100	OPEN FRACTURE MIDDLE PHALANX, TOE
S261J00	OPEN FRACTURE FINGER PROXIMAL PHALANX, MULTIPLE
S012.00	OPEN FRACTURE BASE SKULL WITHOUT MENTION INTRACRANIAL INJURY
Zw02400	[Q] STRESS FRACTURE
SR13.00	FRACTURES INVOLVING MULTIPLE REGIONS OF ONE LOWER LIMB
S3x1.00	OTHER, MULTIPLE AND ILL-DEFINED OPEN FRACTURES OF LOWER LIMB
S300100	CLOSED FRACTURE PROXIMAL FEMUR, TRANSEPIPHYSEAL
S100900	CLOSED FRACTURE ATLAS, COMMINUTED
S351100	OPEN FRACTURES CALCANEUS, INTRA-ARTICULAR
S025.11	FRACTURE OF UPPER JAW, OPEN
S303z00	OPEN FRACTURE OF PROXIMAL FEMUR, PERTROCHANTERIC, NOS
S221400	OPEN FRACTURE PROXIMAL HUMERUS, HEAD
S233200	OPEN FRACTURE OF THE ULNAR SHAFT
S23y000	OPEN FRACTURE OF FOREARM, UNSPECIFIED
S221600	OPEN FRACTURE PROXIMAL HUMERUS, THREE PART
S221700	OPEN FRACTURE PROXIMAL HUMERUS, FOUR PART
S023000	OPEN FRACTURE MANDIBLE (SITE UNSPECIFIED)
S133z00	OPEN FRACTURE OF PUBIS NOS
S201000	OPEN FRACTURE OF CLAVICLE, UNSPECIFIED PART
S4H0100	CLOSED FRACTURE-DISLOCATION, MIDTARSAL JOINT
S221200	OPEN FRACTURE OF PROXIMAL HUMERUS, ANATOMICAL NECK
S303200	OPEN FRACTURE PROXIMAL FEMUR, SUBTROCHANTERIC
S121900	OPEN FRACTURE MULTIPLE RIBS
S125.00	CLOSED FRACTURE LARYNX AND TRACHEA
S022600	CLOSED FRACTURE OF MANDIBLE, SYMPHYSIS OF BODY
S4J2.00	OTHER CLOSED FRACTURE-SUBLUXATION
Syu6400	[X]FRACTURE OF OTHER METACARPAL BONE
S211600	OPEN FRACTURE SCAPULA, NECK
S4H2000	CLOSED FRACTURE-SUBLUXATION, SUBTALAR JOINT
S301100	OPEN FRACTURE PROXIMAL FEMUR, TRANSEPIPHYSEAL
S11..11	FRACTURE OF TRANSVERSE PROCESS OF SPINE + SPINAL CORD LESION

S106000	CLOSED COMPRESSION FRACTURE SACRUM
S251700	OPEN FRACTURE FINGER METACARPAL, MULTIPLE
S231800	OPEN FRACTURE PROXIMAL RADIUS, COMMINUTED
S130000	CLOSED FRACTURE ACETABULUM, ANTERIOR LIP ALONE
S353B00	OPEN FRACTURE METATARSAL BASE
S106100	CLOSED VERTICAL FRACTURE OF SACRUM
S101x00	MULTIPLE OPEN FRACTURES OF CERVICAL VERTEBRAE
S110600	CLS SPINAL FRACTURE WITH UNSPEC CERVICAL CORD LESION, C5-7
S4H2600	CLOSED FRACTURE-SUBLUXATION, IPJ, MULTIPLE TOES
S343100	OPEN FRACTURE ANKLE, LATERAL MALLEOLUS, HIGH
S211400	OPEN FRACTURE SCAPULA, BLADE
Syu7200	[X]FRACTURES OF OTHER PARTS OF FEMUR
S251200	OPEN FRACTURE FINGER METACARPAL BASE
S240F00	CLOSED FRACTURE CARPAL BONES, MULTIPLE
S313x00	OPEN FRACTURE DISTAL FEMUR, COMMINUTED/INTRA-ARTICULAR
S301400	OPEN FRACTURE HEAD, FEMUR
S301y11	OPEN FRACTURE OF FEMUR, SUBCAPITAL
SR1z000	[X]CLOSED MULTIPLE FRACTURES UNSPECIFIED
S110z00	CLOSED FRACTURE OF CERVICAL SPINE WITH CORD LESION NOS
S223z00	OPEN FRACTURE OF HUMERUS, SHAFT OR UNSPECIFIED PART NOS
S130200	CLOSED FRACTURE ACETABULUM, ANTERIOR COLUMN
S105000	OPEN FRACTURE LUMBAR VERTEBRA, BURST
S112z00	CLOSED FRACTURE OF THORACIC SPINE WITH CORD LESION NOS
S121000	OPEN FRACTURE OF RIB, UNSPECIFIED
S4H2200	CLOSED FRACTURE-SUBLUXATION, TARSOMETATARSAL JOINT
S211000	OPEN FRACTURE OF SCAPULA, UNSPECIFIED PART
SR10000	CLOSED FRACTURES INVOLVING HEAD WITH NECK
S114500	CLOSED SPINAL FRACTURE WITH CAUDA EQUINA LESION
S241B00	OPEN FRACTURE SCAPHOID, WAIST, TRANSVERSE
S121700	OPEN FRACTURE OF SEVEN RIBS
S301.00	OPEN FRACTURE PROXIMAL FEMUR, TRANSCERVICAL
S4D3.00	OPEN FRACTURE-SUBLUXATION DIGIT
S33A.00	FRACTURE OF TIBIA
S4B3.00	OPEN FRACTURE-SUBLUXATION ELBOW
S225400	OPEN FRACTURE OF DISTAL HUMERUS, CONDYLE(S) UNSPECIFIED
S313z00	OPEN FRACTURE OF DISTAL FEMUR NOT OTHERWISE SPECIFIED
S120700	CLOSED FRACTURE OF SEVEN RIBS
S4H1200	OPEN FRACTURE-DISLOCATION, TARSOMETATARSAL JOINT
S4H0600	CLOSED FRACTURE-DISLOCATION, IPJ, MULTIPLE TOES
S4D3500	OPEN FRACTURE-SUBLUXATION, INTERPHALANGEAL JOINT THUMB
SR12000	CLOSED FRACTURES INVOLVING MULTIPLE REGIONS OF ONE UPP LIMB
S331011	OPEN FRACTURE OF TIBIAL CONDYLES
S4E2.00	CLOSED FRACTURE-SUBLUXATION, HIP JOINT
N331N00	FRAGILITY FRACTURE
S353900	OPEN FRACTURE TALUS, NECK
S4H3.00	OPEN FRACTURE-SUBLUXATION, FOOT
N331M11	MINIMAL TRAUMA FRACTURE DUE TO UNSPECIFIED

	OSTEOPOROSIS
N331N11	MINIMAL TRAUMA FRACTURE
S261400	OPEN FRACTURE THUMB PROXIMAL PHALANX, BASE
S126300	OPEN FRACTURE OF TRACHEA
S133y00	OTHER SPECIFIED OPEN FRACTURE OF PUBIS
S115.00	OPEN FRACTURE OF LUMBAR SPINE WITH SPINAL CORD LESION
S4J1000	OPEN FRACTURE-DISLOCATION OF STERNUM
S4D1200	OPEN FRACTURE-DISLOCATION IPJ, UNSPECIFIED
S100D00	CLOSED FRACTURE AXIS, TRANSVERSE PROCESS
S311z00	OPEN FRACTURE OF FEMUR, SHAFT OR UNSPECIFIED PART, NOS
S251x00	OPEN FRACTURES OF MULTIPLE SITES OF UNSPECIFIED METACARPUS
S261500	OPEN FRACTURE THUMB PROXIMAL PHALANX, SHAFT
S211z00	OPEN FRACTURE OF SCAPULA NOS
S201z00	OPEN FRACTURE OF CLAVICLE NOS
Zw02500	[Q] REFRACTURE
S117300	OPEN FRACTURE OF SACRUM WITH OTHER SPINAL CORD INJURY
S4J3000	OPEN FRACTURE-SUBLUXATION OF STERNUM
Syu1500	[X]FRACTURE OF OTHER SPECIFIED CERVICAL VERTEBRA
S130100	CLOSED FRACTURE ACETABULUM, POSTERIOR LIP ALONE
S135800	OPEN FRACTURE DISLOCATION OF SACRO-ILIAC JOINT
S4C3300	OPEN FRACTURE-SUBLUXATION, CARPOMETACARPAL JOINT
Syu0400	[X]FRACTURE OF SKULL AND FACIAL BONES, PART UNSPECIFIED
S101A00	OPEN FRACTURE AXIS, ODONTOID PROCESS
S100E00	CLOSED FRACTURE AXIS, POSTERIOR ARCH
Zw02.00	[Q] FRACTURE TYPE QUALIFYING TERMS
S100M00	CLOSED FRACTURE CERVICAL VERTEBRA, POSTERIOR ARCH
S114100	CLOSED SPINAL FRACTURE WITH COMPLETE LUMBAR CORD LESION
S104600	CLOSED FRACTURE LUMBAR VERTEBRA, TRICOLUMNAR
Zw02E00	[Q] OPEN FRACTURE GRADE 2
S100J00	CLOSED FRACTURE CERVICAL VERTEBRA, SPONDYLOLYSIS
S294000	CL FRACTURES INVOLVING MULTIPLE REGIONS OF BOTH UPPER LIMBS
N1y2.00	PARS INTERARTICULARIS STRESS FRACTURE
S4H2400	CLOSED FRACTURE-SUBLUXATION, IPJ, SINGLE TOE
S12y000	CLOSED FRACTURE OF OTHER PARTS OF BONY THORAX
S104500	CLOSED FRACTURE LUMBAR VERTEBRA, POSTERIOR ARCH
S4C0500	CLOSED FRACTURE-DISLOCATION PERI-LUNATE (DORSAL)
S4D3600	OPEN FRACTURE-SUBLUXATION MULTIPLE DIGITS
SR1z100	[X]OPEN MULTIPLE FRACTURES UNSPECIFIED
S117.00	OPEN FRACTURE OF SACRUM WITH SPINAL CORD LESION
S110800	CLS SPINAL FRACTURE WITH ANTERIOR CERVCL CORD LESION, C5-7
S301A00	OPEN FRACTURE OF FEMUR, UPPER EPIPHYSIS
S126100	OPEN FRACTURE OF HYOID BONE
S303011	OPEN FRACTURE OF FEMUR, GREATER TROCHANTER
S102200	CLOSED FRACTURE THORACIC VERTEBRA, SPONDYLOLYSIS
S235400	OPEN FRACTURE OF ULNA, LOWER EPIPHYSIS
Syu8300	[X]FRACTURES OF OTHER PARTS OF LOWER LEG
S135000	OPEN FRACTURE OF ILIUM, UNSPECIFIED
S4D1600	OPEN FRACTURE-DISLOCATION MULTIPLE DIGITS
S013.00	OPEN FRACTURE BASE OF SKULL WITH INTRACRANIAL INJURY

S101211	C2 VERTEBRA OPEN FRACTURE WITHOUT SPINAL CORD LESION
S353500	OPEN FRACTURE INTERMEDIATE CUNEIFORM
Syu4200	[X]MULTIPLE FRACTURES OF CLAVICLE, SCAPULA AND HUMERUS
S107100	OPEN VERTICAL FRACTURE OF SACRUM
Syu9400	[X]FRACTURE OF OTHER TARSAL BONES
S353A00	OPEN FRACTURE TALUS, BODY
S4C1300	OPEN FRACTURE-DISLOCATION CARPOMETACARPAL JOINT
S353D00	OPEN FRACTURE METATARSAL NECK
S225900	OPEN FRACTURE DISTAL HUMERUS, BICONDYLAR (T-Y FRACTURE)

Weight loss

Read code	**Read term**
22A6.00	O/E - UNDERWEIGHT
R032.00	[D]ABNORMAL LOSS OF WEIGHT
1625	ABNORMAL WEIGHT LOSS
1625.11	ABNORMAL WEIGHT LOSS - SYMPTOM
1D1A.00	COMPLAINING OF WEIGHT LOSS
R034800	[D]UNDERWEIGHT
2224	O/E - CACHEXIC
R2y4.00	[D] CACHEXIA
22A1.00	O/E - WEIGHT > 20% BELOW IDEAL
22A2.00	O/E -WEIGHT 10-20% BELOW IDEAL
22K3.00	BODY MASS INDEX LOW K/M2
22A8.00	WEIGHT LOSS FROM BASELINE WEIGHT
2287	ABNORMALLY THIN
R2y4z00	[D]CACHEXIA NOS

APPENDIX 9 - The rule-out value of individual and combinations of investigations for myeloma (FBC instead of haemoglobin).

Investigations	N	TP	FP	FN	TN	Sensitivity (%)	Specificity (%)	LR-
Individual								
ESR	2192	844	546	154	648	85 (82-87)	54 (51-57)	0.28 (0.24-0.33)
PV	433	162	95	39	137	81 (74-86)	59 (52-64)	0.32 (0.24-0.44)
FBC	5934	1309	1109	530	2986	71 (69-73)	73 (72-74)	0.40 (0.37-0.43)
CAL	2809	246	43	893	1627	22 (19-24)	97 (97-0.98)	0.81 (0.78-0.83)
Creatinine	7407	648	1021	1202	4536	35 (33-37)	82 (81-83)	0.80 (0.77-0.83)
CRP	1274	316	281	319	358	46 (42-50)	63 (59-67)	0.87 (0.78-0.94)
Combinations (2 tests)								
FBC + PV	396	176	126	9	85	95 (91-98)	40 (34-47)	0.12 (0.06-0.23)
FBC + ESR	2071	890	638	66	477	93 (91-95)	43 (40-46)	0.16 (0.13-0.21)
FBC + CRP	1193	503	332	99	259	84 (80-86)	44 (40-48)	0.38 (0.31-0.46)
Combinations (3 tests)								
FBC + CREAT + PV	356	165	116	6	69	97 (93-99)	37 (30-45)	0.09 (0.04-0.21)
FBC + CAL + PV	213	128	52	3	30	98 (94-100)	37 (26-48)	0.06(0.02-0.20)
FBC + CREAT + ESR	1864	827	623	50	364	94 (93-96)	37 (34-40)	0.16 (0.12-0.21)
FBC + CAL + ESR	1146	593	301	42	210	93 (91-95)	41 (37-46)	0.16 (0.12-0.22)
FBC + CREAT + CRP	1115	510	356	65	184	89 (86-91)	34 (30-38)	0.33 (0.26-0.43)
FBC + CAL + CRP	722	368	172	59	123	86 (83-89)	42 (36-48)	0.33 (0.25-0.44)
Combinations (4 tests)								
FBC + CAL + CREAT + PV	211	127	53	3	28	98 (93-100)	35 (24.3-46)	0.07 (0.02-0.21)
FBC + CAL + CREAT+ ESR	1103	571	321	37	174	94 (92-96)	35 (31-40)	0.17 (0..12-0.24)
FBC + CAL + CREAT + CRP	706	376	197	41	92	90 (87-93)	32 (27-38)	0.31 (0.22-0.43)

APPENDIX 10 - Pre/Post - test probabilities

Illustration of how post-test probabilities are affected assuming a pre-test probability of the disease of 0.03% and different values of likelihood ratios. The Negative predictive value corresponds to 1-(post-test probability) for LR- and the positive predictive value corresponds to post-test probability for LR+

$p = pretest\ probability\ (prevalence)$

$$pretest\ odds = \frac{p}{1-p}$$

$$posttest\ odds = pretest\ odds\ \times LR$$

$$posttest\ probability = \frac{posttest\ odds}{1 + posttest\ odds}$$

Prevalence = 0.03%			
LR-	Post-test probability (%)	LR+	Post-test probability (%)
0.9	0.027	2	0.06
0.8	0.024	3	0.09
0.7	0.021	4	0.12
0.6	0.018	5	0.15
0.5	0.015	6	0.18
0.4	0.012	7	0.21
0.3	0.009	8	0.24
0.2	0.006	9	0.27
0.1	0.003	10	0.30

Another way is to use the formulae illustrated below to convert to PPV or NP given sensitivity, specificity and prevalence of condition

$$PPV = \frac{sensitivity \times p}{(sensitivity\ \times p) + (1 - specificity) \times (1-p)}$$

$$NPV = \frac{specificity \times (1-p)}{(specificity)\ \times (1-p) + (1 - sensitivity) \times p}$$

APPENDIX 11 – ISAC protocol approval

ISAC EVALUATION OF PROTOCOLS FOR RESEARCH INVOLVING CPRD DATA

FEEDBACK TO APPLICANTS

CONFIDENTIAL			*by e-mail*
PROTOCOL NO:	17_088R		
PROTOCOL TITLE:	Development and validation of clinical prediction rules to help with decision making and early detection of multiple myeloma		
APPLICANT:	Dr. Jason Oke, Senior Statistician University of Oxford, Nuffield department of primary care health sciences Radcliffe Observatory Quarter, Woodstock Road, Oxford. OX2 6GG jason.oke@phc.ox.ac.uk		
APPROVED ☒	**APPROVED WITH COMMENTS (resubmission not required)** ☐	**REVISION/ RESUBMISSION REQUESTED** ☐	**REJECTED** ☐
INSTRUCTIONS: ***Protocols with an outcome of 'Approved' or 'Approved with comments' do not require resubmission to the ISAC.*** **APPLICANT FEEDBACK:** **Protocol 17_088R is approved.**			
DATE OF ISAC FEEDBACK:	06/06/2017		
DATE OF APPLICANT FEEDBACK:			

For protocols approved from 01 April 2014 onwards, applicants are required to include the ISAC protocol in their journal submission with a statement in the manuscript indicating that it had been approved by the ISAC (with the reference number) and made available to the journal reviewers. If the protocol was subject to any amendments, the last amended version should be the one submitted.

***** Please refer to the ISAC advice about protocol amendments provided below*****

APPENDIX 12 - Missing data patterns for blood tests

Full blood count was not included as the selection of the population was based on it. The highlighted row corresponds to the patients that had full data. Table for PV is not shown as only 4% of the patients had that test. Y corresponds to variable recorded and N to variable not recorded

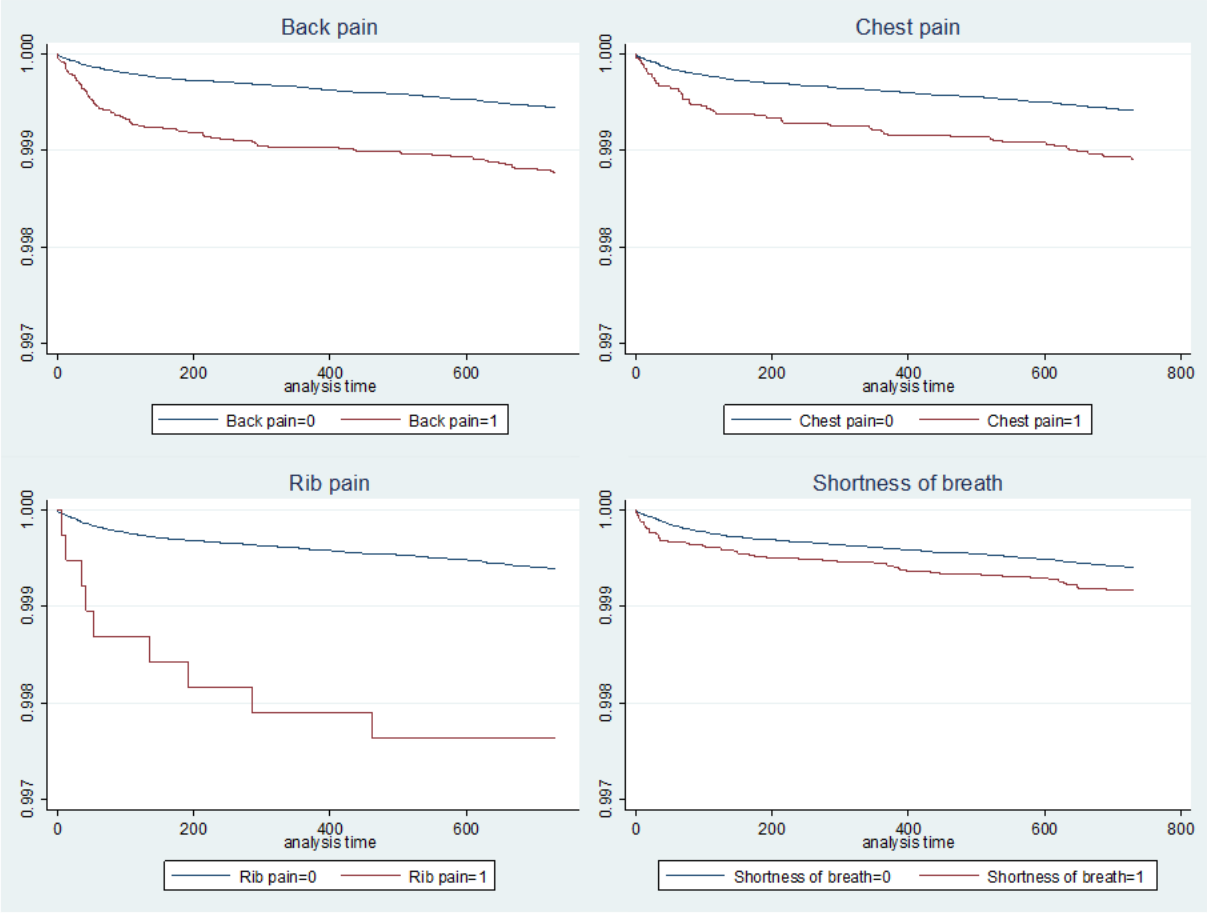
Percent	Creatinine	ESR	Calcium
36%	N	N	N
22%	Y	N	N
10%	Y	N	Y
10%	N	Y	N
8%	Y	Y	N
6%	N	N	Y
6%	Y	Y	Y
3%	N	Y	Y

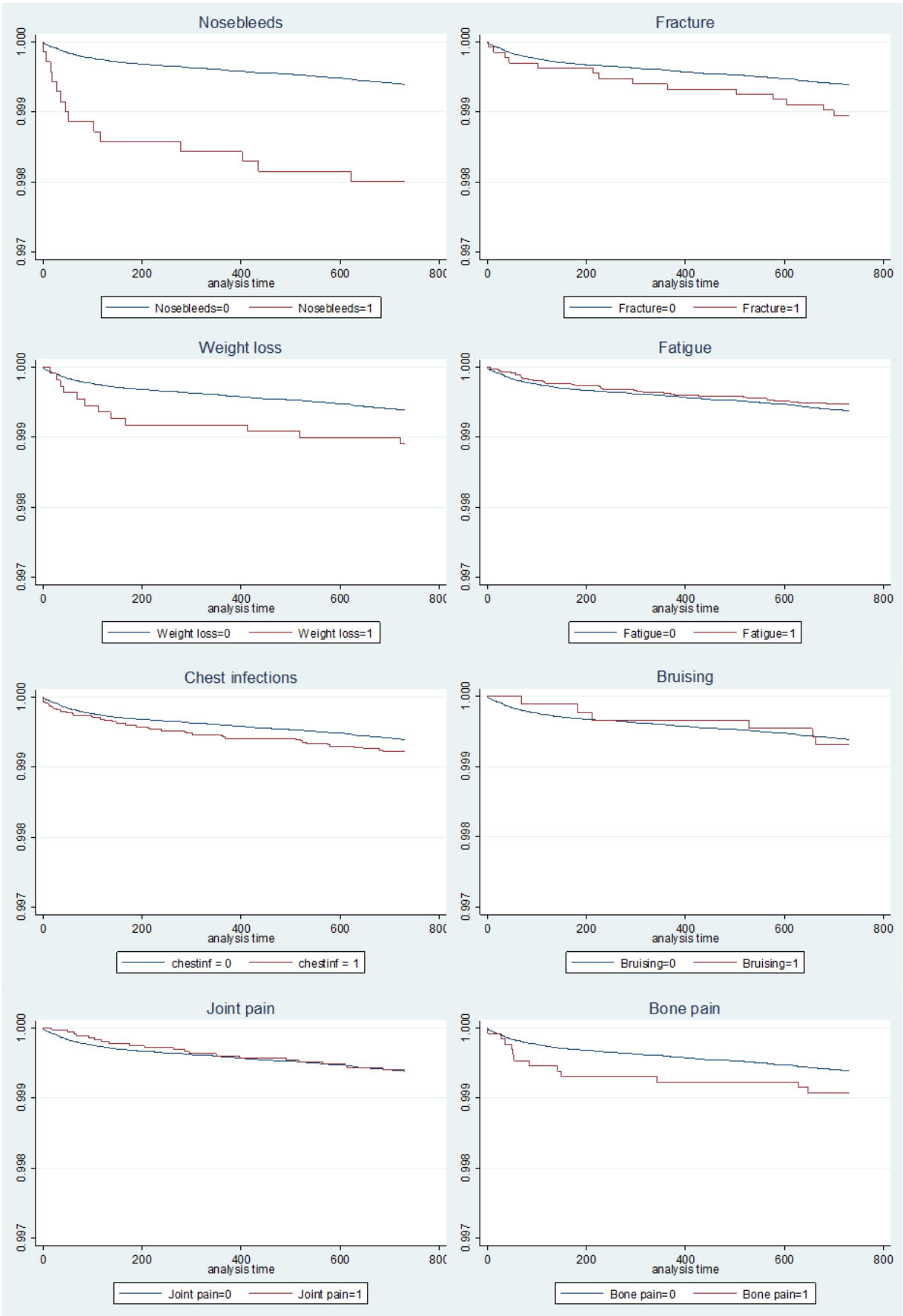
Percent	Creatinine	CRP	Calcium
40%	N	N	N
24%	N	N	N
11%	Y	Y	N
6%	N	Y	N
6%	N	N	Y
6%	Y	N	Y
5%	Y	Y	Y
2%	N	Y	Y

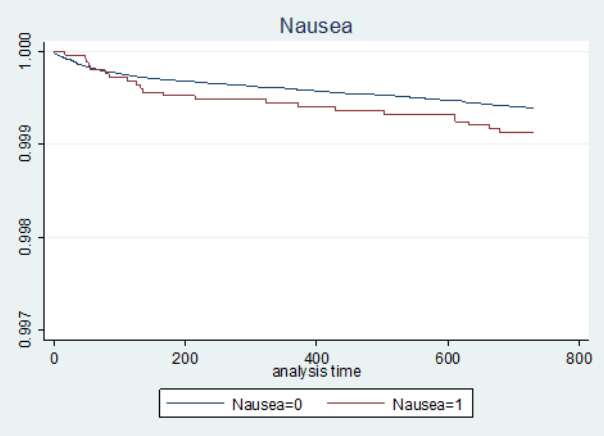
Percent	Creatinine	Any IM	Calcium
32%	N	N	N
19%	Y	N	N
14%	N	Y	N
11%	Y	Y	N
8%	Y	Y	Y
7%	Y	N	Y
5%	N	N	Y
4%	N	Y	Y

APPENDIX 13 - Kaplan Meir plots / proportional hazards assumptions examination for binary predictors

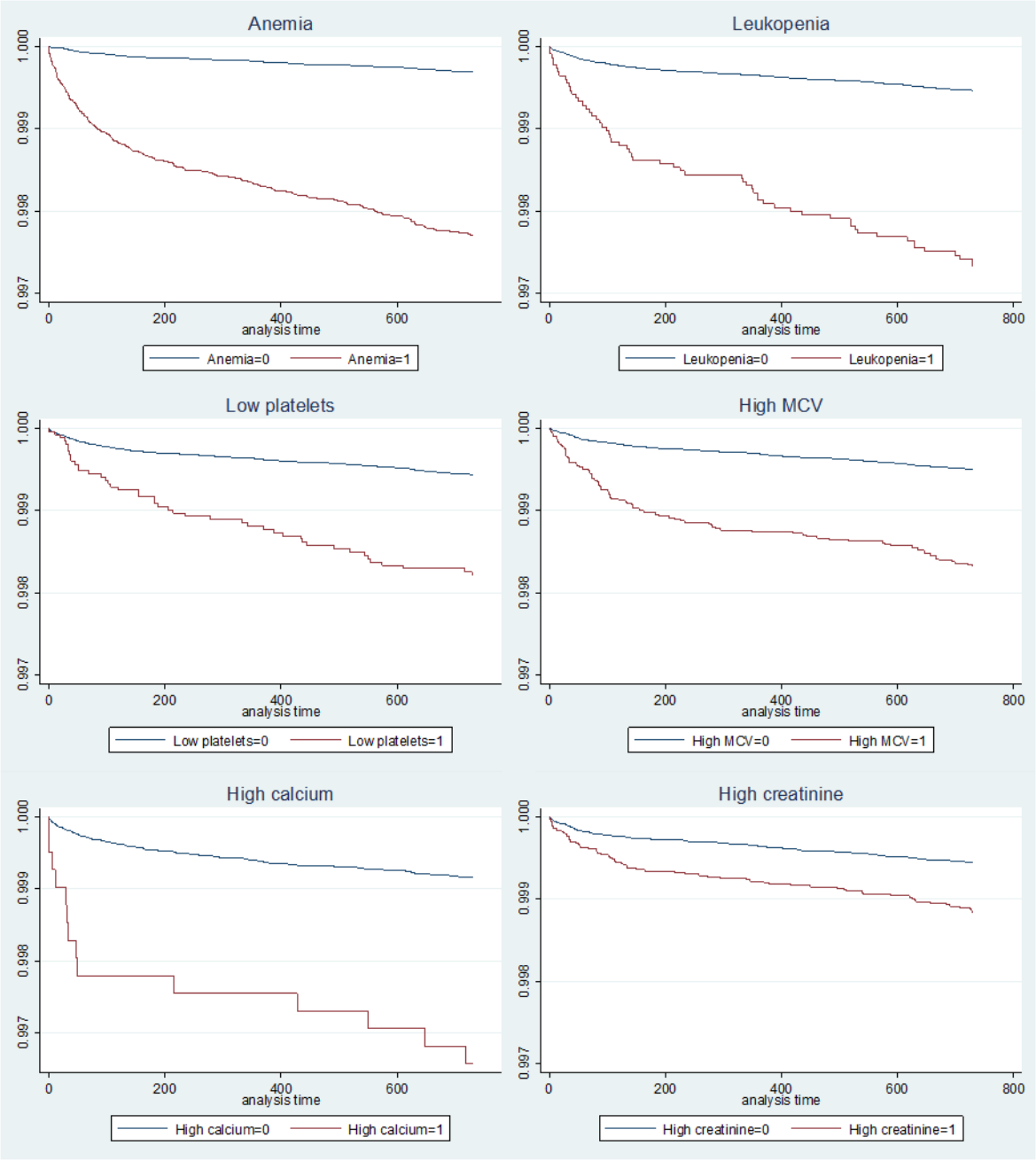
Symptoms



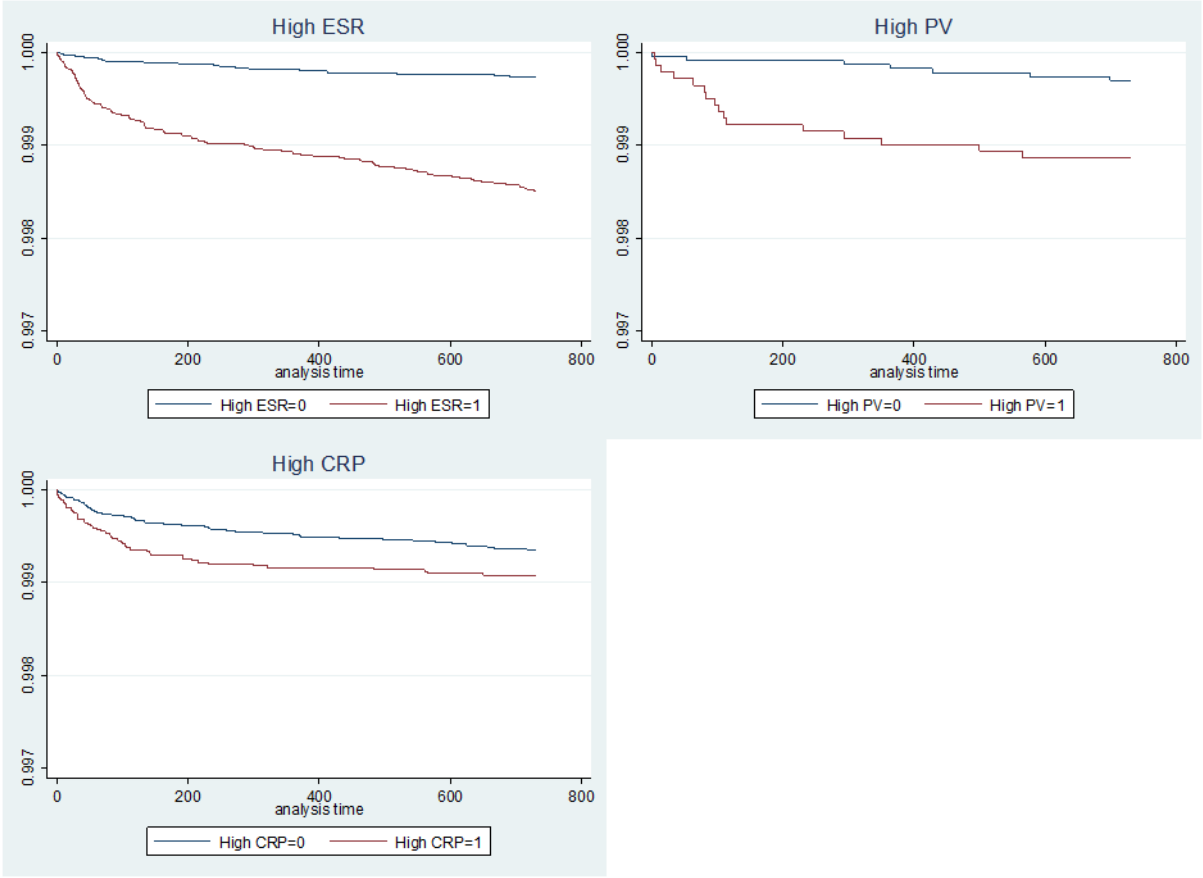

Nosebleeds
1.000
0.999
0.998
0.997
0
200
400
600
800
analysis time
Nosebleeds=0
Nosebleeds=1
Fracture
Fracture=0
Fracture=1
Weight loss
Weight loss=0
Weight loss=1
Fatigue
Fatigue=0
Fatigue=1
Chest infections
chestinf = 0
chestinf = 1
Bruising
Bruising=0
Bruising=1
Joint pain
Joint pain=0
Joint pain=1
Bone pain
Bone pain=0
Bone pain=1



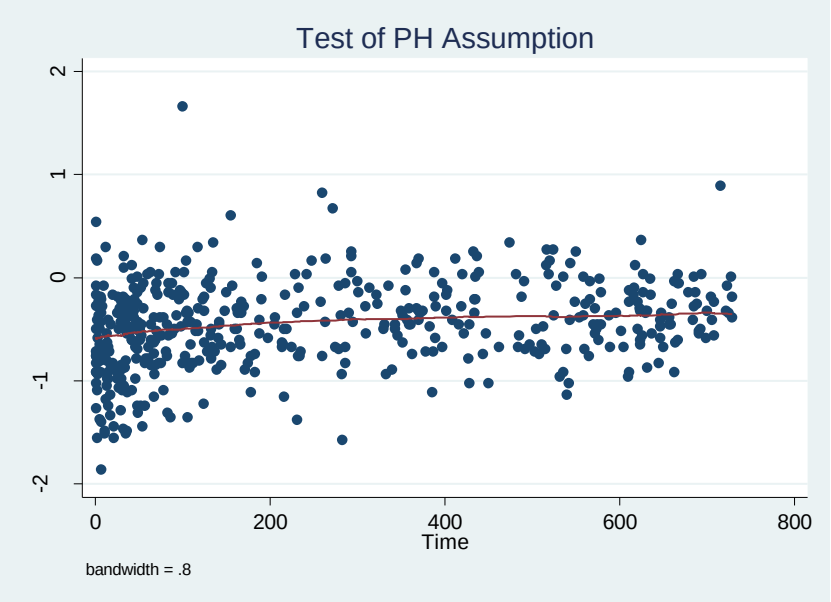
Blood tests categorised as normal/abnormal



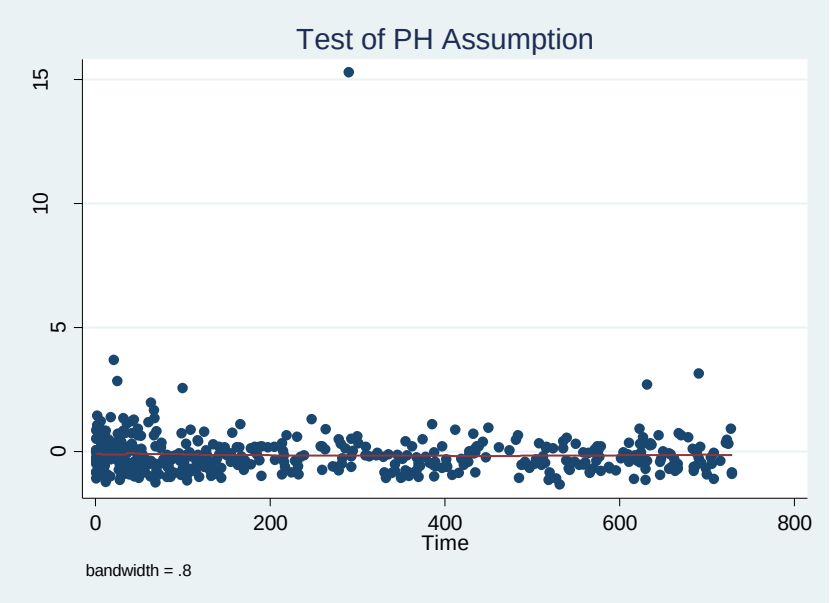
Inflammatory markers



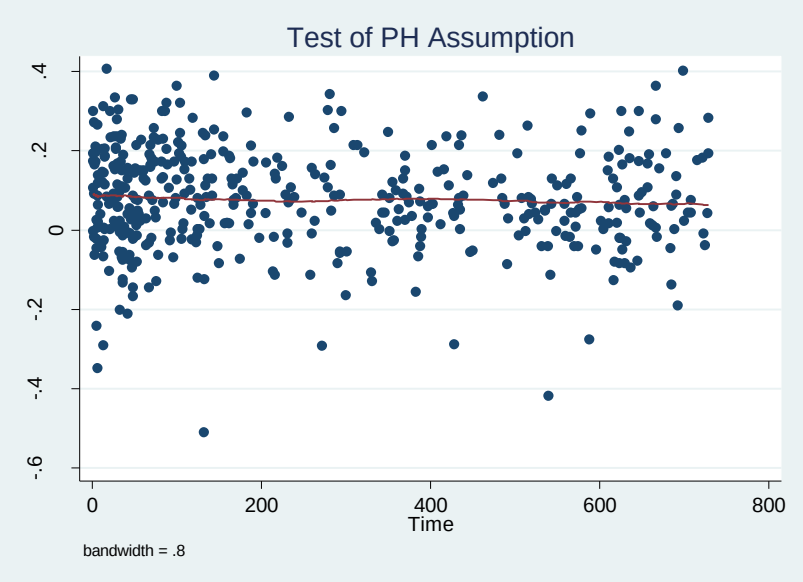
Haemoglobin



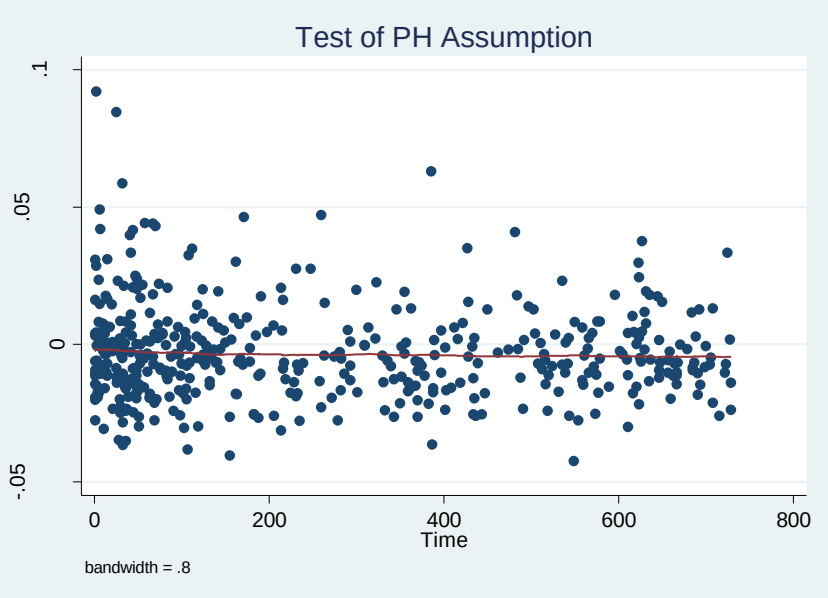
White cell count



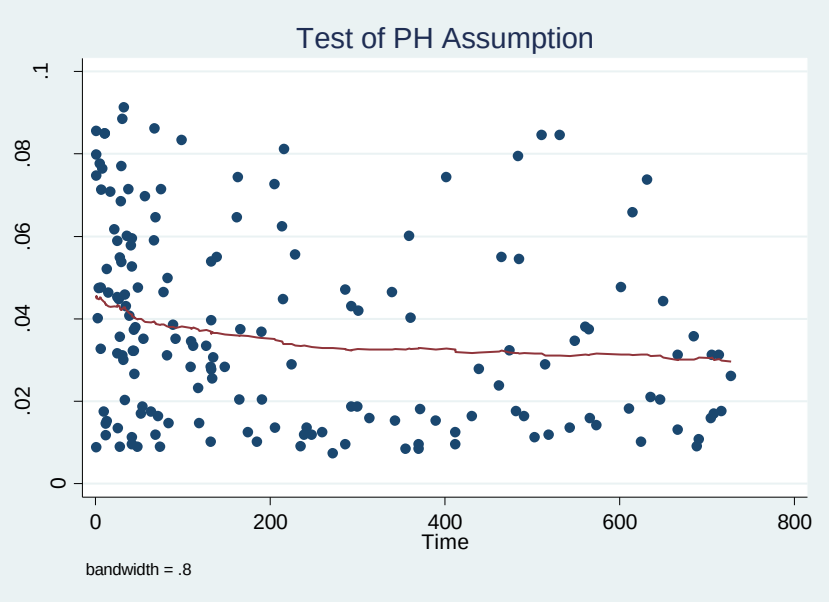
Mean corpuscular volume (MCV)



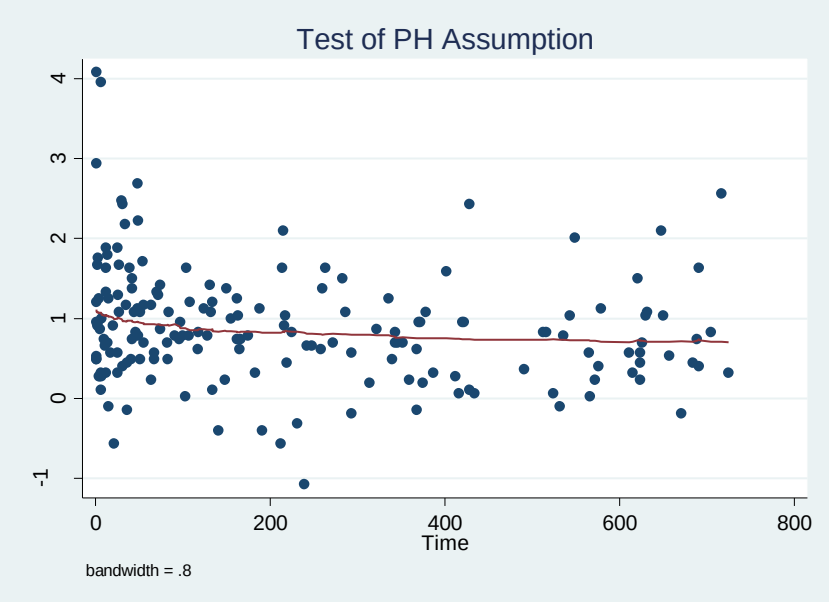
Platelets



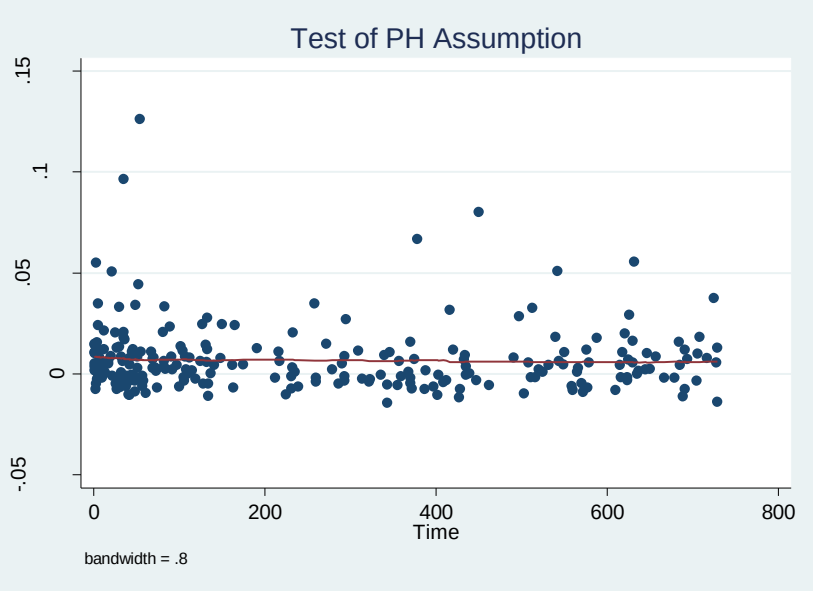
Erythrocyte sedimentation rate (ESR)



Calcium

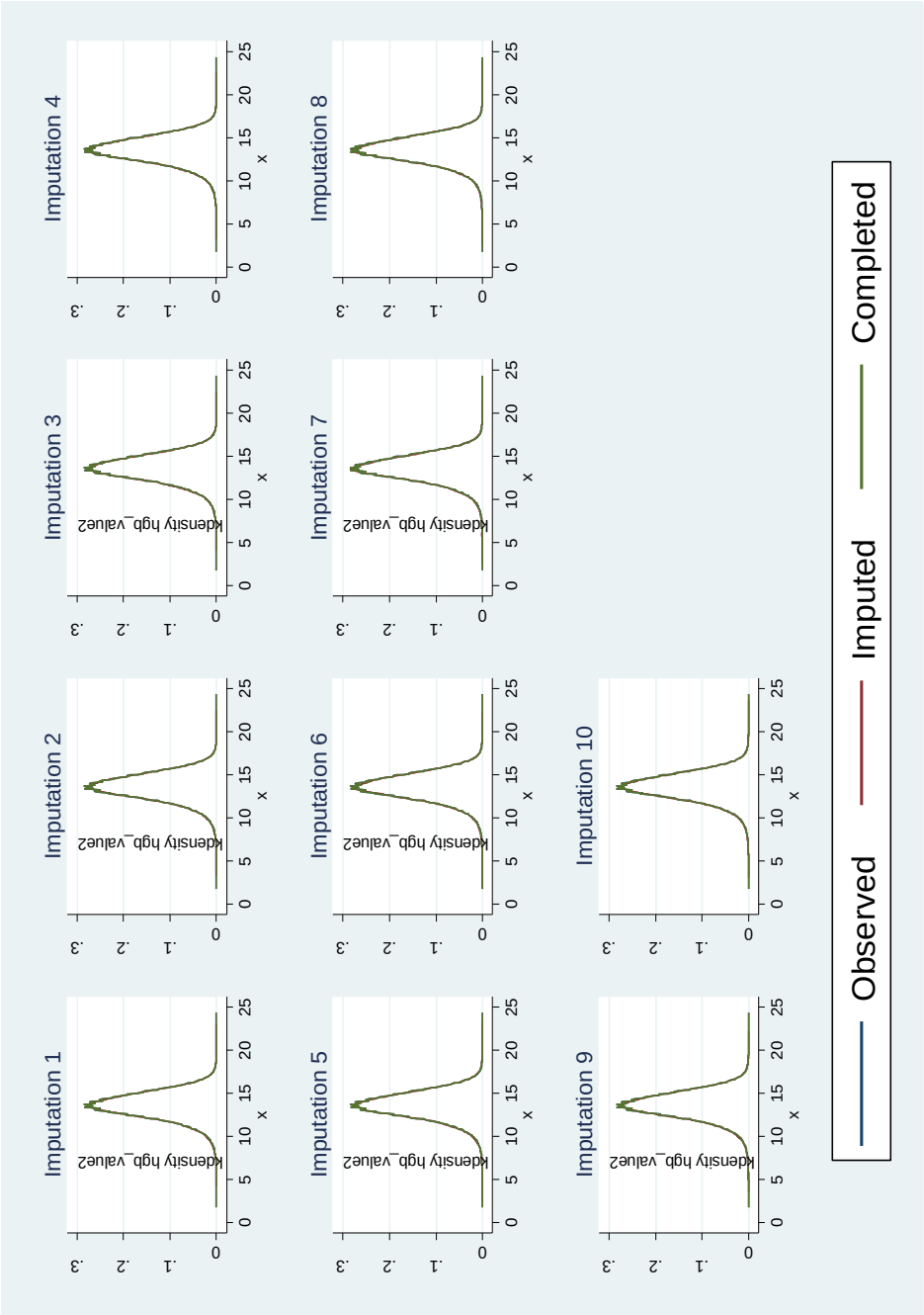


Creatinine

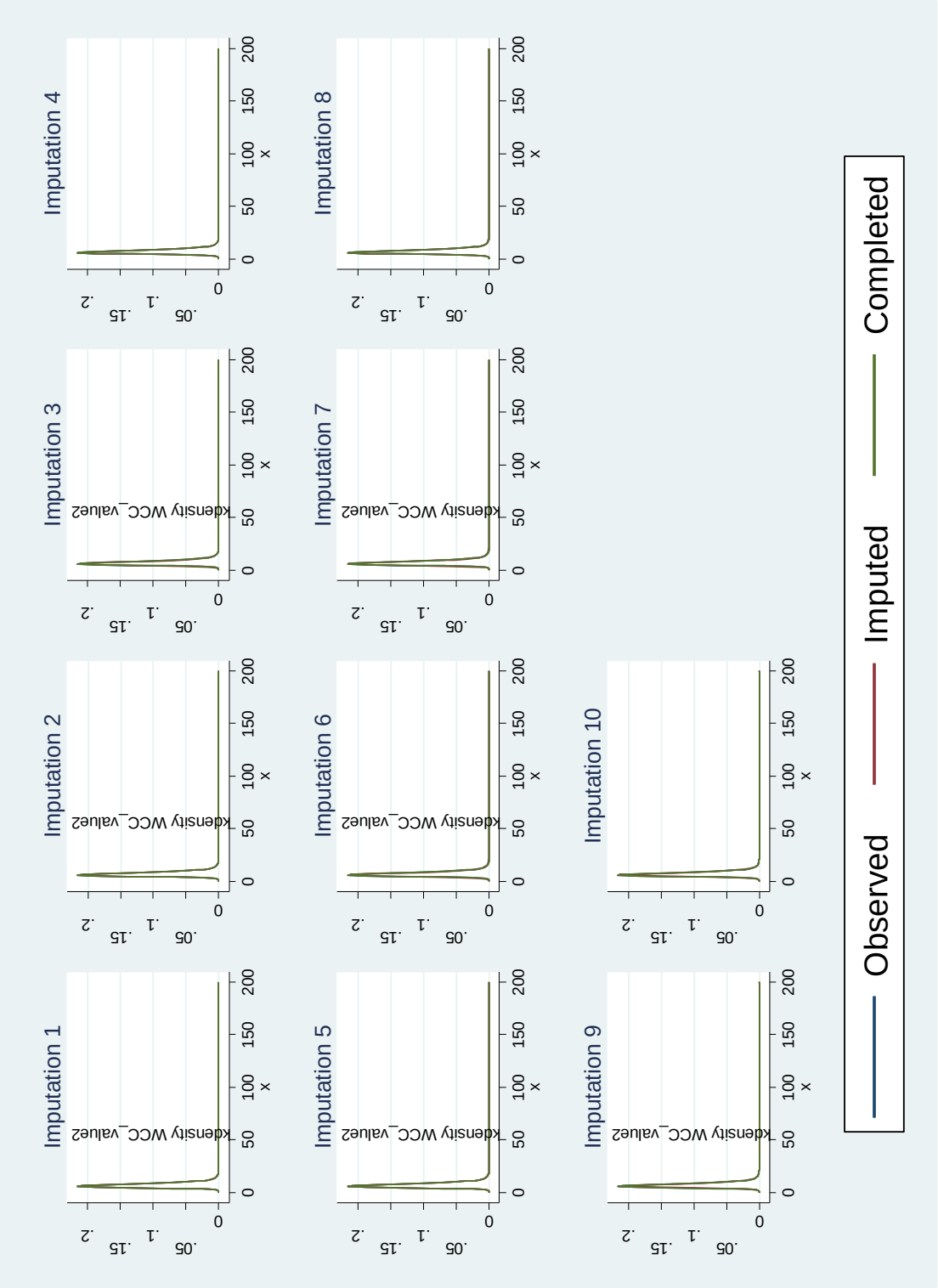

Test of PH Assumption
.15
.1
.05
0
-.05
0
200
400
600
800
Time
bandwidth = .8

APPENDIX 14 - Imputation diagnostics

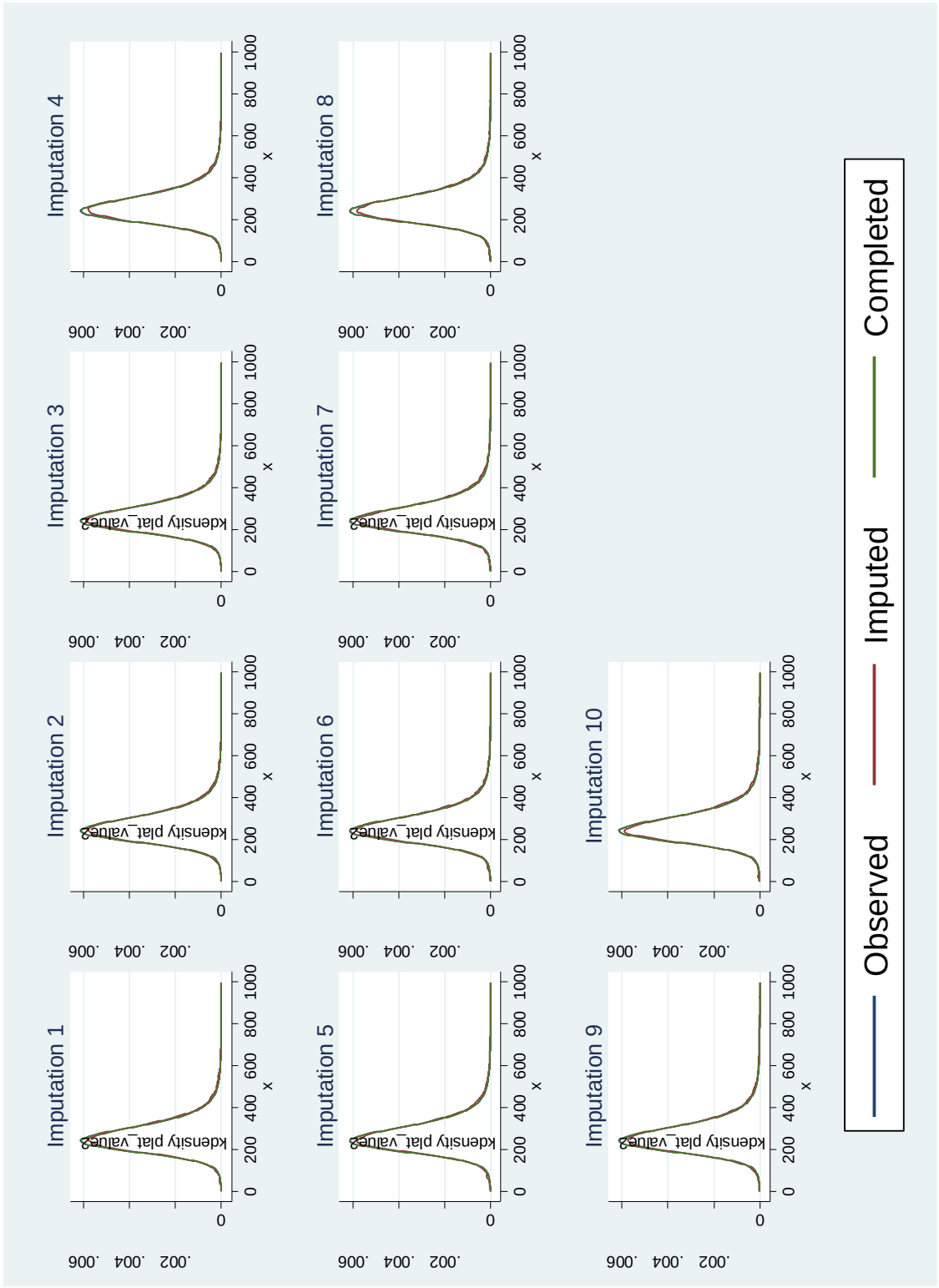
Haemoglobin – index



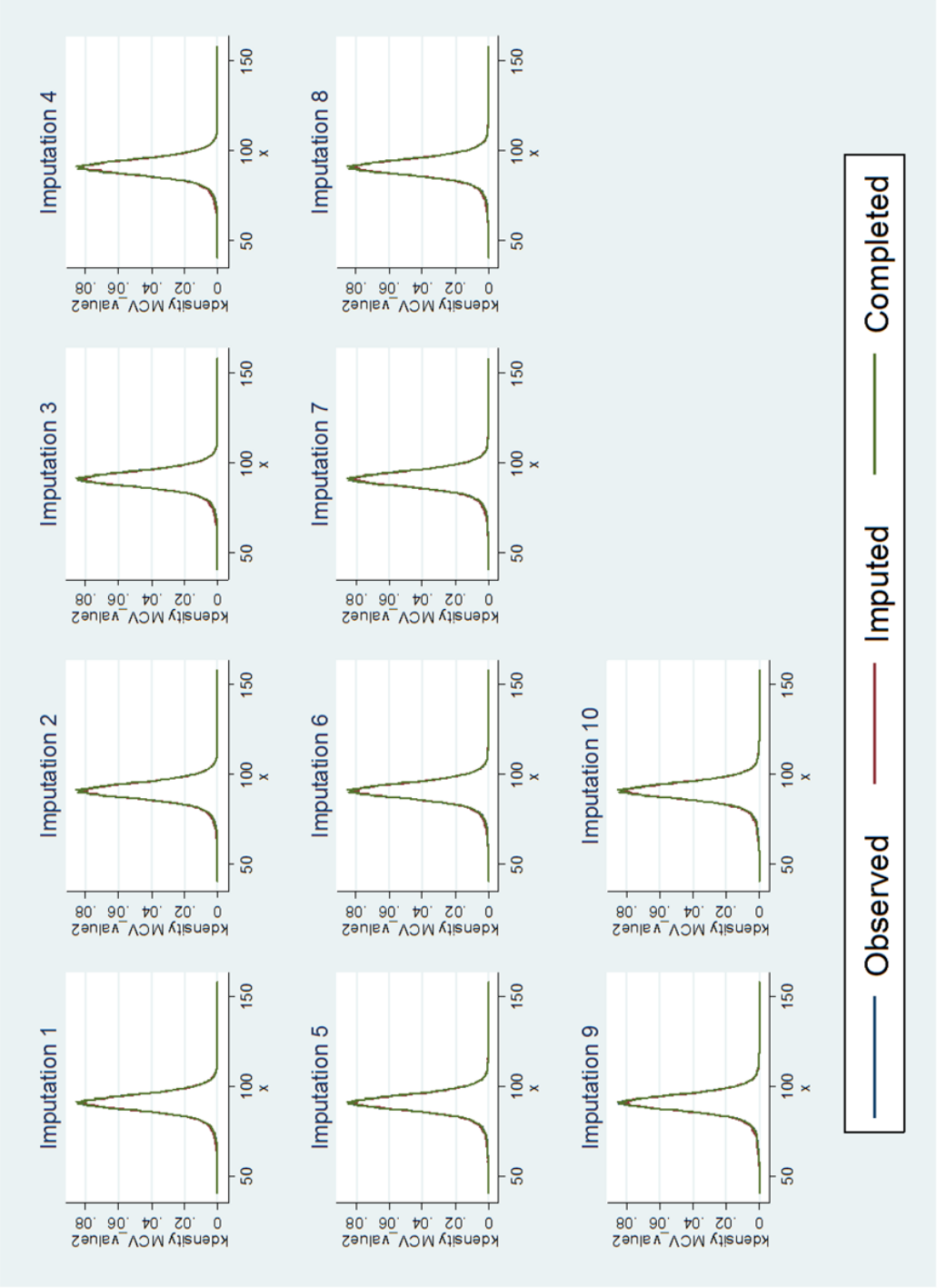
White cell count – index test



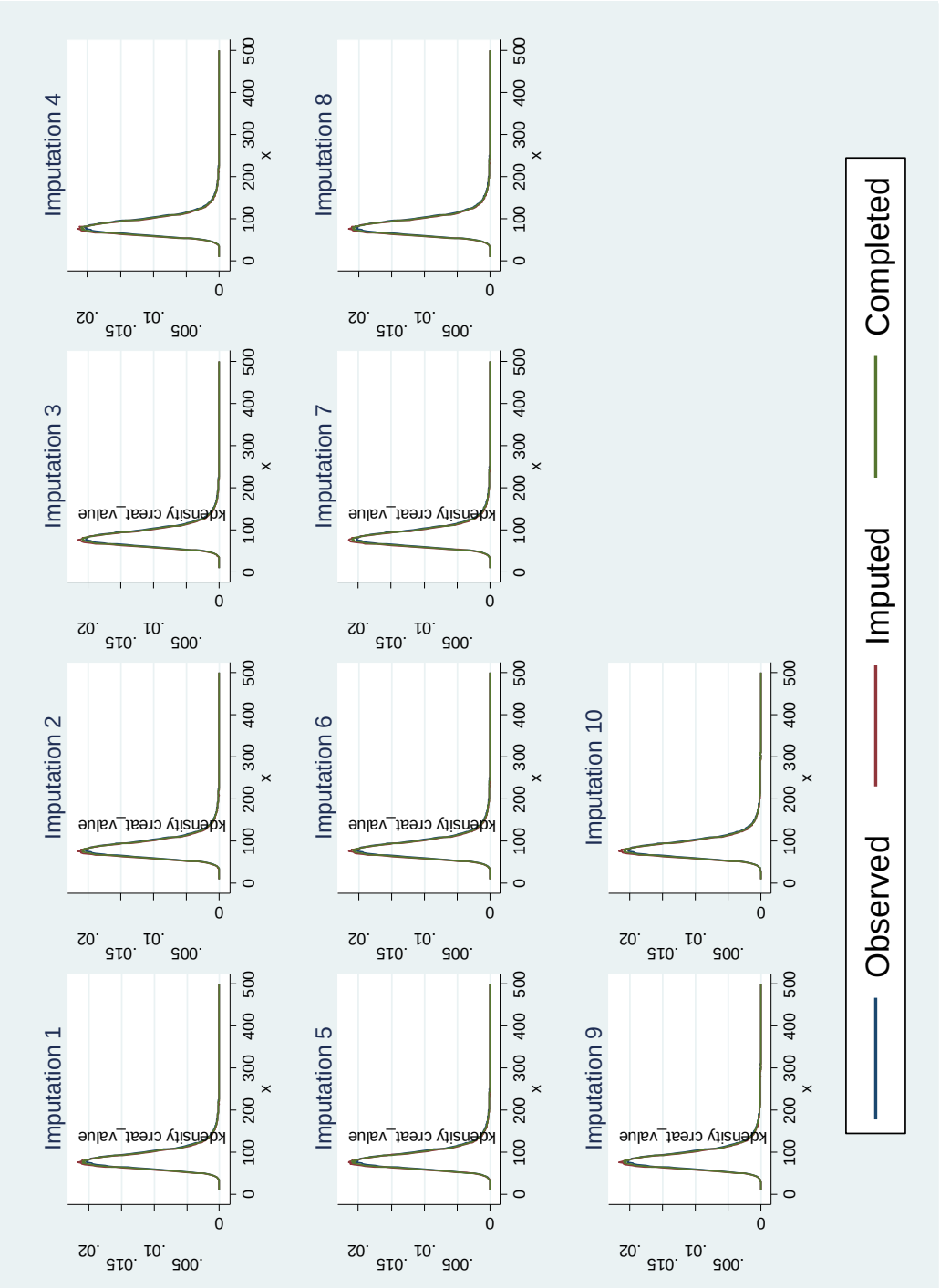
Platelets – index test



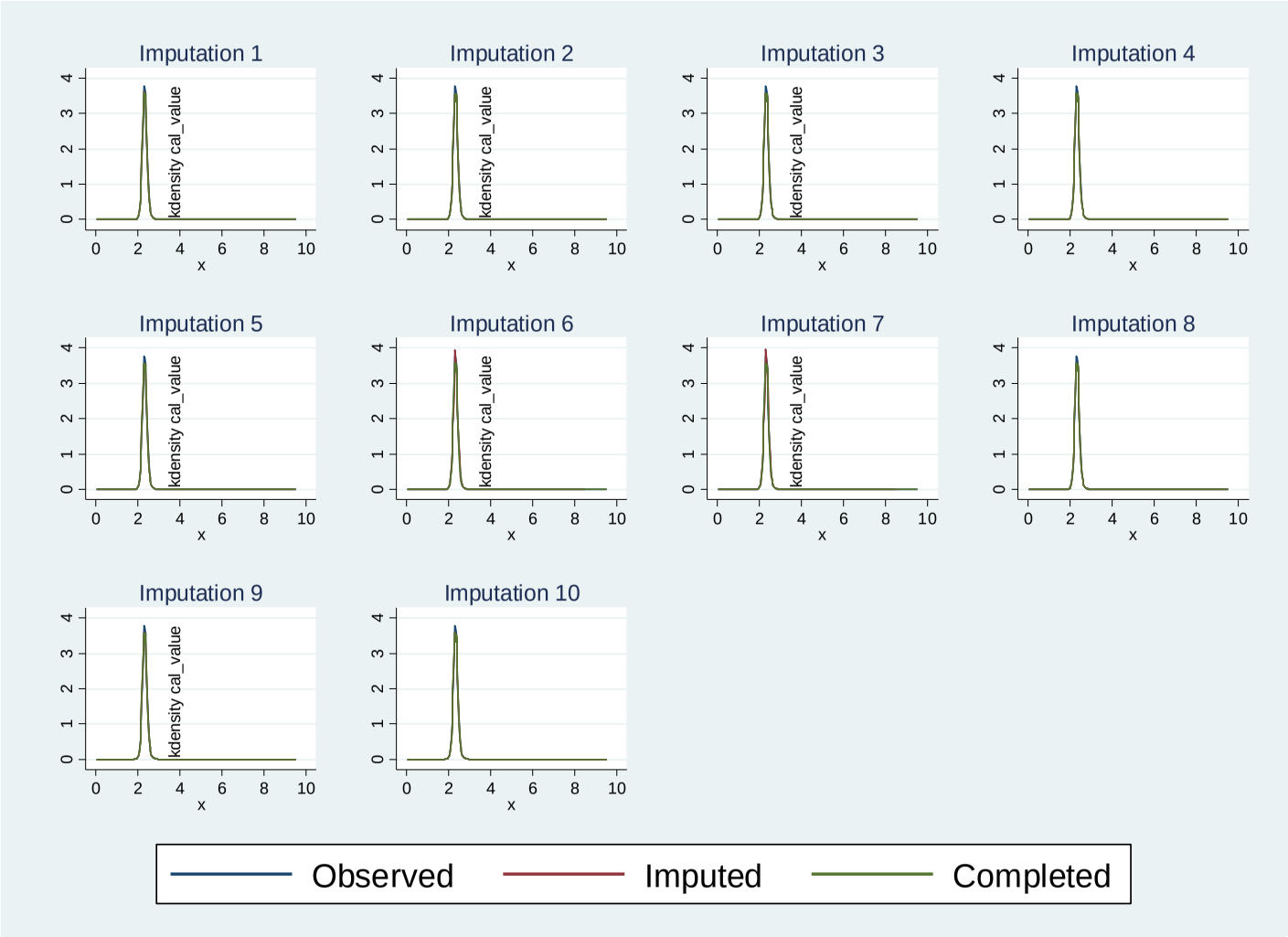
MCV –index test



Creatinine

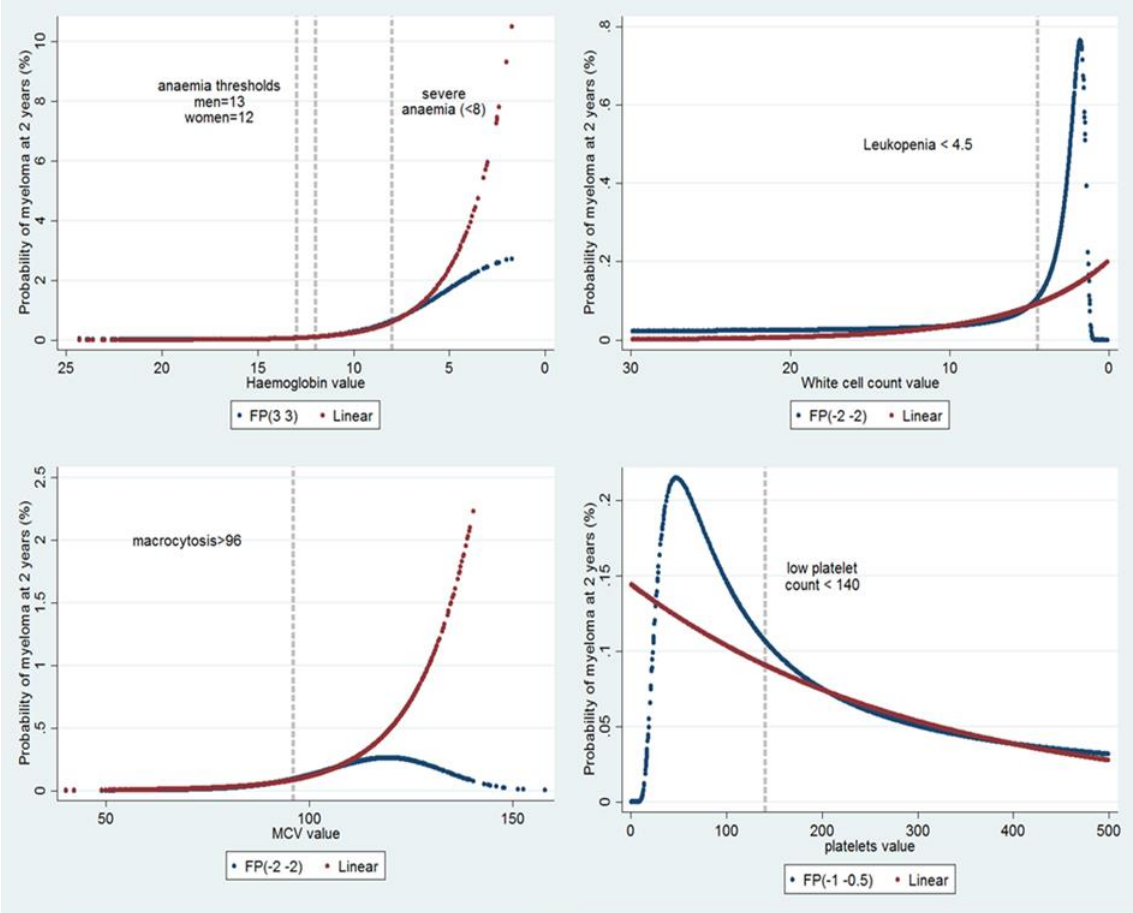


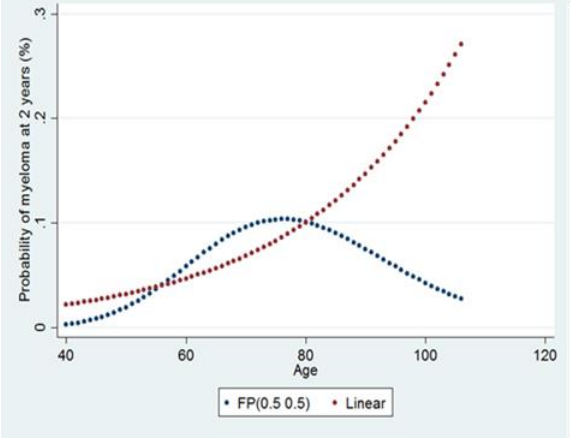
Calcium


Imputation 1
Imputation 2
Imputation 3
Imputation 4
Imputation 5
Imputation 6
Imputation 7
Imputation 8
Imputation 9
Imputation 10
kdensity cal_value
x
0 2 4 6 8 10
0 1 2 3 4
Observed
Imputed
Completed

APPENDIX 15 – Probability plots

Plots comparing the probabilities generated using predictors at the linear scale and the transformed variables.





APPENDIX 16– TRIPOD statement

Section/Topic			Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	90 (D),128 (V)
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	NA
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	21, 90, 128
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	91 (D), 129 (V)
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	90-91 (D, V)
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	90-91 (D, V)
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	90-91 (D,V)
	5b	D;V	Describe eligibility criteria for participants.	91 (D, V)
	5c	D;V	Give details of treatments received, if relevant.	NA
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	91 (D, V)
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	92-94 (D,V)
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	NA
Sample size	8	D;V	Explain how the study size was arrived at.	92 (D,V)
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	99-102 (D,V)
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	96-102 (D)
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	96-102 (D)
	10c	V	For validation, describe how the predictions were calculated.	121, 122-123
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	130-138
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	NA
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	103-104
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	102-103
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	137-138
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	102-103
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	105-107
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	105-107
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	121-123
	15b	D	Explain how to the use the prediction model.	122-123
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	141
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	117-120, 149-150
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	141
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	120-121, 151-153
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	151-154, 169-171

Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	91
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	3