The application of predictive statistical modelling in the investigation of suspected classical Myeloproliferative Neoplasms

Fiona Cullen

PhD

University of York

Health Sciences

September 2018

Abstract

The Myeloproliferative Neoplasms are a clinically heterogenous group of bone marrow haemopoietic disorders that result in the overproduction of myeloid blood cells. The identification of recurrent gene mutations has aided positive identification of disease in a substantial proportion of patients, however, a significant number of individuals with classical MPNs do not have a detectable aberration. The clinical and laboratory presentation of these disorders shows significant overlap with features associated with reactive conditions, which, in patients without detectable genetic mutations, can lead to ambiguity in their diagnosis.

Traditionally, the decision to investigate an individual for a suspected classical MPN has been based upon thresholds in blood count parameters. In this work, current working practices have been audited to ascertain the extent to which diagnostic guidelines are adhered toe. We demonstrate that a significant proportion of referrals for the investigation of suspected classical MPNs do not meet these criteria. Furthermore, this work objectively assesses the diagnostic sensitivity and specificity of current guidelines in the identification of patients with classical MPNs.

The use of predictive statistical modelling is a contemporary approach to the identification of individuals with increased likelihood of suffering from a classical MPN. In this work, several predictive modelling methods were applied to a data set of laboratory and basic demographic information taken from a series of patients investigated for suspected classical MPNs. This work shows that predictive statistical modelling can reproducibly identify those patients who are likely to have a classical MPN from those who do not. These models offer increased specificity and sensitivity compared with the use of published investigatory and diagnostic guidelines. Predictive statistical modelling also offers the ability to triage those patients who are likely to have classical MPNs prior to further investigation, resulting in potentially significant cost savings to both clinical and laboratory services.

List of contents

List of Tables

List of Figures

Acknowledgements

Much of the data used to complete this work were generated by staff employed within the Haematological Malignancy Diagnostic Service (HMDS). All diagnostic data were collected during routine clinical care and diagnostic laboratory investigations.

My own contributions, were the development and validation of the diagnostic assay for the detection of *JAK2* V617F and *CALR* exon 9 mutations, the audit and assessment of protocols employed in the investigation of suspected MPNs, the development, training and testing of predictive modelling approaches and associated impact analysis of introducing predictive modelling in routine practice and the development of an interactive user interface.

I am incredibly grateful to my supervisors: Dr Simon Crouch for introducing me to R and the many hours he spent reviewing my scripts, as well as his academic support and guidance. Dr Richard Kelly, for his support and encouragement, as well as valuable insight into the clinical impact of this work. Professor Eve Roman for support and guidance in the completion of this manuscript, especially during the final weeks.

I would like to extend my gratitude to the following individuals from HMDS: Dr Andrew Jack for the opportunity to undertake this research. Dr Paul Evans and Michael Short for their guidance and supervision in the molecular lab. Dr Richard Jones for his patience whilst teaching me the nuances of structured query language.

None of this would have been possible without my family. My parents, Helen and Paul Bennet, my husband, Matthew, and our children, Nathaniel and Martha – your unwavering belief in my ability to succeed has kept me going at times when it seemed impossible. Thank you.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1. Introduction

1.1 What are myeloproliferative neoplasms?

The term Myeloproliferative Neoplasm (MPN) is used to describe a group of cancers which affect certain blood cells, known as myeloid cells. In human haematopoiesis (blood cell production), the cells can be broadly divided into two groups; myeloid and lymphoid cells. Myeloid cells include populations known as granulocytes, erythrocytes and megakaryocytes (see

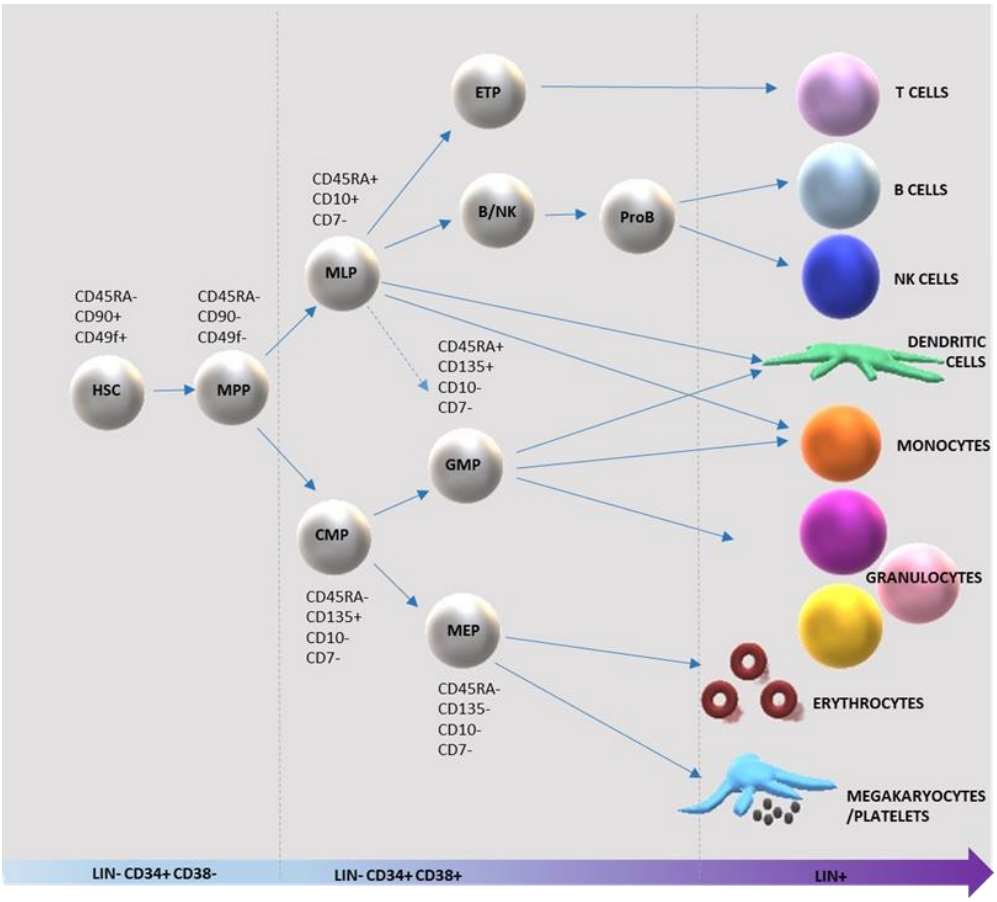

Figure 1-1).

The World Health Organisation (WHO) recognises over 70 different forms of myeloid malignancy (cancer). These are sub-categorised into acute myeloid leukaemias (AML), myelodysplastic syndromes (MDS) and the MPNs. Current WHO 2016 diagnostic criteria categorise 7 entities as MPNs (Arber et al., 2016). These are: - chronic myeloid leukaemia (*BCR-ABL1* positive), chronic neutrophilic leukaemia, chronic eosinophilic leukaemia (not otherwise specified), myeloproliferative neoplasms (unclassifiable), polycythaemia vera, essential thrombocythaemia and primary myelofibrosis. The lattermost three are collectively known as the classical MPNs.

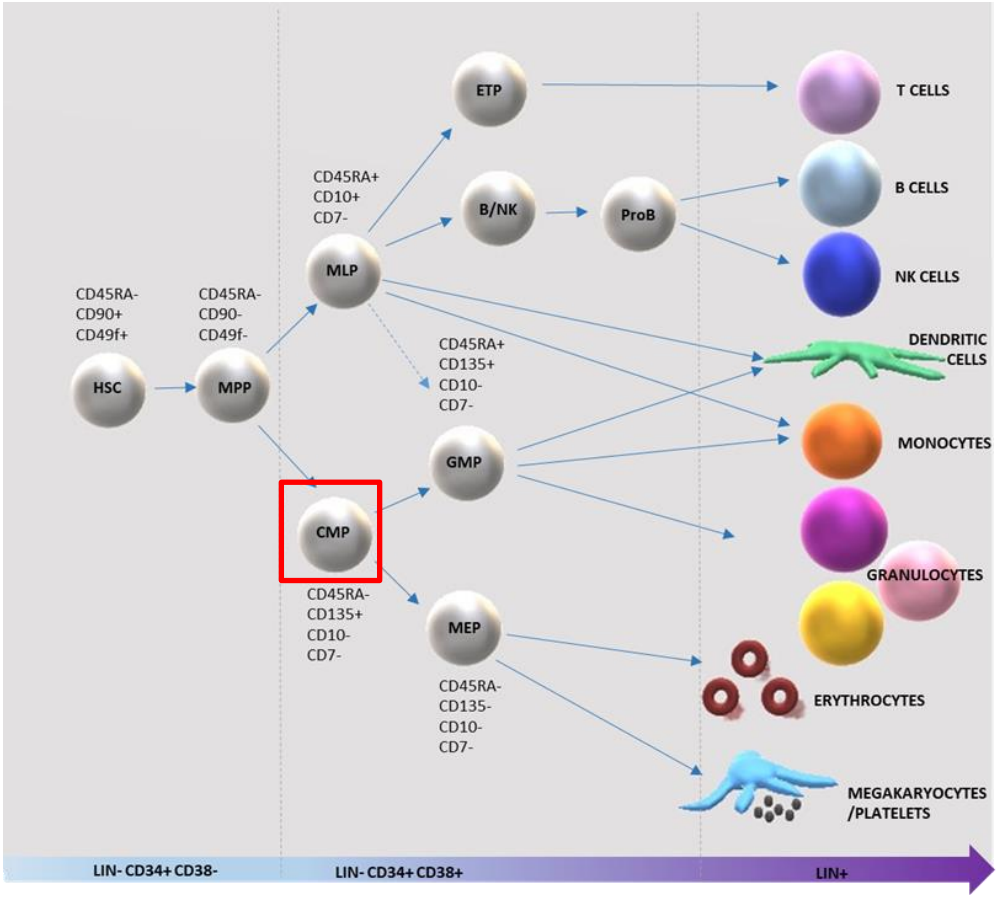

Figure 1-1 Schematic of human haematopoiesis, showing the shared ancestry of myeloid cells. The common myeloid progenitor (CMP) gives rise to the myeloid cell lineages. (reproduced from (Doulatov et al., 2012). CD nomenclature indicates the immunophenotypic characteristics at each stage of development/differentiation.

1.1.1 Essential Thrombocythaemia

Essential Thrombocythaemia (ET) is the most frequently diagnosed of the classical MPNs, with a reported annual incidence ranging from 0.21 – 2.27 per 100,000 (Titmarsh et al., 2014). It is characterised by increased proliferation of megakaryocytes in the bone marrow, and consequently, an increased peripheral blood platelet count (>450 x 10^9/L) (Arber et al., 2016). The incidence of ET increases with age, peaking at 71.5 years (Roman et al., 2016) and it occurs more frequently in females than males (F/M ratio 2.6:1) (Jensen et al., 2000). The female bias in ET diagnoses is an uncommon feature in malignancies, where an excess diagnostic burden in males is well documented across the majority of cancers. It is thought that hormonal, immunological and lifestyle difference contribute to a lower frequency of cancer diagnoses in women. The increased frequency of female ET diagnoses may also be related to these differences. It is reported that females have an increased platelet count in comparison to males, and that this may be related to hormonal differences associated with the female reproductive cycle. Blood loss associated with menstruation and childbirth. The prognosis of ET is generally good, with survival estimates close to those expected in comparable disease-free populations (Barbui et al., 2011b). However, a small number of patients undergo disease progression with the risk of developing overt fibrosis or acute leukaemia reported to be between 1.4 and 3.9% (Passamonti et al., 2008; Wolanskyj et al., 2006)

1.1.2 Polycythaemia Vera

Polycythaemia Vera (PV) primarily affects the production of erythrocytes and is characterised by an unregulated increase in the production of red blood cells. Whilst the erythrocytosis is a defining feature of PV, it is often also accompanied by increased proliferation of both megakaryocytes and granulocytes (Arber et al., 2016). The annual incidence of PV is estimated to be between 0.01 and 2.8 per 100000 in European and North American populations (Mesa et al., 2012; Moulard et al., 2014;

Phekoo et al., 2006; Rollison et al., 2008; Roman et al., 2016). Lower incidence rates are reported in Asian populations (Japan and Korea) (Byun et al., 2017; Kurita, 1974). The median age at diagnosis is similar to that seen in ET (Roman et al., 2016). In contrast to ET, PV is more commonly diagnosed in males than females, with an estimated male to female ratio of 1.2:1 ((Marchioli et al., 2005).

PV has a poorer prognostic outcome than ET. Without treatment, the median survival in PV is reported to be less than 2 years (McMullin et al., 2003). Current therapeutic approaches, have significantly improved this with 10 year survival reaching levels comparable to the general population (Tefferi et al., 2018). Progression to overt fibrosis or acute leukaemia is more frequent in PV than in ET, with studies indicating a risk of progression of around 7% (Marchioli et al., 2005; Tefferi et al., 2014a).

1.1.3 Primary Myelofibrosis

Primary Myelofibrosis (PMF) is the least common of the classical MPNs with an annual incidence estimated to be between 0.5 and 1.5 per 100000 (Arber et al., 2016). Characterised by the increased proliferation of both granulocytic and megakaryocytic cells alongside the deposition of fibrotic connective tissue within the bone marrow (Fujiwara, 2018; Koopmans et al., 2012). With disease progression, bone marrow fibrosis increases and extramedullary haematopoiesis occurs (the formation of blood cells in tissues/organs outside of the bone marrow, such as the spleen and liver) with a resultant leucoerythroblastic blood picture (Laszlo, 1975).

Prognostically, PMF has the poorest prognosis of the classical MPNs, with an estimated median survival of 6 years (Tefferi et al., 2014a), and AML transformation frequency of approximately 14% (Mudireddy et al., 2018).

1.1.4 Gender bias in ET and PV

Gender bias in cancer diagnoses is a well-documented phenomenon (Cook et al., 2011; White et al., 2010). An excess male burden is widely reported in the majority of (non-gynaecological/urological) malignancies, with the exception of thyroid cancer (2016; Cook et al., 2011). This is thought to be, in part, related to differences in epigenetic, hormonal and immunological characteristics between the sexes (Edgren et al., 2012). Lifestyle factors are also implicated, with higher intake of alcohol and tobacco and increased exposure to occupational hazards associated with males (Jaggers et al., 2009; Parkin et al., 2011). The gender bias seen in PV is in keeping with this general observation.

The increased frequency of ET in females is less typical and the reasons for this are not fully understood. Studies have shown that from the onset of puberty, average platelet counts are higher in females than males (Biino et al., 2013; Segal and Moliterno, 2006; Stevens and Alexander, 1977). It is postulated that this may be related to changes experienced during the female reproductive cycle. The physiological requirement for platelets during blood loss, through menstruation and/or childbirth may contribute to an increase in platelet production (Balduini and Noris, 2014). Studies in mouse models has shown oestrogen to promote megakaryocyte production (Nagata et al., 2003). Additionally, the associated reduction in iron stores may promote platelet production (as seen in iron deficiency anaemia) (Kadikoylu et al., 2006). One may postulate that chronic stimulation of platelet production may contribute to the increased frequency of ET in females or conversely, that a proportion of females categorised as ET may have physiologically increased platelet counts.

1.1.5 Overlap between ET, PV and PMF

In their text-book forms, each of the classical MPNs appears easily distinguishable from one another. However, in practice the distinction is less clear; the transformation/progression of ET and PV into fibrotic or blastic phase results in a clinical picture not dissimilar to that seen in PMF (Marchioli et al., 2005; Passamonti et al., 2004) . Likewise, all three diseases can undergo transformation to AML (Mudireddy et al., 2018; Wolanskyj et al., 2006) . This continuum of disease can cause blurring in the distinction between the three conditions.

Overlapping features are not confined to late stage/aggressive disease. In PV, the proliferation of erythroid cells is often accompanied by increases in the production of both granulocytes and megakaryocytes (Dameshek, 1951). In the event that a patient with PV has concurrent iron deficiency, the production of erythrocytes is decreased and as a result polycythaemia is often masked (Shih and Lee, 1994). In these cases, the presentation appears to be that of ET rather than PV (Barbui et al., 2014b). Furthermore, in early PMF, fibrotic changes in the bone marrow can be subtle (referred to as pre-fibrotic PMF), making the distinction between PMF and ET challenging (Giovanni Barosi 2012; Guglielmelli et al., 2017).

1.2 The biological basis of the classical MPNs

Classical MPNs share a common biological ancestry, with the affected cell lineages arising from a common progenitor. As shown in Figure 1-1, all human blood cells are derived from a multipotent haematopoietic stem cell (HSC). HSCs can divide and differentiate into any of the mature haemopoietic cell forms according to the body's requirements.

Briefly, HSCs divide to produce multi-potent stem cells (MPP), which can differentiate into one of two committed progenitors - the common myeloid progenitor (CMP) or the multi-lymphoid progenitor (MLP) (which in turn gives rise to early thymic (ETP) and B/NK cell progenitors (B/NK)) (Tenen et al., 1997). CMPs then undergo further differentiation to form either granulocyte/macrophage progenitor (GMP) or megakaryocyte/erythroid progenitor (MEP) (Doulatov et al., 2010; Doulatov et al., 2012). Classical MPNs primarily affect the production of megakaryocytes and erythrocytes (as well as granulocytes to a lesser extent), indicating a potential common underlying pathobiological basis for all MPNs at either the MEP or GMP stage of differentiation.

1.2.1 Determination of haematopoietic lineage fate

Haematopoiesis is a dynamic process which responds to the requirements of the human body (Kaushansky, 2014). Each type of mature blood cell has a different role within the body, as shown in Table 1-1, and their demand may be increased, for example during infection or following injury. The bone marrow responds to this need by adjusting the proportion of cells which differentiate into a given mature cell form (2014).

Cell type	*Role*
B lymphocyte	Adaptive immune response, antibody production
T lymphocyte	Cell mediated immune response
NK cell	Direct killing of foreign cells
Monocyte	Antigen presentation, removal of damaged and dead cells
Neutrophil	Phagocytosis of bacteria
Basophil	Histamine reaction, vasodilation, anticoagulation
Eosinophil	Kill parasites, allergic reactions
Erythrocyte	Transportation of oxygen to tissues
Platelet	Formation of blood clots, attract cells to site of inflammation/damage

Table 1-1. Basic function of mature blood cell types.

Overlap in the presentation and transformation of the classical MPNs hints at a biological relationship between the conditions. This is further supported by a shared/overlapping spectrum of molecular aberrations associated with the conditions. Mutations in one of 3 genes; *JAK2*, *MPL* and *CALR*, are found in the majority of patients presenting with classical MPNs. The frequency with which they occur varies between diseases, however, significant overlap is seen (see Figure 1-2).

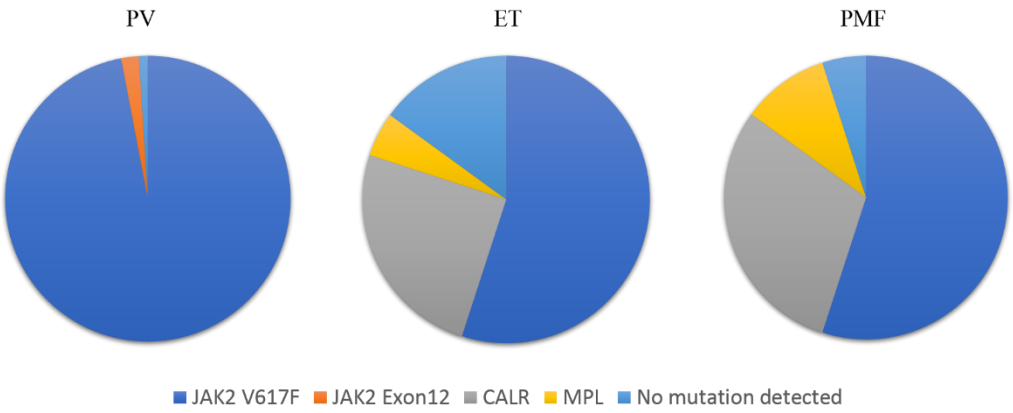

Figure 1-2. Frequency of mutations in *JAK2* (V617F and Exon 12), *CALR* and *MPL* found in the classical MPNs alongside the proportion of patients with no mutation detected in the listed genes. The data used to prepare these illustrations was obtained from (Nangalia and Green, 2014).

The *JAK2* V617F mutation has been reported to be present in more than 97% of patients with PV and in 50% of those patients with ET and PMF (Cross, 2011). A further 65-70% of patients with *JAK2* wild-type ET or PMF have been reported as having a *CALR* mutation (Klampfl et al., 2013b; Nangalia et al., 2013). *MPL* mutations occur less frequently, with W515L/K/A/R mutations reported in 3-4% of patients with ET and approximately 7% of patients with PMF (Beer et al., 2008).

1.2.2 *JAK2* mutations

The most frequently mutated gene in the classical MPNs is *JAK2* (Janus Kinase 2). *JAK2* is a non-receptor tyrosine kinase which is implicated in a wide variety of cellular processes, including cellular growth, development and differentiation (Gocek et al., 2014). JAK2 has many potential binding partners, the most relevant to its role in the development of classical MPNs are the Erythropoietin Receptor (EpoR) and thrombopoietin receptor, commonly referred to as MPL (The UniProt, 2014).

The most frequently identified *JAK2* gene mutation in MPN patients has been characterised as a single amino acid substitution of valine for phenylalanine at codon 617 of the protein (V617F). (Baxter et al., 2005; James et al., 2005; Jones et al., 2005; Kralovics et al., 2005; Vainchenker and Constantinescu, 2005). Mutated JAK2 kinase is constitutively activated, initiating the downstream STAT5 pathway which ultimately results in the uncontrolled proliferation of erythroid and megakaryocytic cells. (James et al., 2005) (Figure 1-3).

Additional mutations have been identified within the exon 12 region of the *JAK2* gene (Butcher et al., 2008; Martinez-Aviles et al., 2007; Pietra et al., 2008). In contrast to the V617F mutation, these mutations are diverse and include; deletions, substitutions and duplications, all of which affect the structure of JAK2 kinase. *JAK2* exon 12 mutations have been identified in approximately two thirds of those cases of PV which lack the *JAK2* V617F mutation (Scott et al., 2007).

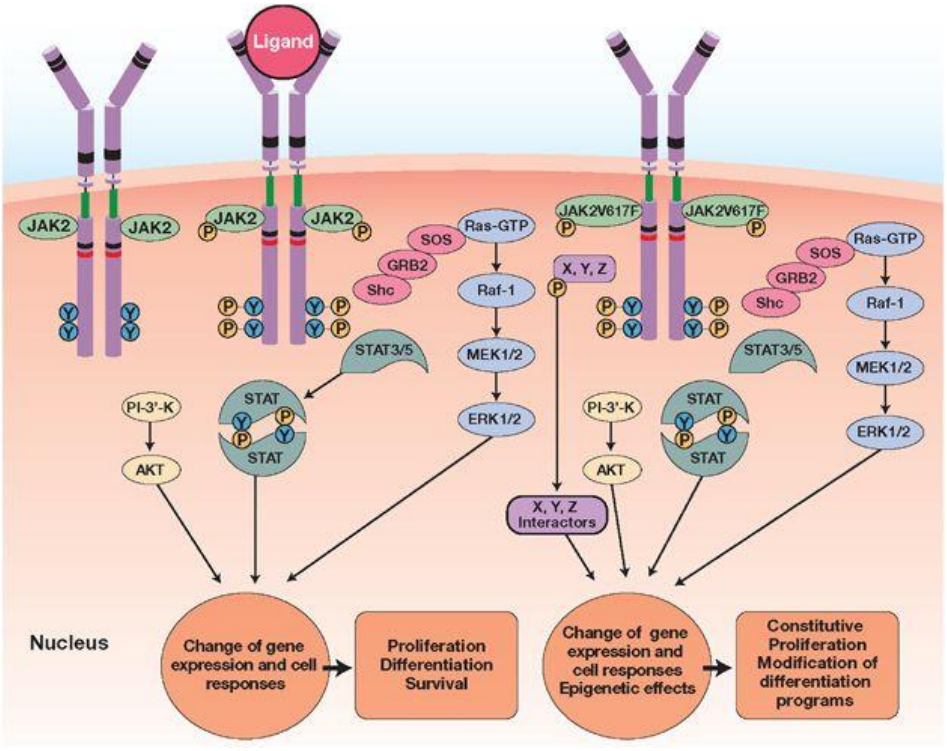

Figure 1-3. JAK-STAT signalling. Normal activation of the JAK-STAT signalling pathway is initiated by the binding of a ligand to a receptor (for example, EPO binding to EPOR). In the presence of the *JAK2* V617F mutation signalling occurs independently of ligand binding (image reproduced from (Vainchenker and Constantinescu, 2012).

1.2.3 Calreticulin gene mutations

In late 2013, two research groups reported the presence of somatic mutations within the calreticulin (*CALR*) gene in ET and PMF patients without *JAK2* mutations (Klampfl et al., 2013; Nangalia et al., 2013). CALR is a multifunctional chaperone protein, localised within the endoplasmic reticulum (ER) and is involved in the folding and localisation of cellular proteins, including the regulation of expression of cell surface receptors (Jiang et al., 2014).

Mutations identified within *CALR* are restricted to the exon 9 region of the gene. Importantly, the mutations are varied, but all cause a net insertion or deletion which alters the reading frame. This consistently results in the production of a truncated protein which is missing a sequence of 4 amino acids at the C terminal end. The missing amino acid sequence has been found to form a KDEL motif (Klampfl et al., 2013; Nangalia et al., 2013; (Chi et al., 2013).

KDEL acts as a signal for the localisation of proteins to the lumen of the ER and is found at the extreme carboxyl terminus of a number of proteins including Bip, calreticulin and protein disulphide isomerase. Loss of the KDEL motif results in the secretion of these proteins (Capitani and Sallese, 2009; Kraus et al., 2007).

Initially, the mechanism by which the cellular secretion of mutated *CALR* would give rise to ET/PMF phenotype was unknown. Studies have now shown that mutated CALR interacts directly with MPL by binding to the N-glycosylation sites on the extracellular domain of the receptor (Araki et al., 2016). This is thought to induce a conformational change which brings the intracellular portions of the receptor into close proximity and stimulates activation of the JAK/STAT pathway independently of TPO (Araki et al., 2016; Chachoua et al., 2016; Marty et al., 2016) (Figure 1-4).

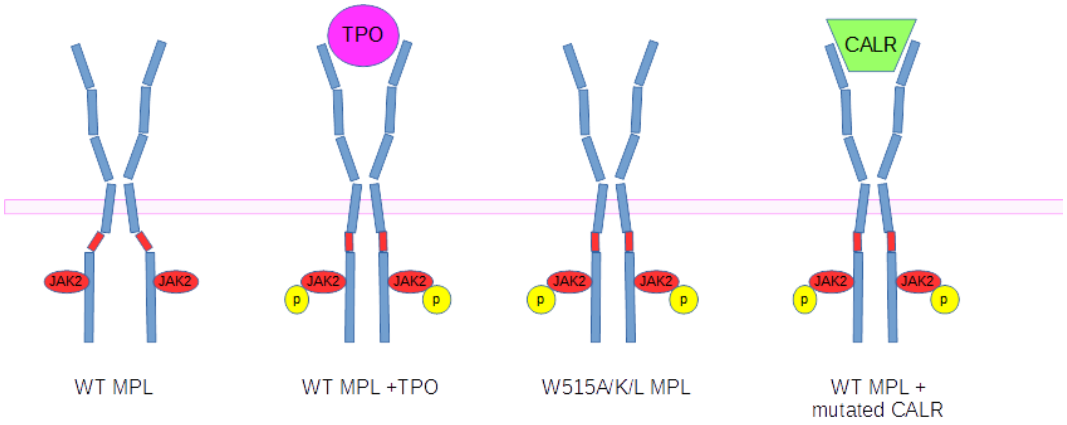

Figure 1-4 Ligand binding induces conformational changes in MPL, resulting in JAK2 phosphorylation and activation. The somatic mutations W515A/K/L alter the shape of MPL, mirroring the effect of ligand binding, allowing JAK2 phosphorylation to take place in the absence of ligand stimulation. In the presence of a *CALR* exon 9 mutation, CALR is secreted by the affected cells. It has been shown to specifically interact with MPL, mimicking TPO binding. (Chachoua et al., 2016; Lee et al., 2011; Mead and Mullally, 2017; Pecquet et al., 2010)

The pattern of mutations occurring in *CALR* exon 9 is highly variable, with over 50 different indels identified, the most frequently reported of these are a 52bp deletion and a 5bp insertion (Chi et al., 2013). Analysis of the resulting amino acid structure shows that these mutations affect the overall charge of the protein, impairing its calcium-binding abilities to different degrees (Pietra et al., 2015). Significantly, a difference in clinical behaviour has been observed in both mouse models and human patients carrying the two different mutations (Pietra et al., 2015). A more aggressive form of ET, with significantly higher risk of transforming to PMF was documented in those carrying the 52bp deletion, in contrast to those with the 5bp insertion who displayed an ET phenotype with little progression or transformation (Marty et al., 2016; Pietra et al., 2015). This work demonstrates the value in both identifying *CALR* mutated MPNs and of the characterisation of the *CALR* mutation itself.

1.2.4 MPL gene mutations

Mutations within the *MPL* gene have been identified in a small proportion of cases of *JAK2* V617F wild-type/*CALR* wild-type ET or PMF, and have been reported to occur in the exon 10 region of the gene (Pardanani et al., 2006; Pikman et al., 2006). A number of these mutations affect codon 515 of the *MPL* gene, resulting in the substitution of the normal amino acid tryptophan (W) with either Leucine (L), Alanine (A), Arginine (R) or Lysine (K) . The presence of a mutation induces a conformational change in MPL (Figure 1-4), which simulates the change which takes place following TPO binding (Mead and Mullally, 2017; Pecquet et al., 2010). This results in ligand-independent activation of the JAK-STAT pathway, leading to cellular proliferation (Pikman et al., 2006)

1.2.5 Unifying feature of MPN associated gene mutations

The unifying feature of these different mutations is their activation of the JAK-STAT signalling pathway. Each of the mutations initiates ligand-independent signalling, which ultimately results in cellular proliferation. To further emphasise the key role of this pathway in the development of MPNs, mutations in genes which regulate JAK-STAT signalling (such as LNK, SOCS and c-CBL) have also been shown to induce an MPN phenotype (Grand et al., 2009; Oh et al., 2010; Sanada et al., 2009; Vainchenker et al., 2011).

Whilst the presence of these JAK-STAT activating mutations is implicated in the development of MPNs, the mutation alone may not always be disease inducing. Low level somatic mutations in haematological malignancy associated genes, including JAK2 V617F mutations, have been reproducibly identified in otherwise healthy individuals (termed clonal haematopoiesis of indeterminant potential) (Genovese et al., 2014; Jaiswal et al., 2014; Steensma et al., 2015). In these studies,

the presence of additional gene mutations is associated with an increased risk of developing a malignancy.

Furthermore, several studies in Denmark have identified low level JAK2 V617F mutations in the general population, with longitudinal analysis showing progression to an MPN was frequent, but not universal (Nielsen et al., 2013; Nielsen et al., 2011; Nielsen et al., 2014). This may suggest that the development of MPNs follows a multi-hit model, with additional mutational events required for the MPN phenotype to develop.

1.2.6 Additional genetic mutation in MPNs

Somatic mutations in epigenetic modulators are widely reported in myeloid malignancies, including MPNs (Bejar et al., 2011; Cross, 2011; Nangalia and Green, 2014; Patel et al.; Shih et al., 2012). These include mutations in genes involved in DNA methylation (such as TET2, DNMT3a and IDH1/2), chromatic structure (ASXL1 and EZH2) and mRNA splicing (SF3B1, U2AF1, SRSF2). Within the MPNs, mutations in TET2, ASXL1 and DNMT3a are most frequently reported (Nangalia and Green, 2014) and are found in varying frequencies among the disorders, with PMF patients showing the highest frequency of additional mutations. The acquisition of additional mutations in MPNs is associated with transformation to acute leukaemia and an adverse prognosis (Abdel-Wahab et al., 2010).

The order in which mutations are acquired has been shown to impact on disease behaviour. Research has shown that a *TET2* mutation acquired prior to *JAK2* V617F mutation is associated with a MPN phenotype (Delhommeau et al., 2009; Ortmann et al., 2015), whereas a *TET2* mutation acquired in an already *JAK2* V617F mutated clone is associated with transformation to acute leukaemia (Abdel-Wahab et al., 2010; Ortmann et al., 2015).

1.2.7 Disease heterogeneity within molecular subgroups

The impact of JAK2/CALR/MPL mutations on the JAK-STAT pathway explains how the different mutations can give rise to the same disease phenotype, and the transitional entities seen in disease progression. However, it does not explain how, conversely, the heterogeneity of disease phenotypes that can arise from each molecular subgroup. For example, the *JAK2* V617F mutation, is reported in all 3 classical MPNs, each with a distinct clinical phenotype.

Studies have identified that *JAK2*, *CALR* and *MPL* mutations are detectable in haemopoietic stem cells as well in differentiated cell populations (Chaligne et al., 2007; Jamieson et al., 2006; Nangalia et al., 2013). This may offer some explanation of the diverse nature of MPN phenotypes. Whilst *CALR* and *MPL* mutations are detected in HSCs, the mutation specifically affects MPL, and its impact, therefore, is restricted to megakaryocytic cells. *JAK2* mutations are less specific in their target, and the development of heterogenous disease phenotypes in *JAK2* V617F mutated MPNs is less well understood.

There are a number of hypotheses as to how different disease phenotypes arise in *JAK2* V617F mutated MPNs. Firstly, PV has been shown to be associated with homozygous *JAK2* V617F mutations suggesting that a higher allelic burden may influence clinical phenotype (Godfrey et al., 2012). Quantitative differences in downstream signalling activation are also implicated, with increased STAT-1 activity inducing megakaryocytic differentiation and reduced STAT-1 activity favouring erythroid development in cell culture experiments (Chen et al., 2010). The acquisition of a *TET2* mutation in a *JAK2* V617F clone has been associated with a PV phenotype, suggesting additional mutations may influence clinical phenotype (Ortmann et al., 2015).

1.3 Clinical Presentation of classical MPNs

The distinction between the various classical MPNs can be challenging, as highlighted by the continuum of disease and the transitional entities as described in 1.1.5. The dynamic nature of haematopoiesis described in 1.2, coupled with the varied, non-malignant causes of increased erythroid/megakaryocytic and fibroblast activity also contribute to the difficulties in the accurate diagnosis of classical MPNs.

1.3.1 Essential thrombocythaemia

A peripheral blood thrombocytosis is the *sine qua non* of ET with thrombocytosis defined as demonstrating a platelet count greater than 450 x 10^9/L in adults (Schafer, 2004). This increase in platelets may be classed as a primary thrombocytosis, due to an underlying abnormality in haematopoiesis processes, or as secondary (also referred to as 'reactive') thrombocytosis, when the increase can be attributed to an external cause (Buss et al., 1994).

The body's requirement for platelets can be increased under a range of circumstances, some examples of which are shown in **Error! Reference source not found.**. This includes circumstances in which there is an increased requirement for platelets, for example during blood loss following injury or surgery. Inflammatory mediators such as interleukin-6 can cause stimulate platelet production (Burmester et al., 2005; Kaser et al., 2001; Wolber et al., 2001), under circumstances where an increase in platelet production is not necessarily required (secondary increase).

Increased platelet requirement	Secondary increase
Tissue damage/vascular injury	Infection
Post-operative	Inflammation
Haemorrhage	Iron deficiency
	Malignancy
	Therapy related

Table 1-2. Causes of reactive thrombocytosis (Harrison et al., 2010).

Thrombocytosis is often first identified as an incidental finding following a routine blood count. The most common cause of a raised platelet count would be a reactive thrombocytosis which is typically transient in nature and resolves without causing clinical complications or requiring intervention (Griesshammer et al., 1999). If platelet counts are assessed at only a single time point, it is impossible to determine whether the thrombocytosis was due to a reactive process.

Primary thrombocytosis is rarer and is most frequently attributable to ET. Complications within this group are more prevalent, with data suggesting that approximately 20% of patients present have symptomatic disease, presenting with features such as headaches, visual disturbances or thromboembolytic events (Brodmann et al., 2000). Differentiating between primary and secondary thrombocytosis has a significant impact on patient management. The identification of underlying reactive causes should be pursued alongside the confirmation of persistence of thrombocytosis. In cases where no reactive cause is identified, evidence of malignant causes should be sought (Harrison et al., 2010).

Symptomatically, ET can be difficult to diagnose. The symptoms associated with ET are primarily related to vascular occlusive events, which can also occur as a result of many other conditions (Rumbaut and Thiagarajan, 2010). Major thrombotic events affecting cerebrovascular, cardiac, pulmonary, hepatic or splenic circulation may be fatal (Sagripanti et al., 1996). Vascular occlusions within the micro-vessels are also common, causing a wide range of symptoms including pain and discolouration of the fingers and toes, headaches and visual dysfunction, which occur as a result of the transient suspension of blood flow (Frewin and Dowson, 2012).

1.3.2 Polycythaemia Vera

Polycythaemia is defined as an elevated haemoglobin (>185g/L in adult males and 165g/L in adult females). This may be classed as a true or apparent polycythaemia on the basis of whether the patient has an increased red cell mass (McMullin et al., 2005). Apparent polycythaemia occurs when plasma volume is decreased; this may be seen in individuals with increased alcohol intake (Biswas et al., 2003). True polycythaemia can be further separated according to whether the cause is primary and secondary (see Figure 1-5). Secondary polycythaemia occurs as a result of hypoxia or in some hereditary conditions (as shown in Table 1-3).

Congenital	Acquired
High oxygen-affinity haemoglobin variant	Cardiac disease
2,3-biphosphoglycerate mutase deficiency	Chronic lung disease
Erythropoietin receptor-mediated	Carbon monoxide poisoning
Chuvash erythrocytosis (vhl mutation)	Smoker's erythrocytosis
	End-stage renal disease
	Hepatocellular carcinoma
	Renal cell cancer

Table 1-3. Example of causes of secondary polycythaemia (adapted from McMullin et al., 2005b).

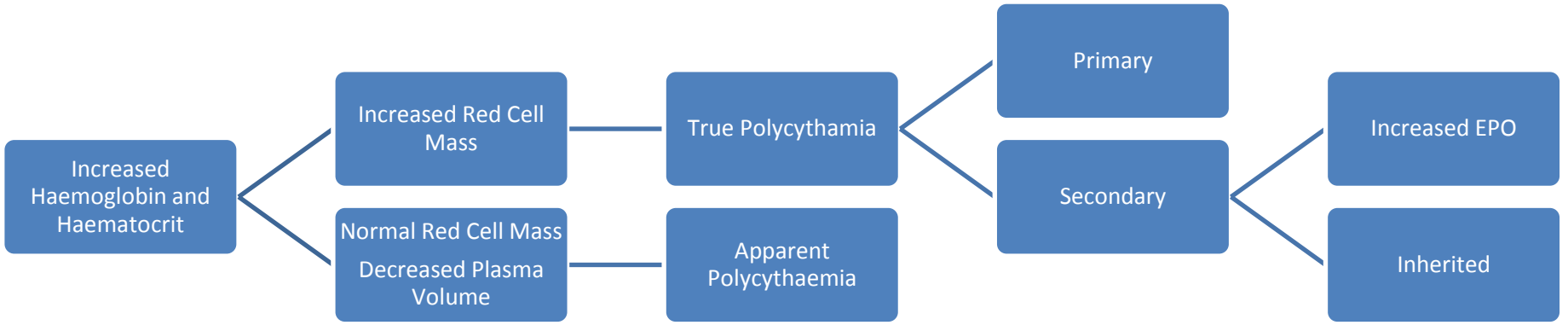

Figure 1-5. Differential diagnosis of polycythaemia. Patients presenting with increased haemoglobulin and/or haematocrit levels can be separated according to the presence or absence of increased red cell mass, allowing those with apparent polycythaemia to be identified. Where true polycythaemia is present, this can be further discriminated into primary polycythaemia (PV) and secondary polycythaemia.

EPO production is stimulated by low tissue oxygen levels. Hypoxia can arise in a wide range of conditions, as shown in Table 1-3, giving rise to increased levels of EPO and subsequently an increase in red cell mass (Lee and Percy, 2011). Polycythaemia may be asymptomatic, and only be identified as an incidental finding following a routine blood count. However, prolonged polycythaemia is associated with a range of symptoms including (Koopmans et al., 2012; Radia and Geyer, 2015);

- Fatigue
- Pruritis (Itching)
- Erythromelalgia (burning sensation in extremities)
- Headaches
- Dizziness
- Reddening of skin
- Splenomegaly
- Thrombosis

The symptoms associated with PV are primarily related to the thickening of peripheral blood and occlusive events which result from hyperviscosity (Emanuel et al., 2012; Geyer et al., 2014; Mesa et al., 2007). Similarly to patients with ET, major thrombotic events affecting cerebrovascular, cardiac, pulmonary, hepatic (Budd-Chiari) or splenic circulation can occur and may be fatal. (Sagripanti et al., 1996). Complications arising from PV are common, with data suggesting that approximately 30-40% of patients present with symptomatic disease and splenomegaly (Geyer et al., 2014).

1.3.3 Primary myelofibrosis

The presentation of PMF is highly variable, as PMF can arise as a *de novo* condition, or as a progression/transformation of an existing PV/ET (Bose and Verstovsek, 2016). Additionally, *de novo* PMF can present in both the prefibrotic stage of disease as well as in its overt form (Tefferi, 2018). As a result of these varying aetiologies patients may present with features ranging from hyper-cellularity (polycythaemia, leucocytosis, thrombocytosis) through to pancytopenia (anaemia, leucopoenia, and thrombocytopenia) and hepato-splenomegaly attributable to extramedullary haematopoiesis (Guglielmelli et al., 2017). Overt PMF is associated with a more severe clinical presentation than is reported in pre-fibrotic PMF (Rumi et al., 2018). The severity of symptoms may be contributed to the increasing hepato/splenomegaly; and can include portal hypertension, pain, and splenic infarction, as well as progressive bone marrow failure due to the increasing areas of fibrosis within the marrow (Guglielmelli et al., 2017).

1.4 Laboratory diagnosis of classical MPNS

Distinguishing between patients with classical MPNs and those with reactive conditions is essential for appropriate patient management. A number of laboratory investigations can be used alongside clinical examinations to achieve this.

1.4.1 Full blood count analysis

Full blood count (FBC) analysis is probably the most accessible and rapid of the laboratory investigations employed in the diagnosis of MPNs. Performed on a sample of peripheral blood, the FBC measures the major components of blood, with FBC reference values denoted in Table 1-4.

The white blood cell (leucocyte) populations (neutrophils, lymphocytes, monocytes, basophils and eosinophils) are quantified, along with erythrocytes (red blood cells) and platelets. The amount of haemoglobin carried within the red blood cells is measured as both a total (haemoglobin level (Hb), and per cell (mean cellular haemoglobin (MCH)). The size and shape characteristics of each cellular component is also measured.

Adult reference ranges	Male	Female	ET	PV	PMF
Haemoglobin (g/l)	130-180	115-165	↔	↑	↓
White cell count ($x10^9$/l)	4-11	4-11	↑	↑	↑
Platelet count ($x10^9$/l)	150-450	150-450	↑	↑	↑
Red cell count ($x10^{12}$/l)	4.5-6.5	3.8-5.8	↔	↑	↓
Mean cell volume (fl)	80-100	80-100	↔	↔	↔
Haematocrit (%)	40-52	0.37-0.47	↔	↑	↓

Table 1-4. Commonly measured full blood count parameters with their normal adult reference range alongside possible changes associated with classical MPNs. (↔ = within reference range, ↑ = increased, ↓ = decreased). (reference range from Leeds Teaching Hospitals NHS Trust, Leeds, UK).

Patients presenting with classical MPNs may show abnormalities within their full blood count parameters (2009). Increased haemoglobin and red cell counts are frequently associated with PV, and can be accompanied by an increased in both platelet and white cell counts (Arber et al., 2016). In cases of PV with concurrent iron deficiency, haemoglobin levels may not appear raised, but evidence of iron deficiency (such as a low mean cell volume) may be documented (Barbui et al., 2014a).

In ET, red blood cell associated parameters are typically within the expected range. Raised platelet counts are characteristically seen, along with increased white cell counts (Arber et al., 2016). The blood picture associated with PMF is more variable, with increased levels of platelets and white cells often present in early stages of disease, however, in patients with progressive disease, impaired bone marrow haematopoiesis can lead to reduced production of all blood cell types (pancytopenia) (Arber et al., 2016; Orazi et al., 2006).

These changes are not restricted to the classical MPNs. As described in 1.2, increased platelet counts and white cell counts commonly seen in patients with reactive conditions and raised haemoglobin and red cell parameters can be found in response to hypoxia.

1.4.2 Peripheral blood morphology

Morphological examination of peripheral blood and bone marrow cells can be valuable in the diagnosis of the classical MPNs. The increased numbers of red blood cells and/or platelets associated with PV and ET are visible on examination of a peripheral blood smear, as shown in Figure 1-6.

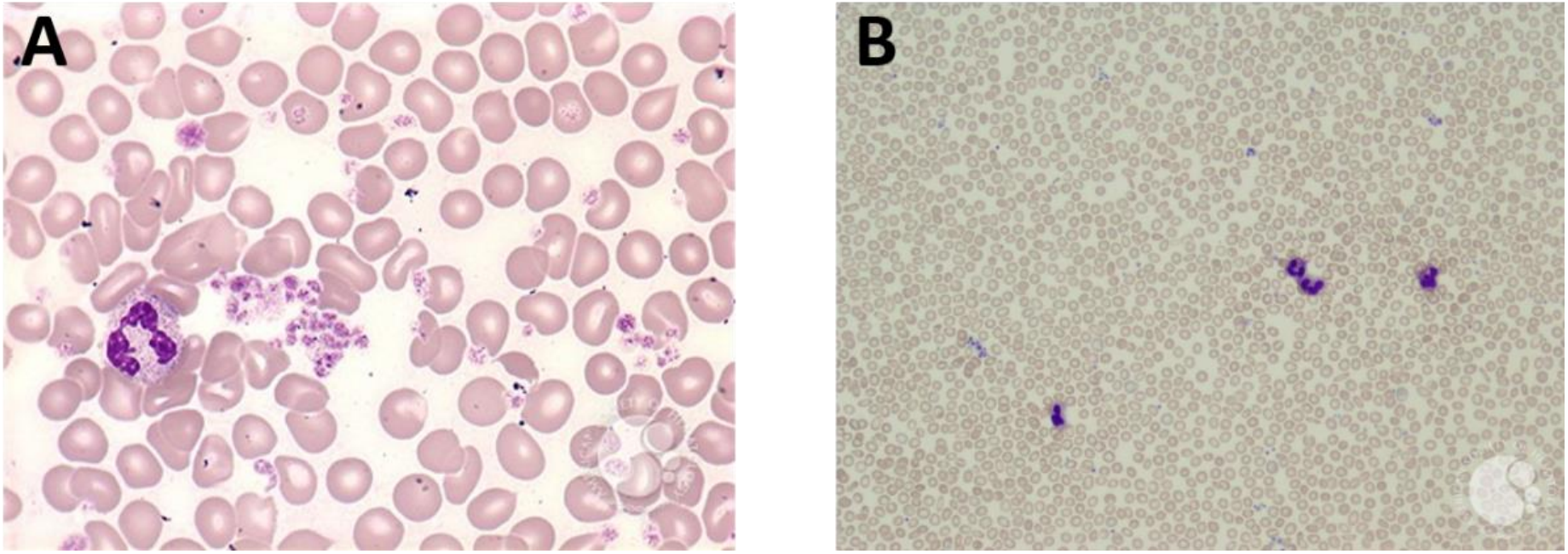

Figure 1-6. Peripheral blood morphology showing typical changes associated with (A) ET (increased platelets and platelet clumping) and (B) PV (increased red blood cells) (images from Ash Image Bank).

In PMF, morphological changes to the red blood cells are visible. The fibrous depositions in the bone marrow are thought to be the cause of damage to the membrane of red blood cells which alters their appearance (2009). Tear-drop poikylocytes (also called dacrocytes) are a characteristic feature of PMF (Figure 1-7), however, their presence is also documented in a wide range of other conditions including severe thalassaemia and haemolytic anaemia (Arber et al., 2016; Orazi et al., 2006; Zhao et al., 2009).

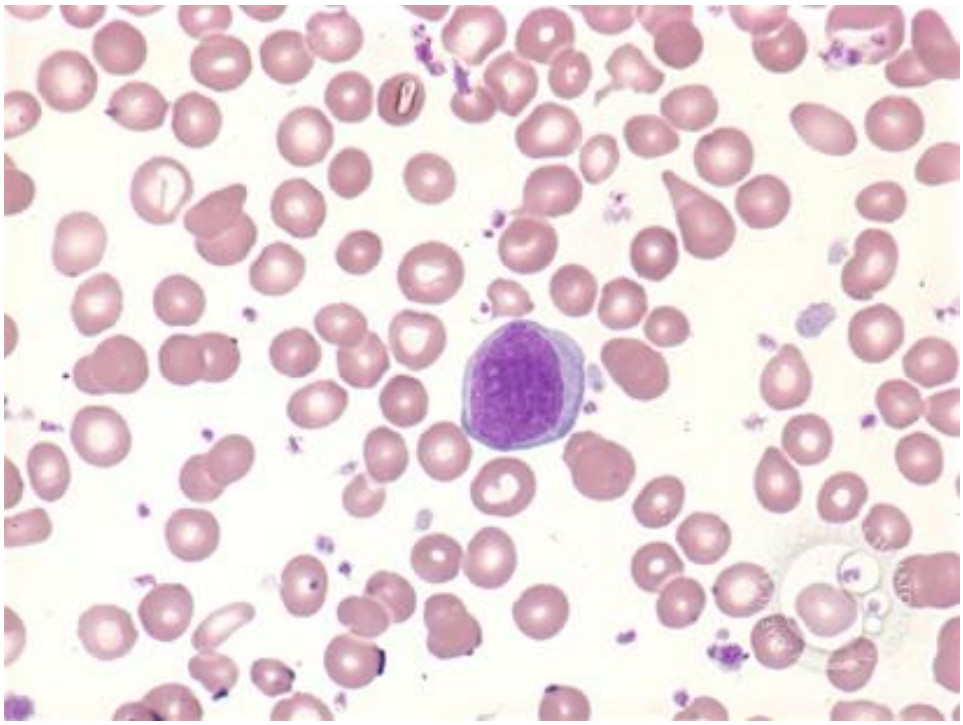

Figure 1-7. Teardrop-shaped red blood cells (teardrop poikylocytes/dacrocytes) are a prominent feature of PMF (image from ASH image bank).

1.4.3 Bone marrow examination

Bone marrow morphology is considered by many to be the gold standard method of diagnosing the classical MPNs (Schalling et al., 2017). Characteristics features present in classical MPNs allow them to be distinguished from reactive conditions as well as discriminating between diagnoses in cases of classical MPNs with ambiguous features from other tests (such as ET vs Prefibrotic PMF) (Arber et al., 2016; Orazi et al., 2006). Figure 1-8 shows examples of bone marrow morphology associate with the 3 classical MPNs.

1.4.4 Red cell mass

Polycythaemia can be seen as a result of reduced plasma volume (hyperviscosity), this phenomenon is known as apparent or relative polycythaemia. In these cases, full blood count analysis shows an increase in haematocrit (the proportion of the blood sample composed of red blood cells), however, the overall mass of red blood cells is not increased. Both smoking and alcohol consumption are associated with hyperviscosity and therefore, can present with an apparent polycythaemia (Biswas et al., 2003).

The measurement of red cell mass is performed using a radio-isotopic assay. This requires the incubation of peripheral blood with a radio-nucleotide, such as chromium 51, followed by reinjection back into the patient. Further blood samples are taken at 10, 20- and 40-minutes post reinjection. The proportion of chromium 51 labelled red cells is measured and used to determine overall red cell mass. Unfortunately, despite being a very useful assay in distinguishing between apparent and true polycythaemia, the complexity of this assay means it is not widely available in UK blood science or specialist pathology laboratories.

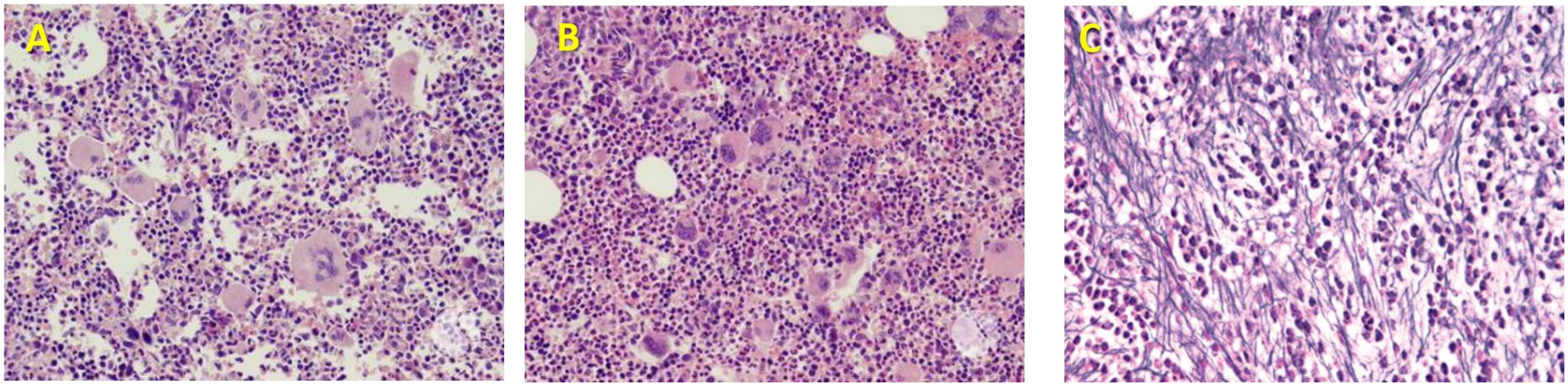

Figure 1-8. Typical bone marrow morphology seen in classical MPNs. Image A shows BM from a patient with ET, the predominant feature of which is the increased number of large megakaryocytes, with no associated increase in erythroid activity or fibrosis present. Image B shows the bone marrow of a patient with PV. The characteristic features of this bone marrow are the erythroid hyperplasia, accompanied by an increase in large megakaryocytes (similar to those seen in ET). Image C is an example of the morphology associated with overt PMF, with deposition of reticulin fibrosis throughout the biopsy. (All images obtained from the ASH image bank).

1.4.5 Serum Erythropoietin

The quantification of serum EPO levels is a useful tool in distinguishing between primary and secondary causes of polycythaemia. As shown in Table 1-3, secondary polycythaemia primary occurs as a result of hypoxia associated with a range of medical conditions and behavioural traits (such as smoking and high altitude training) (Biswas et al., 2003; Haase, 2013). Hypoxic conditions stimulate the secretion of EPO from the renal cortex resulting in increased serum EPO levels. In true polycythaemia, there is no increase in serum EPO levels (McMullin et al., 2007).

1.4.6 Molecular screening assays

As described in **Error! Reference source not found.**, the classical MPNs are f requently associated with mutations in the *JAK2*, *CALR* and *MPL* genes, amongst others. The frequency of each mutation in the classical MPNs is shown in Figure 1-2. Screening for these mutations is an essential aspect of the diagnosis of classical MPNs as the majority of patients will have one of these abnormalities. However, although none of the mutations are restricted to a specific classical MPN, the presence of a mutation in combination with the results of the other tests detailed in this chapter allows for accurate diagnosis of disease.

1.5 Diagnostic guidelines in classical MPNs

1.5.1 Historical aspects

The concept of MPNs was first proposed in 1951 by William Dameshek, whereby he referred to them as myeloproliferative disorders (Dameshek, 1951). In this work, Dameshek proposed that 5 conditions (chronic granulocytic leukaemia (CGL), polycythaemia vera (PV), idiopathic or angiogenic myeloid metaplasia of the spleen, megakaryocytic leukaemia and erythroleukaemia (including diGuglielmo syndrome)) should be considered as an interrelated family of conditions (Dameshek, 1951). Prior to this publication, each of these disorders had been described individually, in papers published as early as 1845 (see Figure 1-9).

At the time of Dameshek's publication there were no formally recognised guidelines for the diagnosis of haematological malignancies. In the years which followed, there began to be a recognition that variation existed in not only the clinical and laboratory features of haematological malignancies, but also in their responsiveness to treatments and movement towards formal classification schemes began.

In 1967 the International Polycythaemia Study Group (PVSG) was founded with the aim of undertaking a prospective study to assess the long-term effects of different treatment regimes in PV (Wasserman, 1971). In order to undertake this, it was necessary to develop diagnostic criteria for the inclusion of patients into the study. The group published its first recommendations for the diagnosis and management of PV in 1971 (Wasserman, 1971). The main features required for a diagnosis of PV are shown in Table 1-5.

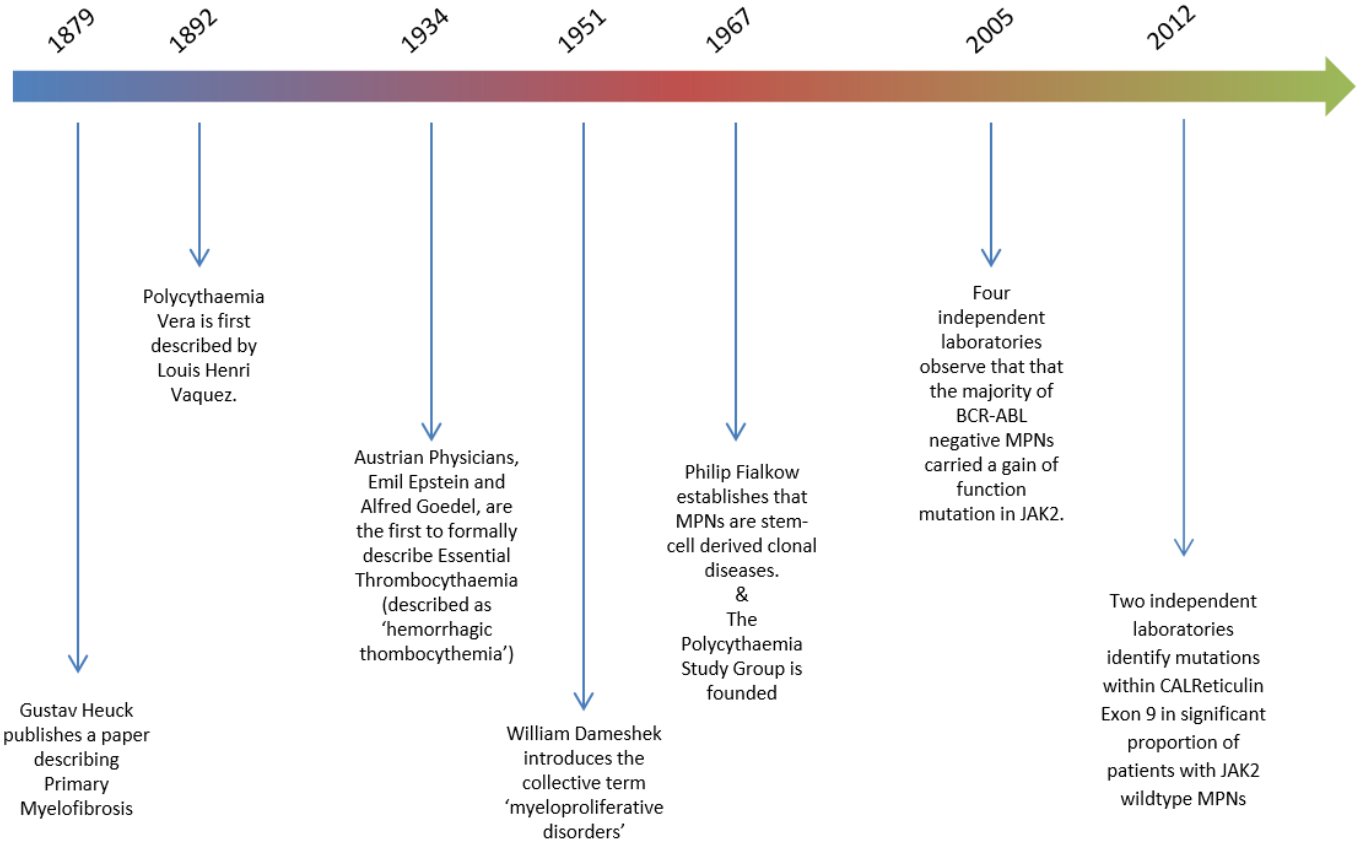

Figure 1-9. Timeline of major events in the understanding of classical Myeloproliferative Neoplasms. (Koumas et al., 2013; Nangalia et al., 2013; Tefferi, 2008).

Major criteria	Elevated Red Cell Mass
	Normal Arterial O_2 Saturation
	Splenomegaly
Minor criteria*	Leucocytosis >12000
2 minor criteria must be met in absence of splenomegaly	Thrombocytosis >400,000
	Leukocyte alkaline phosphatase >100
	Serum B_{12} >900 or B_{12} binding capacity >2200

Table 1-5. 1971 PVSG diagnostic criteria for PV (Wasserman, 1971).

Recommendations for the diagnosis of ET followed much later with interim guidelines for the diagnosis of ET published by the PVSG in 1986 (Murphy et al., 1986) followed by definitive recommendations for the diagnosis and management of ET in 1997 (Murphy et al., 1997). Concurrently, the recently formed Thrombocythaemia Vera Study Group (TVSG) also published a set of diagnostic guidelines for ET (Michiels and Juvonen, 1997). The two sets of guidelines differed mainly in the increased emphasis placed upon the bone marrow histology within the TVSG guidelines thereby allowing a lower platelet count threshold to be used in the diagnosis of ET. A summary of the two is shown in Table 1-6.

PVSG	TVSG
Platelet count >600x10^9/l	Platelet count >400x10^9/l
No evidence of reactive causes	No evidence of fever or infection
Normal red cell mass and normal mcv	Normal ESR
Normal BM iron stores	Normal bone marrow cellularity with an increase in megakaryocytes
No evidence of fibrosis or dysplasia	
No bcr/abl	

Table 1-6. 1997 PVSG and TVSG criteria for the diagnosis of ET.

The first formal guidance for the diagnosis of Myeloid Metaplasia, more commonly referred to as PMF was published in 1975 by one of the founding members of the PVSG, John Laszlo. The main criteria for diagnosis included; fibrosis involving more than a third of the section of bone marrow examined, splenomegaly, leucoerythroblastic features on morphological assessment, no evidence of increased red cell mass and the absence of the Philadelphia chromosome (Laszlo, 1975).

1.5.2 Current classification schemes

Current classification of the classical MPNs has evolved from these original guidelines. As our understanding of the underlying biology of these diseases has advanced, the guidelines have been adapted to accommodate this knowledge in order to refine and improve diagnostic processes.

There are two main publications currently in use for the diagnosis of MPNs in the UK. The diagnostic criteria as defined within the World Health Organisation (WHO) Classification of Tumours of Haematopoietic and Lymphoid Tissues, most recently the revised 2016 classification (Arber et al., 2016) and the recommendations published by the British Committee for Standards in Haematology (BCSH) (Harrison et al., 2010; Harrison et al., 2014; McMullin et al., 2005; McMullin et al., 2007; Reilly et al., 2012a; Reilly et al., 2014; Reilly et al., 2012b).

The WHO classification scheme is intended to provide a detailed description of the features of disease and define criteria for their identification. They are developed by leading experts within each clinical field and based upon both previous guidelines and published clinical and scientific evidence (Swerdlow et al., 2008). However, the criteria for classification do not cover best practice approaches for the implementation of the diagnostic and clinical processes, nor do they necessarily reflect the practical challenges faced by different geographic regions in terms of clinical and laboratory resources.

The BCSH publish guidelines which set the benchmark for the diagnosis and management of patients with Haematological Conditions in the UK and include, but are not restricted to, malignancy (BCSH, 2016; GRADE, 2016). Each set of BCSH guidelines is developed by a panel of clinical and scientific professionals with expert knowledge of the disorder. The panel use the recommendations of the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group for the assessment of quality of published evidence to inform the guidelines (GRADE, 2016).

There are subtle differences within the guidelines between the two groups which reflect the different evidence and opinion of the panel of experts authoring the guidelines as well as the difficulties in diagnosing these conditions from reactive processes and from each another.

1.5.2.1 WHO recommendations

The current 2016 WHO classification scheme for MPNs is an update of the previously published 2008 WHO classification of myeloid neoplasms (Vardiman et al., 2009a) which, in turn, are based upon the original PVSG/TVSG recommendations in conjunction with more recent scientific and clinical findings. The Myeloproliferative Neoplasms category, within the WHO classification scheme, contains 9 different diagnostic entities, as shown in Table 1-7. Similarly to Dameshek's original proposal for the myeloproliferative disorders, the WHO grouping of these entities is based upon the shared feature of the increased proliferation of morphologically normal mature cells of myeloid lineage.

The main revisions from the 2008 version of these guidelines are the inclusion of additional molecular markers and introduction of a new, lower, threshold for haemoglobin introduced (reduced from 185g/L to 165g/L in males and from 165g/L to 160g/L in females) (Barbui et al., 2015). These reduced thresholds have been selected to minimise the risk of missing cases of PV with mildly elevated haemoglobin levels and to avoid confusing these cases with ET (Barbui et al., 2014a; Barbui et al., 2014b).

These guidelines also see the separation of PMF into two distinct forms. The first of these is Pre-fibrotic Myelofibrosis (PrePMF), in which the fibrotic changes associated with PMF have not yet occurred and which previously would have been classified as ET on the basis of their lack of fibrosis. The inclusion of this subclassification is based on several studies by a group in Cologne, which have demonstrated that subtle changes are present and that patients displaying these Pre-fibrotic characteristics have an inferior outcome to those without pre-fibrotic characteristics (true ET) (Barbui et al., 2011a; Barbui et al., 2015; Gisslinger et al., 2016; Thiele et al., 2011). The second class of PMF is named "overt PMF" and represents those cases which would have met the criteria for PMF in earlier iterations of the guidelines. A summary of the WHO diagnostic criteria for each of the classical MPNs can be seen in Tables 1-8.

Myeloproliferative Neoplasms
Chronic Myeloid Leukaemia (CML) *BCR-ABL*$^{+}$
Chronic Neutrophilic Leukaemia (CNL)
Polycythaemia Vera (PV)
Primary Myelofibrosis (PMF)
PMF, prefibrotic/early stage
PMF, Overt fibrotic stage
Essential Thrombocythemia (ET)
Chronic Eosinophilic Leukaemia not otherwise specified (NOS)
MPN, unclassifiable

Table 1-7. 2016 WHO classification of MPNs with the classical MPNs shown in bold (Arber et al., 2016).

	PV	ET	PrePMF	Overt PMF
Major criteria	Haemoglobin >165g/dL (M), >160g/dL (F) *or* Haematocrit >49% (M), >48% (F) *or* Increase red cell mass (RCM)	1. Platelet count ≥450 × 109/L	1. Megakaryocytic proliferation and atypia, without reticulin fibrosis >grade 1, accompanied by increased age-adjusted BM cellularity, granulocytic proliferation, and often decreased erythropoiesis	1. Presence of megakaryocytic proliferation and atypia, accompanied by either reticulin and/or collagen fibrosis grades 2 or 3
	2. BM biopsy showing hypercellularity for age with trilineage growth (panmyelosis) including prominent erythroid, granulocytic, and megakaryocytic proliferation with pleomorphic, mature megakaryocytes (differences in size)	2. BM biopsy showing proliferation mainly of the megakaryocyte lineage with increased numbers of enlarged, mature megakaryocytes with hyperlobulated nuclei. No significant increase or left shift in neutrophil granulopoiesis or erythropoiesis and very rarely minor (grade 1) increase in reticulin fibers	2. Not meeting the WHO criteria for BCR-ABL1+ CML, PV, ET, myelodysplastic syndromes, or other myeloid neoplasms	
	3. Presence of *JAK2*V617F or *JAK2* exon 12 mutation	3. Not meeting WHO criteria for BCR-ABL1+ CML, PV, PMF, myelodysplastic syndromes, or other myeloid neoplasms	3. Presence of *JAK2*, *CALR*, or *MPL* mutation or in the absence of these mutations, presence of another clonal marker, or absence of minor reactive BM reticulin fibrosis	3. Presence of *JAK2*, *CALR*, or *MPL* mutation or in the absence of these mutations, presence of another clonal marker, or absence of reactive myelofibrosis
		4. Presence of *JAK2*, *CALR*, or MPL mutation		
Minor criterion	Subnormal serum erythropoietin level	Presence of a clonal marker or absence of evidence for reactive thrombocytosis	a. Anaemia not attributed to a comorbid condition b. Leucocytosis ≥11 × 109/L c. Palpable splenomegaly d. LDH increased to above upper normal limit of institutional reference range	

Table 1-8. Summary of the WHO classification criteria for the classical MPNs (Arber et al., 2016).

1.5.2.2 BCSH recommendations

The BCSH do not published a single document which encompasses all MPN's. Separate guidelines are produced for PV, ET and PMF and include best practice recommendations for investigation and clinical management. The guidelines for the investigation and management of PV were published in 2005 (McMullin et al., 2005) and updated in 2007 (McMullin et al., 2007) following the identification of a recurrent mutations with in the *JAK2* gene (Baxter et al., 2005; James et al., 2005; Kralovics et al., 2005; Levine et al., 2005; Scott et al., 2007). Diagnostic recommendations were amended to include molecular screening for these defects.

Guidance for the investigation and management of ET were originally published in 2010 (Harrison et al., 2010) and further modified in 2014 (Harrison et al., 2014) to include recommendations for additional mutational screening following the identification of mutations with the *CALR* gene (Klampfl et al., 2013a; Nangalia et al., 2013; Rotunno et al., 2013a).

Recommendations for the investigation and management of PMF were published in 2012 (Reilly et al., 2012b), with an amended version released in 2014 (Reilly et al., 2014). As in the case of ET, the updated recommendations include screening for *CALR* gene mutations alongside updated clinical recommendations. A summary of the BCSH diagnostic guidelines for each of these disorders are shown in Table 1-9 to 1-11.

***JAK2*-positive Polycythaemia Vera**

A1	High haematocrit (52 in men, 48 in women) or raised red cell mass (>25% above predicted) *
A2	Mutation in *JAK2*
***DIAGNOSIS REQUIRES BOTH CRITERIA TO BE PRESENT**	

***JAK2*-negative Polycythaemia Vera**

A1	Raised red cell mass (>25% above predicted) or haematocrit >60 in men, >56 in women.
A2	Absence of mutation in *JAK2*
A3	No cause of secondary erythrocytosis
A4	Palpable splenomegaly
A5	Presence of an acquired genetic abnormality (excluding *bcr-abl*) in the haematopoietic cells
B1	Thrombocytosis (platelet count >450 × 10^9/l)
B2	Neutrophil leucocytosis (neutrophil count > 10 × 10^9/l in non-smokers; >12.5 × 10^9/l in smokers)
B3	Radiological evidence of splenomegaly
B4	Endogenous erythroid colonies or low serum erythropoietin
***DIAGNOSIS REQUIRES A1 + A2 + A3 + EITHER ANOTHER A OR TWO B CRITERIA**	

Table 1-9. BCSH diagnostic criteria for PV (McMullin et al., 2005).

A1	Sustained platelet count >450 × 10^9/l
A2	Presence of an acquired pathogenetic mutation (e.g. in the *JAK2*, *CALR* or *MPL* genes)
	No other myeloid malignancy, especially PV, PMF, CML or MDS
A3	
A4	No reactive cause for thrombocytosis and normal iron stores
A5	Bone marrow aspirate and trephine biopsy showing increased megakaryocyte numbers displaying a spectrum of morphology with predominant large megakaryocytes with hyperlobated nuclei and abundant cytoplasm. Reticulin is generally not increased (grades 0–2/4 or grade 0/3)
***DIAGNOSIS REQUIRES A1–A3 OR A1 + A3–A5**	

Table 1-10. BCSH diagnostic criteria for ET(Harrison et al., 2014).

A1	**Bone marrow fibrosis ≥3 (on 0–4 scale)**
A2	Pathogenetic mutation (e.g. in *JAK2*, *CALR* or *MPL*), or absence of both BCR-ABL1 and reactive causes of bone marrow fibrosis
B1	Palpable splenomegaly
B2	Unexplained anaemia
B3	Leuco-erythroblastosis
B4	Tear-drop red cells
B5	Constitutional symptoms (drenching night sweats, weight loss >10% over 6 months, unexplained fever (>37·5°c) or diffuse bone pains).
B6	Histological evidence of extramedullary haematopoiesis
***DIAGNOSIS REQUIRES A1 + A2 AND ANY TWO B CRITERIA**	

Table 1-11. BCSH diagnostic criteria for PMF (Reilly et al., 2014).

1.5.3 Differences between classification systems

The most significant difference between the two groups is the emphasis placed upon bone marrow histology in the WHO classification (Arber et al., 2016). The assessment of morphological features in the bone marrow is a major criterion in each of the classifications shown. Within the BCSH guidelines, bone marrow histology is a major criterion in the diagnosis of PMF, however, in the diagnosis of PV and ET, it may not be clinically indicated (Harrison et al., 2014; McMullin et al., 2007; Reilly et al., 2012a).

This difference, particularly in relation to the diagnosis of ET, has caused some controversy, as it does not allow for the diagnosis of so-called Pre-PMF, where all other features would mimic those seen in ET (Gisslinger et al., 2016). A large independent cohort study had been able to identify morphological subgroup with the histological features said to indicate Pre-PMF, although longitudinal analysis did not show any evidence of increased fibrotic transformation within those patients and the adverse clinical impact of the original finding could not be confirmed (Wilkins et al., 2008).

There are small differences in the full blood count thresholds set out by each group. Within the criteria for the diagnosis of PV, the WHO includes a raised haemoglobin, whereas it is not included in the BCSH guidelines (Arber et al., 2016; McMullin et al., 2007). Differences are also present in the threshold level of haematocrit in the two sets of guidelines (Arber et al., 2016; McMullin et al., 2007). Numerical thresholds are used within all of these guidelines; however, it should be noted that there is no reference within the guidelines as to the underlying statistical basis of these values.

1.5.4 The use of thresholds in diagnostic guidelines

The use of thresholds is commonplace within diagnostic criteria and disease scoring schemes. Within the wider diagnosis of myeloid malignancy, the use of arbitrary thresholds applied to parameters such as blast percentage can make the difference between a diagnosis of myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML), albeit, a demonstrable clinical and outcome differences between a patient presenting with 19% blasts (MDS with excess blasts) versus 21% blasts (AML) has not been established (Lichtman, 2013).

Within the diagnosis of MPNs thresholds applied to laboratory parameters are equally vague. For example, the use of a platelet count threshold of 450 x 10^9/L reflects the upper limit of the recognised reference range. However, by definition, a reference range only encompasses 95% of the population and as such 2.5% of individuals will have a platelet count exceeding the upper value under normal circumstances. To date there are no published studies within the field of MPNs which systematically assess the probability of diagnosing an MPN with increasing laboratory and clinical parameters to establish the most appropriate threshold to use in diagnostic guidelines.

1.6 Clinical outcomes in classical MPN

Prognostic outcomes differ between the diseases, as well as showing variability within the disease itself. To enable clinicians to give patient's appropriate information about their prognosis, and make plans for their management, an accurate diagnosis is critical.

1.6.1 Prognosis according to disease type

The prognosis of patients diagnosed with PMF is significantly worse than in both ET and PV. Data published by Price *et al* from a series of 3364 MPN patients diagnosed in the USA shows inferior overall survival in all three classical MPNs compared to a control population (Figure 1-10) (Price et al., 2014). The most significant of which was associated with PMF, where median survival was 24 months (vs 106 months in the control group) (Price et al., 2014).

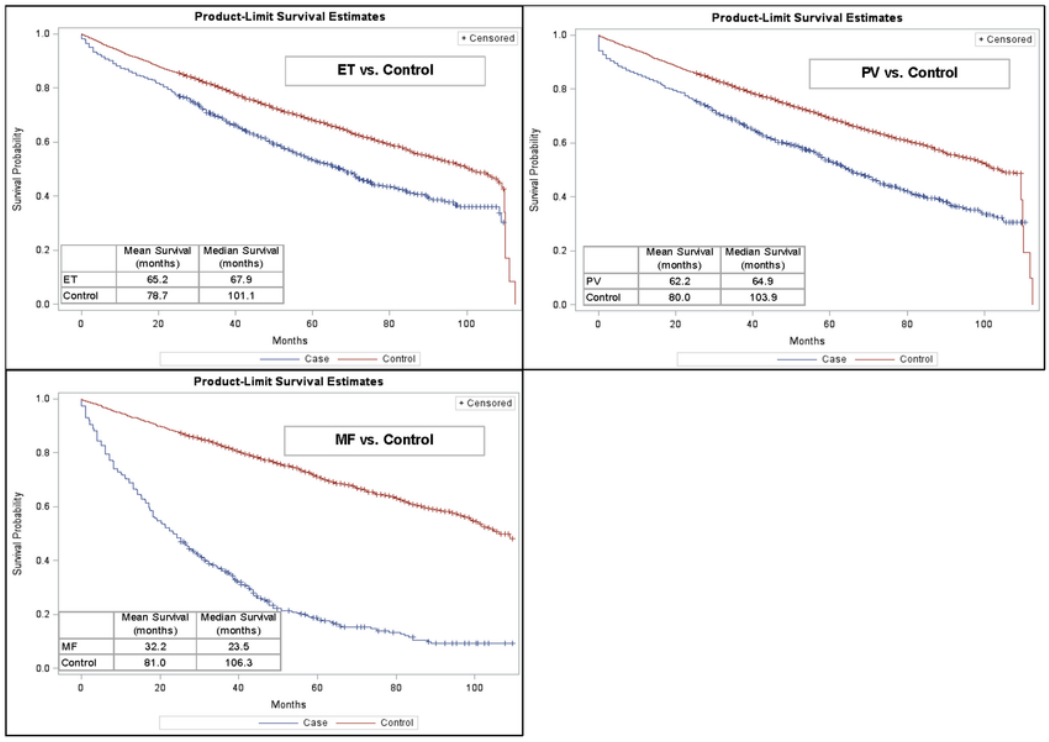

Figure 1-10. Kaplan-Meier survival estimates by MPN subtype compared to a control population. Figures reproduced from (Price et al., 2014).

1.6.2 Prognostic significance of genetic mutations

The prognosis for patients with ET and PMF also varies according to the underlying genetic lesion. As shown in Figure 1-11, ET patients with *JAK2* V617F mutations have an inferior prognosis to those with *CALR* and *MPL* mutations (Price et al., 2014). Patients with ET, who do not have a detectable mutation in any of these three genes, are referred to as triple negative or wild type (WT), and are shown to have superior outcomes (Rotunno et al., 2013a; Tefferi et al., 2014c) compared to those with a *JAK2* V617F, *CALR* or *MPL*.

Conversely, patients with triple negative PMF are shown to have a poorer prognosis than those with a mutation in either *JAK2*, *MPL* or *CALR*, with a median overall survival (OS) of 2.5 years (Tefferi et al., 2014b). In contrast, in those patients with PMF, *CALR* mutations are associated with the longest OS (median 8.2 years), followed by *JAK2* and *MPL* which have similar survivals (median OS 4.3 and 4.1 years respectively) (Tefferi et al., 2014b).

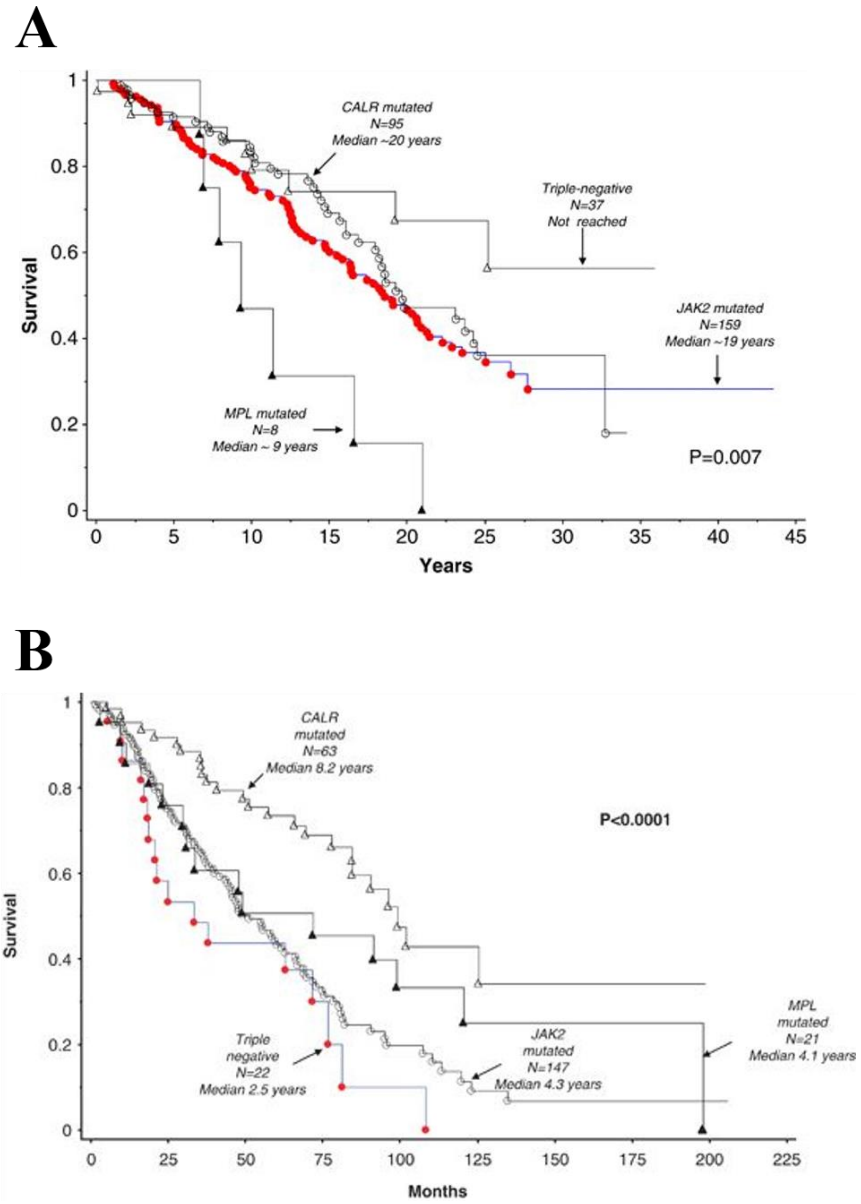

Figure 1-11. Kaplan-Meier survival estimates in (A) ET and (B) PMF according to molecular subtype (Tefferi et al., 2014b; Tefferi et al., 2014c)

1.6.3 Importance of distinguishing between ET and Prefibrotic PMF

As described in 1.5.2.1, the most recent revisions to the WHO classification of MPNs has separated PMF into two distinct entities – overt PMF and Prefibrotic PMF (Arber et al., 2016). Historically, many cases of prefibrotic PMF would have been classified as ET due to their similar morphological appearances and failure to meet the previous criteria for diagnosis of PMF (Barbui et al., 2018; Gisslinger et al., 2016; Thiele et al., 2011). Studies of patient with pre-fibrotic PMF have reproducibly shown an inferior prognosis in this group (Figure 1-12) (Giovanni Barosi 2012; Guglielmelli et al., 2017).

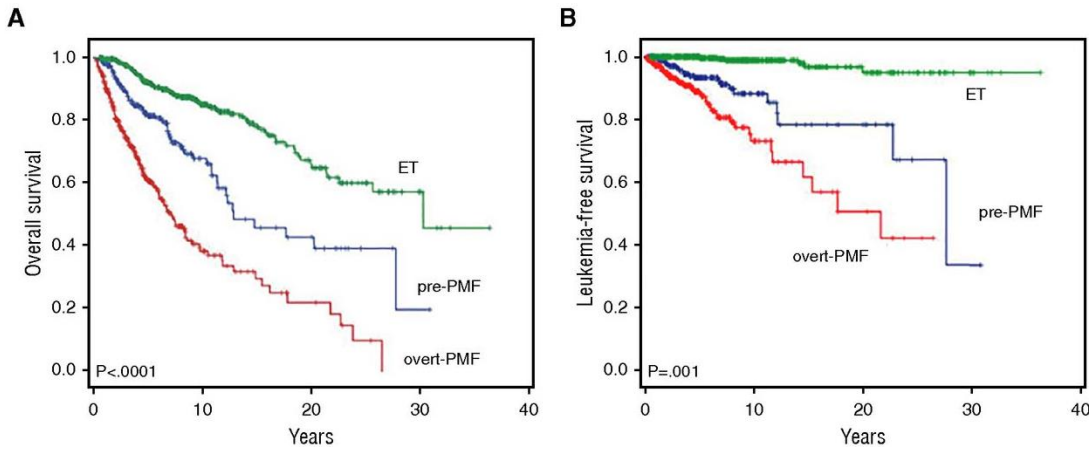

Figure 1-12. Overall (A) and leukaemia free (B) survival in ET vs Pre- and Overt-PMF (figure reproduced from (Guglielmelli et al., 2017). This data shows superior outcomes for individuals with ET compared to those with pre- and overt-PMF.

1.7 The diagnostic challenge

1.7.1 Understanding the practical limitations of investigating suspected classical MPNs

The data presented in the previous section highlights the importance of making an accurate diagnosis. This is particularly so in the case of those patients presenting with peripheral blood features consistent with ET where a mutation within *JAK2*/*CALR* or *MPL* is not detected. This is made difficult due to the aforementioned significant overlap between the presentation features of classical MPNs with those of reactive conditions. Whilst the presence of a raised platelet count could indicate that a patient has ET or Prefibrotic PMF, it is more likely to be the result of an underlying reactive process. In these instances, the absence of a mutation in *JAK2*, *CALR* or *MPL* would not exclude a diagnosis of ET or Prefibrotic PMF. In order to confirm or exclude these diagnostic entities, the patient would have to undergo a bone marrow biopsy for morphological assessment. If this approach was taken in all cases of thrombocytosis, it would not only be costly, but would also result in a large number of non-diagnostic, invasive bone marrow aspirate procedures. Which, whilst considered a relatively safe procedure is a cause of significant discomfort and distress to the patients undergoing investigation

1.7.2 Managing patient expectations without giving false reassurance

It is important to balance the practical implications of testing patients in whom clinical suspicion is low with the expectations of the patients and the reassurance that performing blood tests can offer. It is widely perceived that patients expect blood tests to be performed during clinical investigations and consider active testing approaches to be an indicator of quality of care (Hartley et al., 1987; Prochazka et al., 2005; van Bokhoven et al., 2006). However, there is also a perception that the

results of blood tests will be conclusive in the diagnosis or exclusion of disease (van der Weijden et al., 2002). In the case of suspected MPNs, this would be an unrealistic expectation. Clinicians may experience difficulties in conveying the limitations of performing *JAK2* V617F and *CALR* mutational screening to patients who are anxious, and who may have a limited understanding of biology. Furthermore, there is also evidence that when a patient is given the 'all clear' following cancer-related investigations, they can become dismissive of new or worsening symptoms prolonging the time taken to seek medical advice (Renzi et al., 2015; Renzi et al., 2016). This may occur in the investigation of suspected classical MPNs as peripheral blood screening alone is not capable of fully excluding the presence of disease. If a *JAK2* V617F wild-type/*CALR* wild-type result is conveyed as being an 'all clear' result, and no further investigations are performed, a patient who may have an MPN associated with another mutational profile (such as a mutation in the *MPL* gene or triple negative disease) may be falsely reassured and dismissive of symptoms which may indicate progressive disease.

One possible solution would be to better identify those individuals within the population who should be investigated for a possible MPN. Additionally, a tool capable of identifying those patients who would benefit from further testing when a definitive diagnosis cannot be made from peripheral blood screening would be valuable in addressing the challenges faced by clinicians.

2. Aims and objectives

Although clinical guidelines exist for the referral of patients with a suspected classical MPN, the actual referral pathways are not prescriptive, and it is unclear how efficient these guidelines are in identifying classical MPN patients in a real-world setting. It is hypothesised that the referral of patients with suspected classical MPN represents an increasing proportion of the haematopathology laboratory workload without a concomitant increased proportion or absolute number of confirmed classical MPN diagnoses. As any increase in samples referred for laboratory-based investigatory testing places a significant burden on laboratory and clinical resources, the overall aim of the study is to better develop methods to target patients with suspected MPNs for further investigation.

To achieve this aim, this study will:

- Develop, implement, and validate an assay for the detection of *CALR* mutations in patients referred for the investigation of a suspected MPN.
- This assay will then be applied in conjunction with *JAK2* V617F mutation screening to create an assay which is applicable in a routine, clinical laboratory-based setting.
- Evaluate both the numbers of referred cases, and the existing clinical referral pathway, to examine the potential to offer better value and increase efficiency in testing.
- Quantify the performance of current clinical guidelines and the extent to which they are followed in a real-world setting.
- Develop, test and validate predictive statistical models to ascertain whether such an approach would offer a benefit over existing referral strategies.
- Assess the potential impact a change in referral practise would have on an independent, real world cohort.
- Develop an interactive user interface to enable the proposed predictive model to be accessed prospectively by the diagnostic community at large.

3. Materials and methods

3.1 Rationale and introduction

This chapter describes the setting in which this research has taken place along with the relationship between the populations included in the analyses. The study population and methods of selection are discussed, along with the rationale for inclusion/exclusion criteria. In depth protocols are provided for each of the laboratory methodologies used in this thesis, with details of their interpretation. Details of the study designs chosen, and statistical methods employed, along with their interpretation and levels of significance are also included.

3.2 Setting

The work described in this thesis was performed at the Haematological Malignancy Diagnostic Service (HMDS) laboratory based within St James's Institute of Oncology; a part of Leeds Teaching Hospitals NHS Trust in the UK. Established in 1992, the department specialises in the diagnosis of haematological malignancies, including MPNs. Today, HMDS provides a comprehensive, diagnostic service to a population of over 6 million people (Figure 3-1) and receives approximately 35,000 patient referrals per annum, making it the largest department of its kind in the UK.

The department brings together all the laboratory techniques required to diagnose haematological malignancies in a single, multidisciplinary centre and all the data generated by the HMDS laboratory is stored centrally in a purpose-designed laboratory database, HMDS Integrated Laboratory Information System, 'HILIS'.

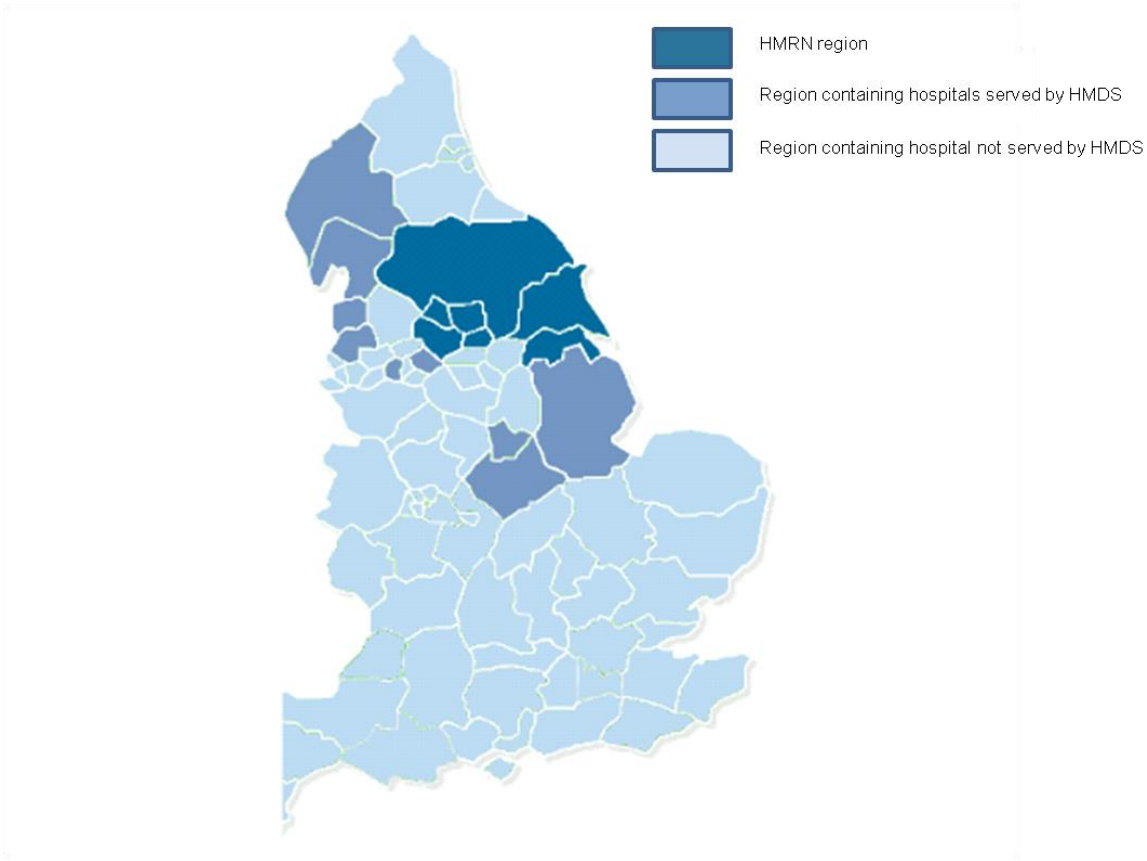

Figure 3-1. Map indicating the geographical regions containing hospitals which are served by HMDS, including those with the HMRN region.

The UK 2007 Cancer Reform Strategy recognised that integrating these technologies into a single diagnostic service was the best way to ensure diagnostic accuracy and optimal treatment for patients with haematological malignancies (DOH, 2007) Since its inception this approach has been strengthened by the rapid development of diagnostic technology and is now regarded as the national standard of care by the National Institute for Health and Care Excellence (NICE, 2015).

In 2004 The Haematological Malignancy Research Network (HMRN) was established, a collaboration between HMDS, a unified clinical network of 14 hospitals (comprising the Yorkshire & Humber and Yorkshire Coast Cancer Networks) and the Epidemiology and Cancer Statistics Group (ECSG) at the University of York (Figure 3-2) (Smith et al., 2010).

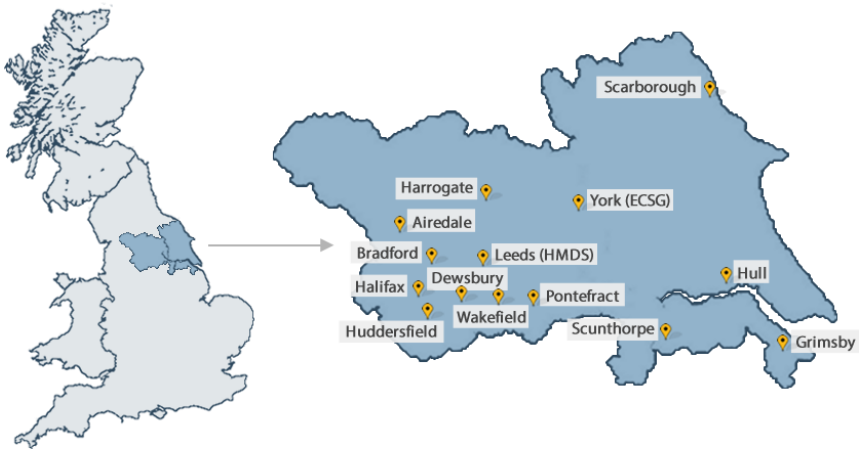

Figure 3-2. The geographical area and 14 hospitals served by both HMDS and HMRN.

The HMRN region covers a population of approximately 3.8 million people. Measures of the socio-demographic profile of individuals within the catchment region, as well as age and sex distribution are comparable to data collected from across the UK (Figure 3-3). This enables researchers to extrapolate data findings from the HMRN region to the national population (Smith et al., 2010).

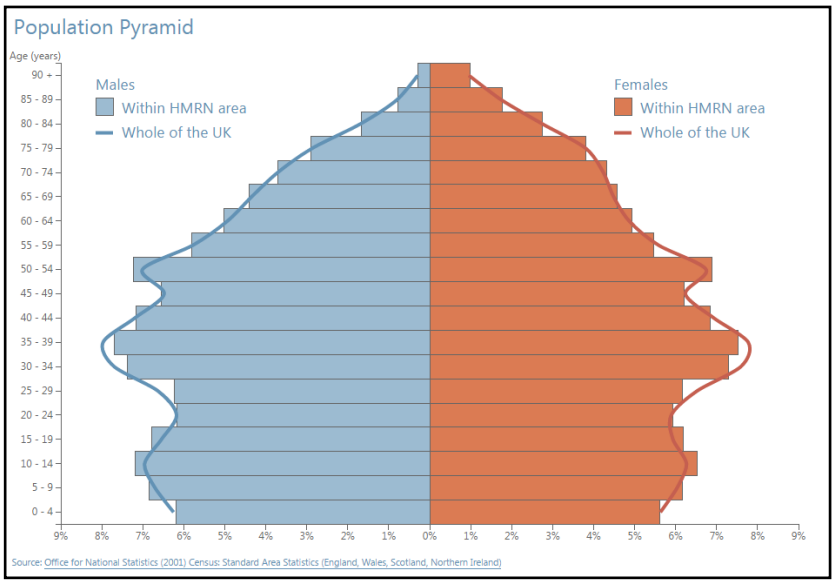

Figure 3-3. The Age and Sex distribution of HMRN population vs UK average (Smith et al., 2018).

Specimens from any patient with a suspected haematological malignancy visiting a hospital within the HMRN are referred to HMDS for diagnostic investigations. HMRN then tracks all patients diagnosed with a haematological malignancy within this region; obtaining information from clinical and laboratory records and linking to nationally compiled administrative records (mortality, cancer registration and hospital episode statistics), as well as collecting self-reported information from patients (Smith et al., 2018).

3.3 Clinical pathway for patients with suspected classical MPNs

HMDS receives specimens from hospitals both within and outside of the geographical region forming HMRN, the majority of these are sent by haematologists working within haematology outpatient clinics. An overview of the pathway a patient with a suspected MPN follows in Leeds Teaching Hospitals is shown below (Figure 3-4) with approximate timescales shown alongside each step. Discussion with clinical colleagues indicated that the same, or similar, approach was used throughout the HMRN region and beyond.

The patient is initially referred to the haematology clinic by a health care professional, generally a general practitioner. The first clinic appointment usually takes place within 14 days of this referral, at this visit, the patient is seen by a haematologist who takes a detailed medical history and performs a physical examination. The clinician also reviews the most recent peripheral blood counts available on the hospital result service. Following this consultation, the clinician requests the appropriate investigations to be undertaken and the patient is sent to the phlebotomist for venepuncture. It is at this point that a JAK2/CALR mutation screen would be requested, along with any other blood tests indicated (full blood count, biochemical markers).

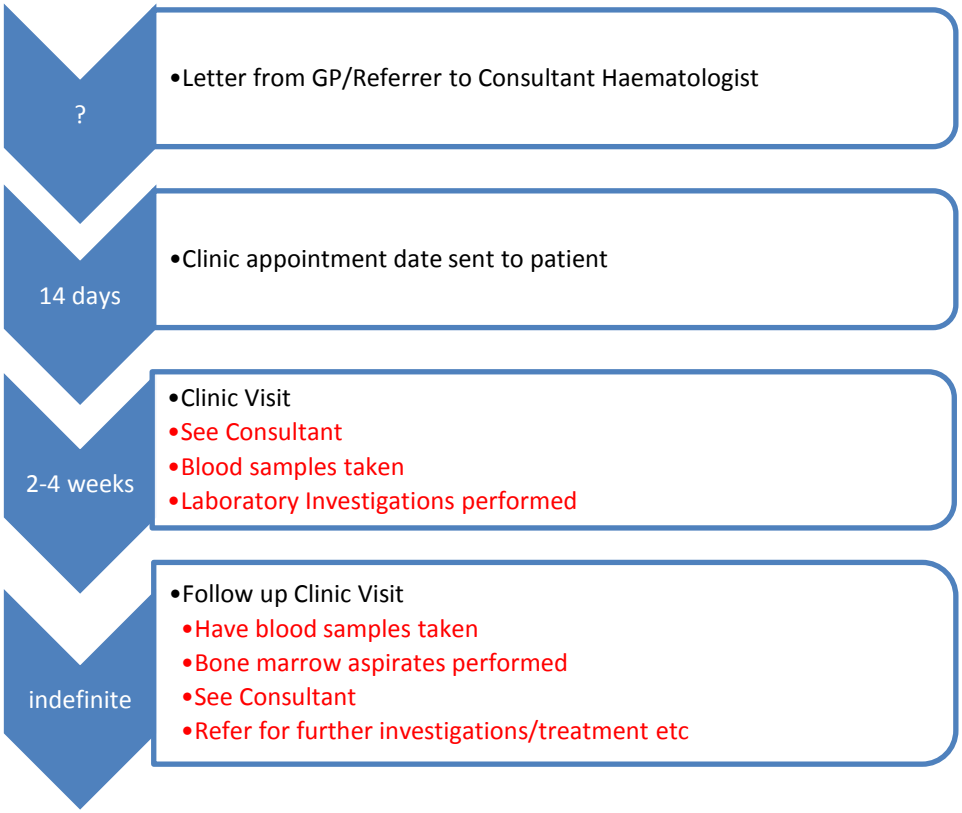

Figure 3-4. Clinical investigatory pathway followed by suspected MPN patients at Leeds Teaching Hospitals NHS Trust.

The patient returns for a follow-up appointment around 2-4 weeks later where the results of these investigations are reviewed. On the basis of these results, further tests may be arranged (such as bone marrow aspirates) and treatment may be initiated (if appropriate). Patients are then seen at intervals depending their clinical need.

3.4 Laboratory processes in HMDS

Upon their arrival at HMDS, specimens are prepared for morphological examination. A small volume of blood or bone marrow is spread onto a glass microscope slide and stained with cellular dyes (May-Grünwald Geimsa) before being viewed by light microscopy. Further laboratory tests are requested based on clinical information provided alongside morphological features and peripheral blood count data (where available). Details of each request are registered into HILIS and include the patients name, date of birth and NHS number as well as the referring hospital/clinician. A record is made of any laboratory tests requested and performed, the results of which are added to HILIS as they become available. If the patient has been investigated previously by HMDS, the computer records are linked to allow chronological monitoring.

The department is arranged in sections according to laboratory techniques employed – flow cytometry/immunophenotyping, histopathology/immunohistochemistry, molecular diagnostics (including Fluorescent *in situ* hybridisation, PCR based assays, gene sequencing and SNP analysis). Each section of the laboratory reports the results of its own tests and when every section has completed their tests, a single unified report is collated by specialist haematopathologists.

3.4.1 Diagnostic workflow for suspected classical MPNs

In HMDS, the diagnosis of a classical MPN is a multistep process, which may require the analysis of a second specimen in some patients. The workflow process followed by the HMDS laboratory is shown in Figure 3-5, briefly, where a blood sample is received with a suspected diagnosis of a classical MPN, *JAK2* V617F and *CALR* mutational screening is performed. If a mutation is identified, a diagnosis of a classical MPN can be made, if clinical features support this. If no mutations are detected, it is not possible to diagnose or exclude a classical MPN and a further sample (bone marrow) is requested for analysis. If a bone marrow is received by

HMDS from such a patient, it is tested for the presence of rarer mutations (*MPL* and/or *JAK2* Exon 12) and the bone marrow is morphologically assessed for features consistent with a classical MPN.

HMDS introduced *JAK2* V617F mutational screening in October 2005, and *CALR* mutational screening in February 2014. The department currently receives approximately 2500 referrals per annum for the investigation of suspected MPNs. The laboratory data stored by HMDS offers an unparalleled resource for the analysis of data related to the investigation of suspected classical MPNs and provides an excellent framework for this research.

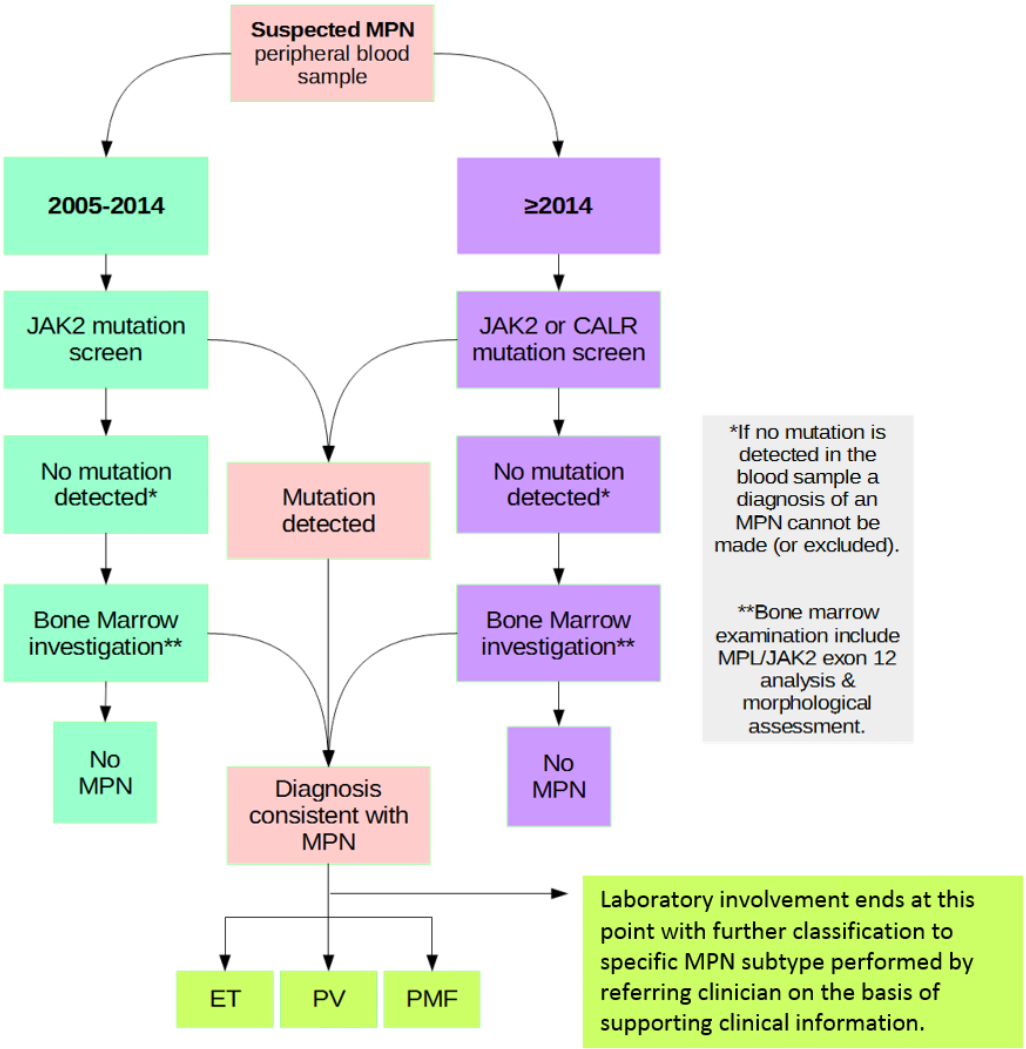

Figure 3-5. HMDS laboratory workflow process for the investigation of a suspected classical MPN.

3.5 Sample groups

Several groups of patient specimens have been analysed in this work. These have been divided into "Experimental" specimens, that is those which underwent additional laboratory analysis in order to generate data presented in this work, and "Statistical" specimens, that is those where only stored data has been used to perform the statistical analyses presented in this work.

3.5.1 Experimental specimens (Sample group 1)

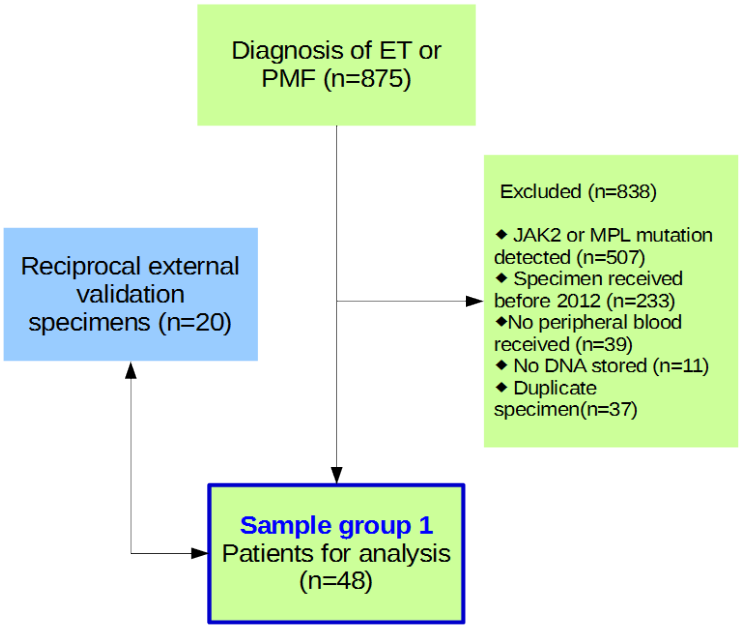

Figure 3-6. Summary of specimens used to develop *CALR* mutational screening assay.

The identification of *CALR* mutations in *JAK2* V617F wild type classical MPNs necessitated the development of a diagnostic assay capable of detecting these mutations for routine use in HMDS. HMDS records were searched for bone marrow specimens with a diagnosis of ET (n=875). Any specimen in which a *JAK2* or *MPL* mutation had been detected was excluded from further analysis (n=507). For

convenience and accessibility of stored DNA, specimens referred prior to 2012 were excluded (n=233). The remaining referrals (n=135) were reviewed and excluded if a peripheral blood sample was not received at the same timepoint (n=39) or DNA was not stored (n=11) (Figure 3-6). Paired PB and BM specimens from the remaining 48 patients were used to validate the PCR assay.

To comply with UKAS Medical Laboratory Accreditation (ISO 15189) standards, further external validation of the assay was performed through a mutual specimen exchange with Nottingham City Hospital's molecular diagnostic laboratory. Following analysis in HMDS, 20 of the specimens which had been tested were sent to Nottingham and a further 20 samples were received by HMDS in exchange. Following testing by each respective laboratory, the samples were returned, and the test results compared.

In addition to this reciprocal sample exchange, 30/48 of the HMDS patients included in this group had been analysed by the Wessex Regional Genetics Service as part of their Genome Wide Association Studies (GWAS) which led to the identification of *CALR* mutations. The results of the *CALR* analysis performed on these samples were made available to us for comparison.

3.5.2 Statistical Specimens (sample groups 2-6)

3.5.2.1 Sample group 2: All suspected MPNs (2005-2014)

Specimens received by HMDS for the investigation of a suspected MPN were identified from HILIS using structured query language searches. All records received between the beginning 2005 and the end of 2014 (when the search was performed) were included. The data collected included age, gender, molecular screening results and full blood count parameters (where available) as well as a unique patient identifier (NHS number) which was used to remove duplicate requests

on the same individual and to positively identify those patients who had a subsequent bone marrow examination to confirm or exclude a diagnosis of an MPN. Where the final diagnosis was not given a specific MPN subtype, data was reviewed with a clinician (Dr C Cargo or Dr R Kelly) to determine the most appropriate diagnosis. These specimens (see Figure 3-7) form a core sample group from which two further subgroups of specimens have been generated (Sample groups 3 and 4).

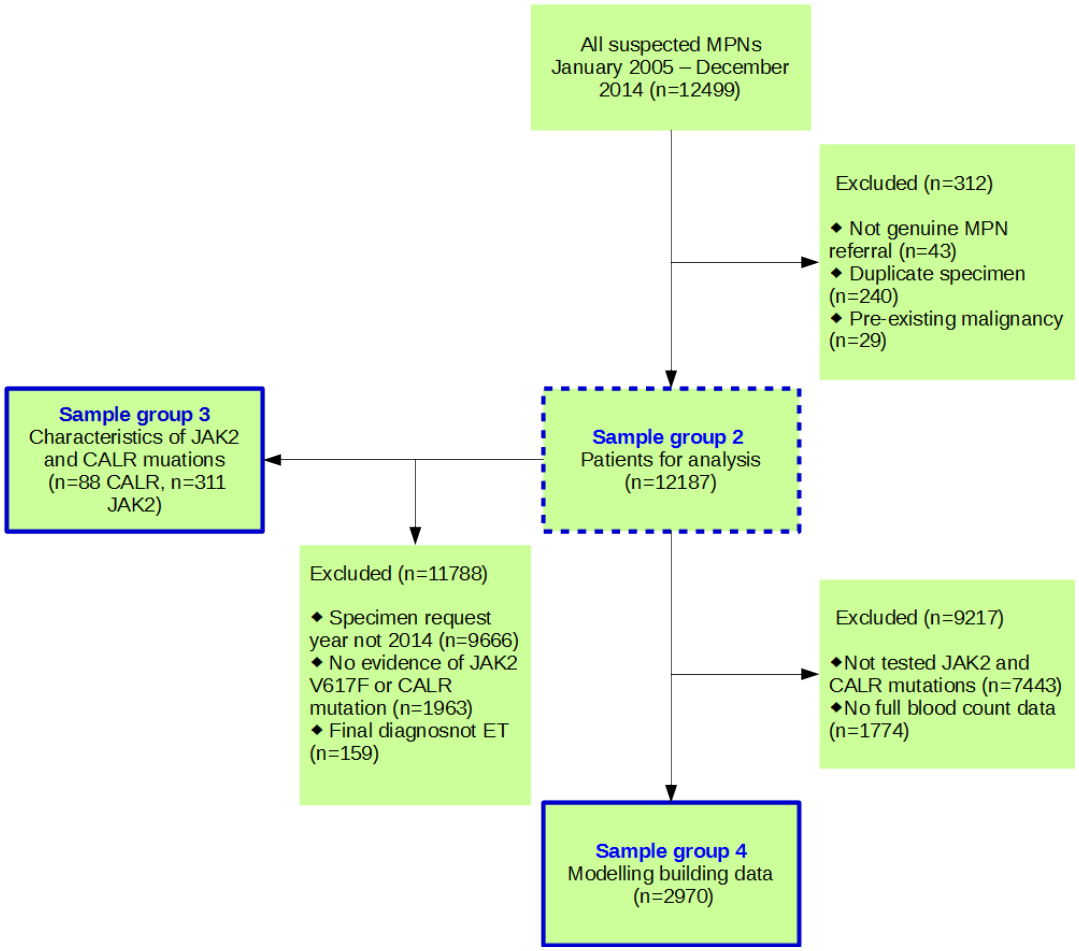

Figure 3-7. Summary of sample groups used in statistical analyses.

3.5.2.2 Sample group 3: Comparison of *JAK2* and *CALR* mutated ET

To compare the demographic and laboratory features associated with *CALR* mutation against those seen in *JAK2* V6147F mutated counterparts, diagnoses of ET with a *CALR* mutation identified between January and December 2014 (n=88) was compared to the *JAK2* V617F mutated ET diagnoses made during the same period (n=311).

3.5.2.3 Sample group 4: Predictive model development

This work aimed to develop a statistical model which could be used to aid in the investigation of suspected MPNs received by HMDS. Statistical models were developed using data collected from all suspected MPN referrals which had undergone both *JAK2* V617F and *CALR* mutational screening in HMDS with full blood count data available at the time of referral (n=2970).

3.5.2.4 Sample group 5: Model validation

Each of the statistical models developed using sample group 4, was then validated on a further dataset using samples received between January and March of 2015 (n=515) (see Figure 3-8).

3.5.2.5 Sample group 6: model assessment

Finally, the best performing model, was then applied to a third dataset, containing all suspected MPN requests received between January and December 2016 (n=2174) (see Figure 3-8).

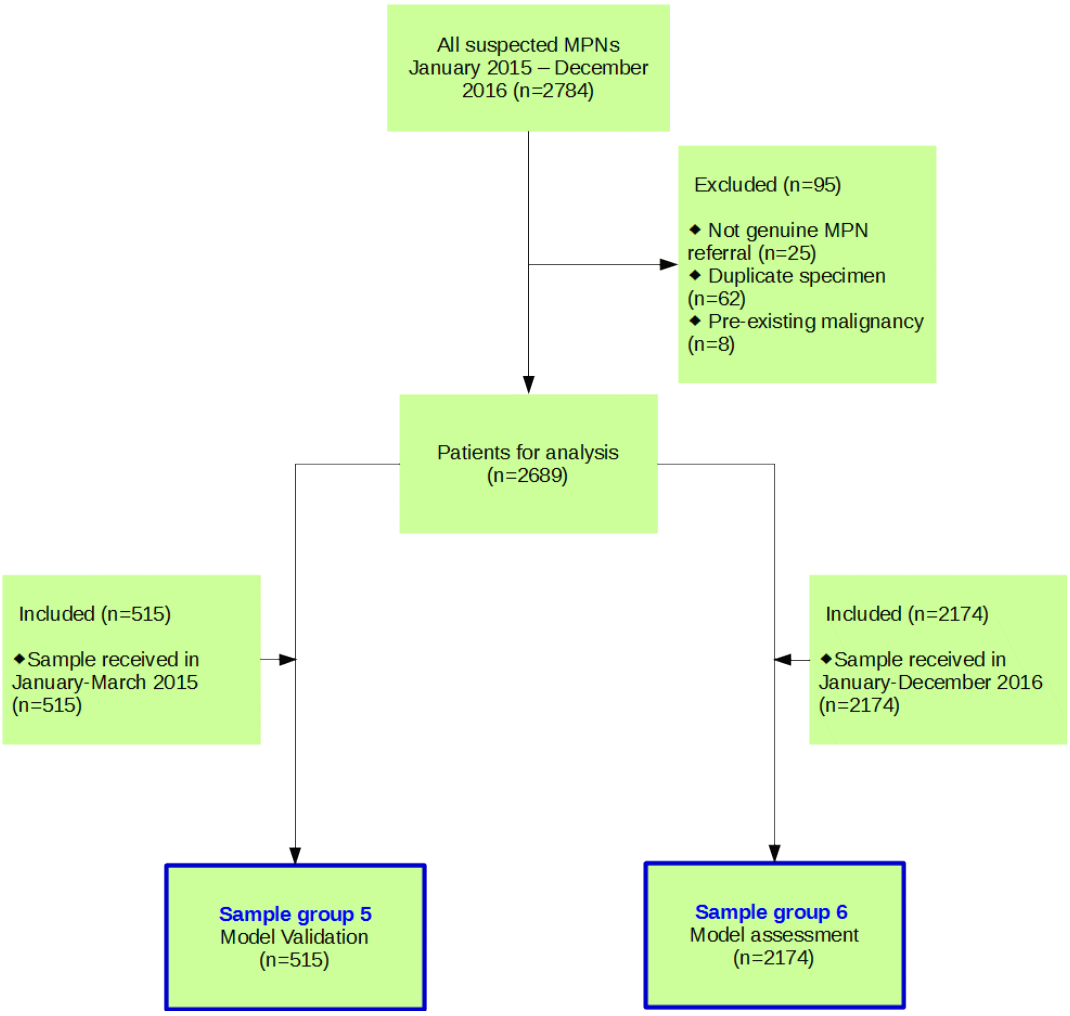

Figure 3-8. Overview of samples used in the validation and assessment of predictive modelling in the investigation of suspected MPNs.

3.6 Laboratory investigations

The laboratory methods described in this thesis were undertaken in the HMDS laboratory, Leeds, United Kingdom. Development of the fluorescent PCR and fragment analysis assay used in the determination of *JAK2* V617F and *CALR* mutations was undertaken by the author. Following development, the assay has been incorporated in routine laboratory practice and the data used in this thesis has been collected by all members of the HMDS Molecular Diagnostics team, including the author. Full details of reagents and protocols are available in the appendix.

3.6.1 Isolation of leucocytes

Leucocytes were isolated by selectively removing erythrocytes by differential lysis with 0.86% ammonium chloride solution (NH_4CL). A 1mL aliquot of Ethylenediaminetetraacetic acid (EDTA) anticoagulated peripheral blood or bone marrow was incubated with 9mL of ammonium chloride solution in a 13mL Röhren tube, which was then incubated for 10 minutes at room temperature. The suspension was then centrifuged for 4 minutes at a relative centrifugal force (RCF) of 470 and the supernatant discarded. The resulting cell pellet was washed twice with 10mL volumes of FACSFlow containing 0.3% Bovine Serum Albumin (BSA), centrifuging at 470 RCF for 4 minutes and discarding supernatant between washes. The washed cells were then re-suspended in 200µL of FACSFlow prior to genomic Deoxyribonucleic acid (DNA) extraction.

3.6.2 Genomic DNA extraction

Genomic DNA was extracted from the 200µL aliquot of leucocytes using the QIAamp DNA Mini Kit (QIAGEN Ltd, U.K.) according to the manufacturer's instructions. In this method, the aliquot of cells was added to a 1.5mL Eppendorf tube. Twenty microlitres of proteinase K (activity of 600U/mg protein) and 200µl of lysis buffer (buffer AL) were added together and the mixture was then incubated in a water bath at 56°C for 10 minutes. Following incubation, the samples were briefly centrifuged to remove droplets from the inside of the lid and 200µL of ethanol (96-100% concentration) was added and the sample mixed by pulse-vortex. The mixture was then applied to a QIAamp Spin Column and centrifuged for 1 minute at 6000 RCF to bind the DNA to the silica-gel column membrane. The supernatant was decanted and 500µL of wash buffer (AW1) applied to the column. The column was then centrifuged for 1 minute at 6000 RCF and the effluent discarded. A further 500µL of wash buffer (AW2) was added to the column, which was then centrifuged for 3 minutes at 18,400 RCF. The QIAamp Spin Column was then transferred into a fresh collection tube and 200µL of elution buffer (buffer AE) added. The column

was left to stand for 2 minutes, prior to centrifugation for 1 minute at 6000 RCF to elute the extracted genomic DNA into the filtrate. DNA was stored at room temperature prior to amplification.

3.6.3 *CALR* / *JAK2* V617F mutation screening

For *CALR* mutational screening, amplification of exon 9 is performed using a fluorescently labelled forward primer and unlabelled reverse primer. An allele specific oligonucleotide (ASO) polymerase chain reaction (PCR) assay is traditionally used for the detection of the *JAK2* V617F mutation, in combination with agarose gel electrophoresis. In this work, the ASO strategy has been modified by the author, to incorporate a fluorescent dye into the reverse primer. The exon 14 region of the gene was amplified using the fluorescently labelled reverse primer in combination with an unlabelled mutant specific forward primer and a consensus *JAK2* wild type forward primer. Full primer sequences and reagent mixes are shown in the appendix (Table 10-3Table 10-4). This allows the simultaneous visualisation of the PCR products from both the *CALR* and *JAK2* V617F assays using fragment analysis software. Amplifications were performed in separate PCR reactions and the resulting products were combined in equal volumes for analysis.

3.6.3.1 Thermal cycling conditions

Both PCR reactions were performed simultaneously using a unified programme. Cycling Conditions for *JAK2* V617F / *CALR* exon 9 PCR amplification are included in the appendix (Table 10-5).

3.6.4 Fluorescent fragment analysis

A 1µL aliquot of each of the paired PCR products was added to a 96 well microtitre plate along with 10µL of Hi-Di formamide solution (Applied Biosystems) and 1µL of Rox-500 size standard (Life Technologies). The microtitre plate was centrifuged briefly to ensure that samples were in the bottom of the well and then loaded onto the ABI PRISM 3130 Avant Genetic Analyser. Results were analysed using the generic microsatellite analysis protocol (ABI, Warrington, U.K.).

3.6.5 Expected results and interpretation

Data was interpreted using Genemapper Analysis Software (v3.1) (Applied Biosystems). Product peaks were measured in relation to the ROX-500 size standards to identify wild type and mutated PCR products. The expected size of both wild type and mutated PCR products are shown in Table 3-1.

Product	**Fragment size**
***JAK2* wild type**	362bp
***JAK2* V617F mutated**	202bp
***CALR* wild type**	292bp
***CALR* mutated**	Various - most common 240bp (52bp del) and 297 (5bp ins)

Table 3-1. Expected fragment sizes for *JAK2* V617F and *CALR* PCR products.

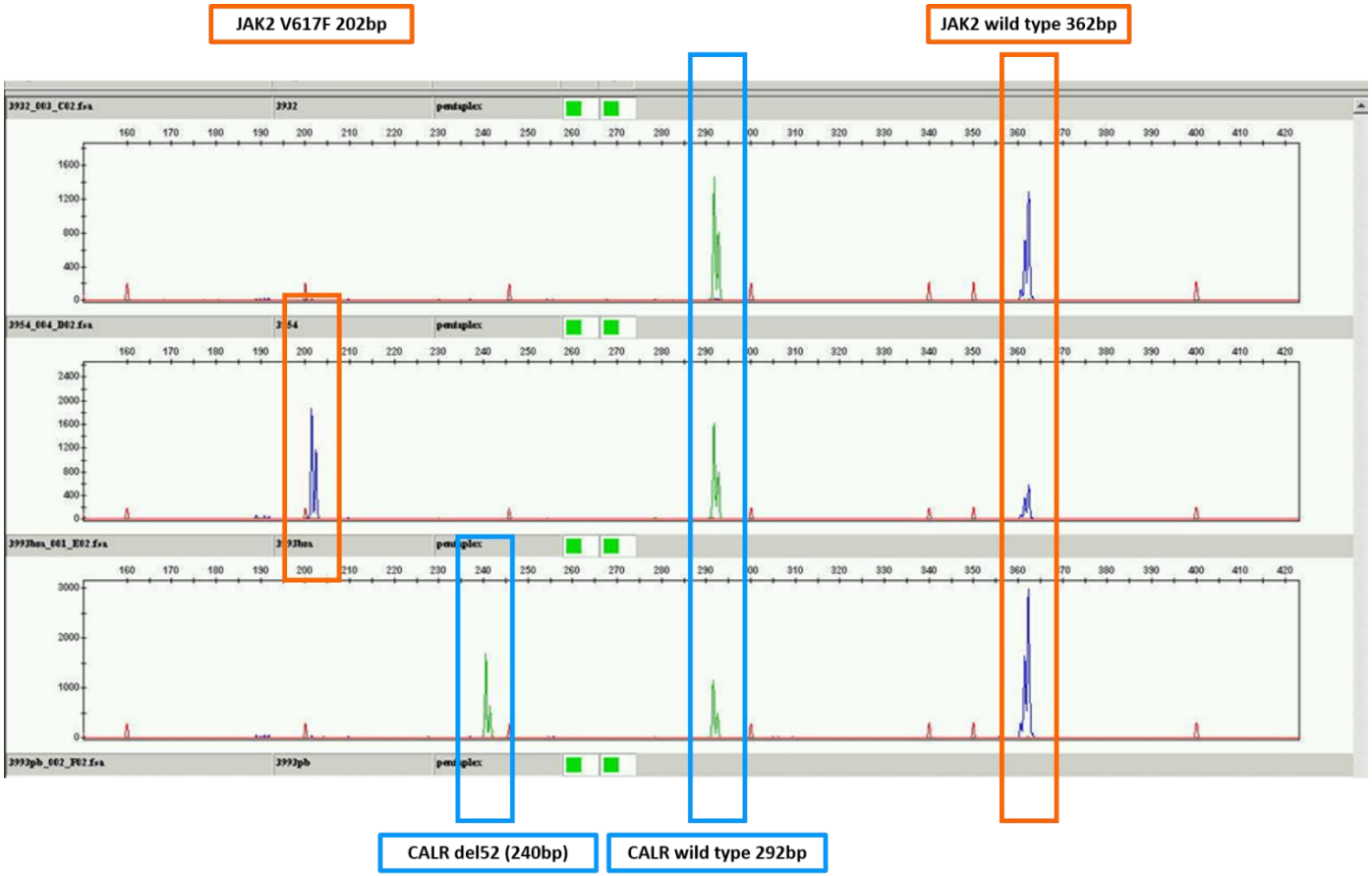

Figure 3-9. Electrophoretogram examples of *JAK2* and *CALR* mutational analysis. Genemapper software allows visualisation of PCR products according to their relative size. In this case the wild type JAK2 product is 362bp and V617F mutant is 202bp (shown as blue peaks); wild type CALR is 292bp and mutant products vary (the example shown is 240bp) (shown as green peaks).

Figure 3-9 shows examples of the results of this PCR assay. The top panel shows 1results from a specimen with wild type *JAK2* V617F and *CALR* amplification. The green peak at 292bp corresponds to the wild type *CALR* product and the blue peak at 362bp the wild type *JAK2* product, no further peaks are present. The middle plate shows a case with mutated *JAK2* V617F showing blue peaks representing both the wild type product at 362bp and a second peak at 202bp, the mutated product. The *CALR* gene is normal in this case with only the 292bp green peak present. The bottom plate shows a case with mutated *CALR*. The wild-type 292bp PCR product can be seen in green alongside a mutant peak measuring 240bp – this represents a 52bp gene deletion. The *JAK2* gene is normal in this case with only the 362bp peak present.

3.7 Statistical modelling methodologies

3.7.1 Basic analyses

The demographic and laboratory features of *CALR* mutated classical MPNs were compared against *JAK2* V617F mutated counterparts using standard statistical tests. Mann-Whitney U tests were used to establish whether statistically significant differences existed between groups. A p value of <0.05 was used to indicate significance in these analyses. The sensitivity and specificity of the current clinical guidelines were calculated using the metrics described below in 3.7.4

3.7.2 Approach

To determine which modelling approaches to test in this work a preliminary side by side comparison of a range of supervised predictive modelling methods was performed using the automated software, WEKA (Waikato Environment for Knowledge Analysis). This software platform contains a collection of machine

learning classifiers/algorithms which can be applied to user imported data sets. Using default settings, a set of 8 different supervised classification methods was applied to sample group 4 (see 3.5.2.3). Performance measures were compared between approaches and the three which gave the 'best' results were chosen for further development and tuning. These were logistic regression, random forest analysis and gradient boosted analysis.

3.7.3 Software

Further, detailed statistical analysis was undertaken in R studio using R version 3.5.1 (Team, 2018), using packages shown in Table 3-2.

PACKAGE	REFERENCE
SURVIVAL	(Therneau, 2015; Therneau and Grambsch, 2000)
MASS	(Venables and Ripley, 2002)
FARAWAY	(Faraway, 2014)
RGL	(Adler and Murdoch, 2015)
CARET	(Kuhn et al., 2015)
GMODELS	(Warnes et al., 2013)
ROCR	(Sing et al., 2005)
RMS	(Harrell Jnr, 2015)
RANDOMFOREST	(Liaw and Wiener, 2002)
GBM	(Ridgeway, 2015)
SHINY	(Chang et al., 2018)

Table 3-2. Statistical Packages used in data analysis.

3.7.3.1 Logistic regression

In this work, the specified outcome was a diagnosis of MPN, and the candidate explanatory variables were the following measured laboratory and demographic attributes:

- Age (years)
- Gender (male(M) or female(F))
- Platelet count (Plt) $x10^9$/L
- Haemoglobin (Hb) g/L
- Red cell count (RBC) $x10^6$/L
- Haematocrit (HCT) %
- Mean cell volume (MCV) fl
- White cell count (WBC) $x10^9$/L
- Neutrophils (Neutr) $x10^9$/L
- Lymphocytes (Lymph) $x10^9$/L
- Monocytes (Mono) $x10^9$/L
- Mean cell haemoglobin (MCH) pg
- Mean corpuscular haemoglobin concentration (MCHC) g/L

The explanatory variables were evaluated for collinearity using paired distribution plots and where identified one variable of the pair was chosen to be excluded based on biological information and frequency of missing data. The selected explanatory variables were also assessed for statistical significance using a Least Absolute Shrinkage and Selection Operator (LASSO) analysis. LASSO regression performs attribute selection to enhance the predictive accuracy of the model by shrinkage. This involves penalising the absolute size of the regression coefficients, and in doing so, some of the coefficients were shrunk to zero, at which point they were removed from the model.

Overfitting may occur when a statistical model is too complicated, contains larger numbers of variables than necessary and is consequently too specific to the dataset used to develop it. Whilst a model may perform exceptionally well during development, if overfitted to the data, it will perform poorly on unseen data (Hawkins 2004). Cross validation was performed to estimate the effect of overfitting within the logistic regression model. This approach, an example of a repeated holdout method, randomly divides the original dataset into a specified number of smaller datasets as shown in Figure 3-10 **(**Lantz 2013 page 319).

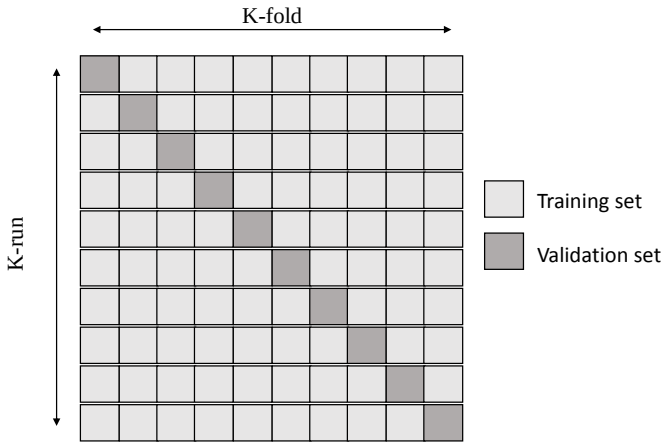

Figure 3-10. Schematic representation of cross-validation.

In this case the data was randomly divided into 10 new datasets. The regression analysis was repeated using 9 of the 10 data subsets to train the model and the remaining 1 used to validate the performance of the model. The logistic regression model tested on each of the datasets in turn. The statistical significance of each explanatory variable was measured for each sampling, and the overall effect of overfitting estimated.

3.7.3.2 Random forest analysis

The random forest is an example of an ensemble technique. It is based on the common decision tree method, where data is separated into branches depending upon how it responds to the set criteria (see example in Figure 3-11). Random forest utilised a plethora of randomly selected decision trees and combines their output to provide an overall prediction of outcome. Overfitting does not occur as the large number of decision trees provide better generalisation (Breiman 2001) In this case, 500 individual decision trees were created to build the random forest.

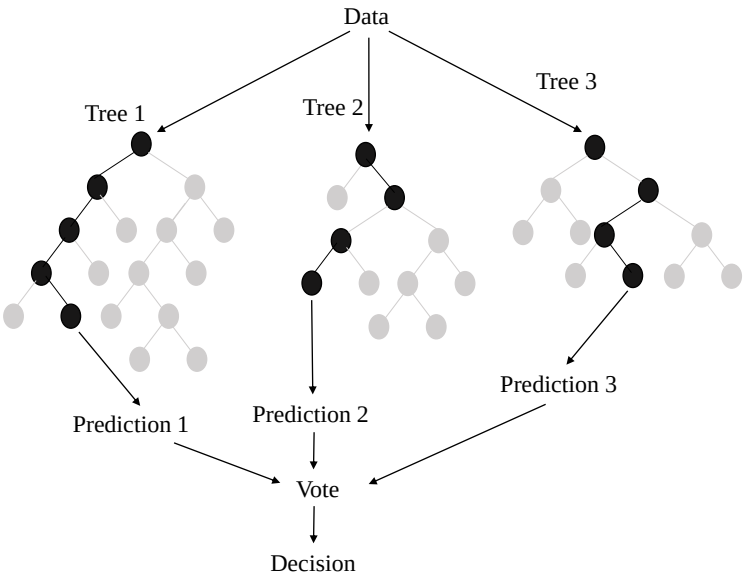

Figure 3-11. Random forest decision tree process.

Decision trees are created which bifurcate samples using simple threshold values and yes/no responses relating to the predictor variables, for example; is the patient male? (yes/no), is the patient over 50 years old? (yes/no).

3.7.3.3 Gradient-boost

Similarly to Random Forest analysis, Gradient Boost is an ensemble technique in which a series of weak classifiers, typically decision trees, are combined to produce an overall prediction. In this approach, each observation is initially assigned an equal weighting. When a model is applied to the data, misclassified observations gain weight and correct classifications lose weight. With the addition of each subsequent classifier there is increased focus on misclassified observations, due to their increasing weighting (Schapire and Fruend 2012). The overall effect is that each classifier contributes proportionally to the prediction based on its predictive accuracy.

3.7.4 Model performance evaluation

Each of the statistical models developed were then applied to the original dataset and the model's performance was tested. Several measurements were assessed to evaluate the model as shown in Table 3-3.

	Predicted MPN	**Predicted no mutation**	**Total**
Actual MPN	True Positive (TP)	False Negative (FN)	TP+FN
Actual no mutation	False Positive (FP)	True Negative (TN)	FP+TN
Total	TP+FP	FN+TN	

Table 3-3. Confusion Matrix.

Performance measurements are centred on measuring how often the model gets the correct answer. The focus is not only on how many times a positive case is correctly identified, but also how often a negative case is mistakenly called positive and visa-versa, a confusion matrix separates the predicted data into 4 categories when compared to the actual diagnostic status (Table 3-3). Depending upon the intended use of the model, it may be desirable to weight the model towards one of these categories. For example, if the consequences of missing a case were severe, there emphasis may be placed on minimising false negatives, which may have a consequential impact on the frequency of false positives predicted by the model. Using the categories, a series of metrics can be calculated which are reproducible between modelling approaches and allow different models to be directly compared. The performance measures and their calculations are shown in Table 3-4.

3.7.5 Model validation

Validation is a key step in the assessment of a statistical predictive model and essential prior to introduction into clinical practice. Ideally, the model should be applied to a 'new' dataset, that is, data which did not form part of the original dataset on which the model is built. This is especially important where a model has been built using a sample of data in which the proportion of 'MPNs' and 'No mutations' does not necessarily reflect the actual frequency of each group in the population to which the model will be applied. For these reasons, each model was applied to a test dataset and the performance measures reassessed to determine the applicability to 'real world' data (Figure 3-12).

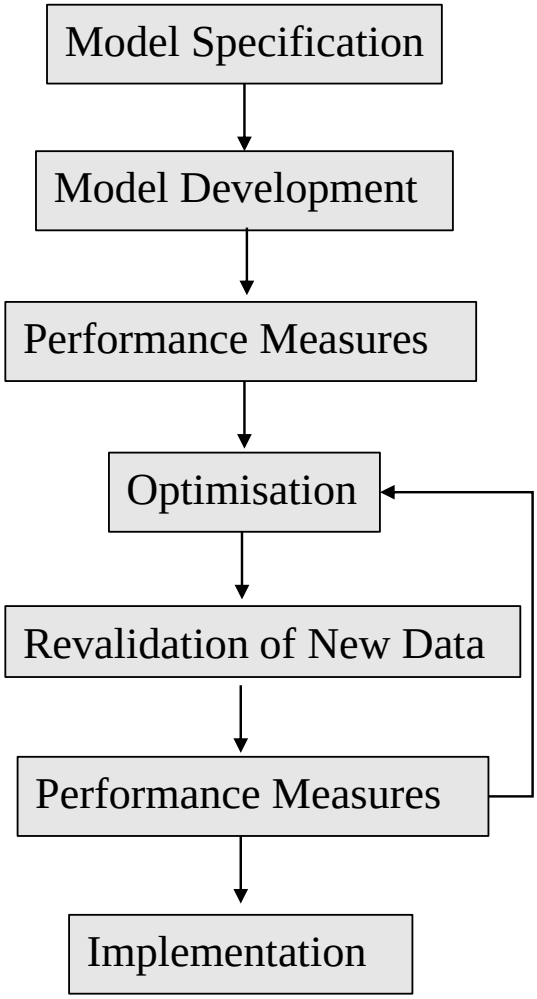

Figure 3-12. Process of model development and validation.

Measure	Description	Formula
Accuracy	Degree of closeness to the actual results from the predicted results	= (TP + TN)/(TP+FP+FN+TN)
Sensitivity	Proportion of actual cases predicted to be cases	= TP/(TP + FN)
Specificity	Proportion of controls predicted to be controls	= TN/(FP + TN)
Kappa Statistic	Compares the accuracy of the classifier compared to the random accuracy of the classifier	See (Cohen, 1960)
F-measure	A measure of the test's accuracy taking into account sensitivity and precision	= 2 x TP/(2 x TP + FP + FN)
Area under the Receiver Operating Curve (AUROC)	Reduces the ROC performance (FN and FP for every cut-off) to a single value thereby allowing classifier comparison	See (Fawcett, 2006)

Table 3-4. Standard performance characteristics and their formulae.

4. Development and validation of a screening assay for the simultaneous detection of *JAK2* V617F and *CALR* Exon 9 mutations.

4.1 Rationale and introduction

Molecular screening for *JAK2* V617F mutations was introduced into routine practice in HMDS shortly after it was described (Baxter et al., 2005), using an Allele-specific oligonucleotide (ASO) PCR technique based upon the method presented in the literature. *CALR* mutations were later identified through whole exome sequencing experiments. This technique is not suitable as a screening tool in routine diagnostic practice – due to the cost implications, processing and data analysis time. To be able to identify *CALR* mutations for the diagnosis of suspected MPNs in HMDS an alternative laboratory approach was required. To maximise sample processing efficiency, an approach which would be used to detect mutations both *JAK2* V617F and *CALR* mutations simultaneously was most desirable.

4.2 Current approach to detecting *JAK2* V617F mutations

At the outset of this work, *JAK2* V617F mutations were detected using an ASO-PCR assay. This technique utilised two different forwards primers, in conjunction with a consensus reverse primer as shown in Figure 4-1. The first was a consensus forward primer (a) binds upstream of the mutation point and will amplify independently of the mutational status. The second was a mutant specific forward primer (b), designed to terminate at the point of the mutation, this primer would only anneal and amplify in the presence of the mutant sequence. A consensus reverse primer was used to amplify both products.

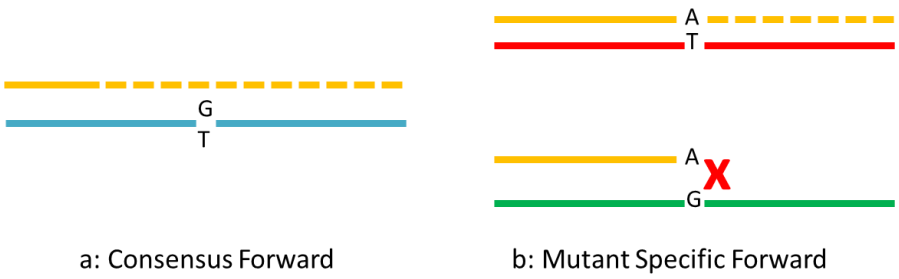

Figure 4-1. Schematic showing ASO-PCR.

Following amplification, the PCR products were visualised using agarose gel electrophoresis. The wild-type *JAK2* PCR product is 364bp in length, whereas the product produced by DNA carrying the *JAK2* V617F mutation is 203bp in length. This difference in size allowed easy distinction between the wild type and mutant *JAK2* PCR products.

ASO-PCR followed by agarose gel electrophoresis was not deemed to be a suitable approach for the detection of *CALR* mutations. Mutations within the exon 9 region of the *CALR* gene are varied and therefore a single mutant specific primer would not be applicable. The published data indicated that *CALR* mutations were always associated with a change in amplicon size, although the reported mutations varied in size and included compound mutations resulting in size changes of as little as 1bp (Chen et al., 2014; Klampfl et al., 2013; Nangalia et al., 2013). The resolution of electrophoresis of a 2% agarose gel is widely accepted to be in the region of 20-30bp. This limitation excluded gel electrophoresis from being utilised in this setting as a method of product visualisation.

4.3 Fragment size analysis for the detection of *CALR* exon 9 and *JAK2* V617F mutations

Both the existing approach to *JAK2* V617F analysis and amplification of the *CALR* exon 9 region would result in PCR products which differ in size between wild type and mutant DNA. For this reason, fragment size analysis was considered as a potentially suitable approach. The technique had been used extensively in our laboratory across a wide range of assays and was sufficiently sensitive to detect single base pair size changes as would be required for *CALR* mutation screening. The technique can be used to analyse multiple products simultaneously, to allow for both *JAK2* V617F and *CALR* mutations to be identified in a single assay.

The approach utilised fluorescent markers which were incorporated into the 5' end of one of the primers used to amplify the target DNA. The products were mixed with a size standard and analysed using capillary electrophoresis. Instrument software measured the size of the DNA fragments relative to the corresponding size standards and displayed this information for user interpretation.

Validation of a *JAK2* V617F / *CALR* mutation screening assay was critical. As a new diagnostic technique, extensive validation was undertaken prior to the introduction of this assay into routine laboratory use. Standards for the validation of new diagnostic techniques are outlined in the standards set out by ISO 15189 (BSI, 30 November 2012).

4.4 Primer selection

As the *JAK2* V617F assay was already well established, the same primer sequences were used for the fluorescent fragment analysis, with the addition of a 6-carboxyfluorescein (6-FAM) label at the 5'end of the consensus reverse primer. These primers retained a melting temperature (Tm) of 60˚C which allows the use of the existing thermal cycling conditions for amplification.

The genomic sequences described in the supplementary information from the original manuscript by Nangalia et al. (Nangalia et al., 2013) were used to design primers for the amplification of *CALR* exon 9. The appropriate genomic region was identified using the University of California Santa Cruz (UCSC) genome browser, BLAT search (http://genome.ucsc.edu/). Once identified the corresponding sequence information spanning the whole of exon 9 was retrieved and imported into the Primer 3 software package (http://simgene.com/Primer3). Primers were designed to have a Tm of 60˚C to allow amplification in the same reaction as those for *JAK2* V617F. A hexachlorofluorescein (HEX) label was incorporated into the 5' end of the *CALR* forward primer to enable visualisation.

Following design, the primers were checked to ensure suitability. Firstly, the primer sequence was checked to ensure that they would not bind to any other region of DNA using the Basic Local Alignment Search Tool (BLAST) from the National Centre for Biotechnology Information (NCBI) https://blast.ncbi.nlm.nih.gov/Blast.cgi). Neither the *CALR* not *JAK2* V617F primers were found to be homologous to any other region of human genomic DNA. The primer binding region of DNA was also checked to ensure that it did not contain a recognised single nucleotide polymorphism (SNP) site using the online SNPCheck database search facility (www.snpcheck.net/), which would potentially prevent binding. Again, neither set of primers was found to bind over a reported SNP site.

4.5 Confirming amplification of *CALR* exon 9 gene region

Firstly, it was demonstrated that the primers worked satisfactorily under the selected PCR conditions, and furthermore, were amplifying the correct region of DNA. To determine this, 6 DNA samples were selected from archived material and analysed using the assay described. Following amplification, the PCR products were visualised by agarose gel electrophoresis to determine the presence of the amplicon. Amplification was evident at a satisfactory level using the existing thermal cycling conditions and these were used for all subsequent work. The PCR products were analysed by Sanger sequencing and the resulting data compared to the corresponding *CALR* exon 9 gene region as detailed in the UCSC genome browser, BLAT search. The primers were found to be amplifying the appropriate region of DNA and were validated as being suitable for use.

4.6 Validating assay in *JAK2* wild type ET/PMF samples

4.6.1 Demographic details and diagnoses

A critical aspect of validation was to confirm that the assay would discriminate mutant *CALR* products from wild type. To achieve this, the assay needed to be performed on samples with varying *CALR* mutations. As no suitable reference material was available, the published data was used to identify a group of samples which would be most likely to have mutations present. The literature indicated that *CALR* mutations were found in a significant proportion of patients with *JAK2* V617F wild-type ET or PMF (Nangalia et al., 2013) and so a group of patients with these diagnoses were identified from laboratory records.

Sample group 1 (Figure 3-6 and section 3.5.1) was used to validate the assays ability to detect mutations in *CALR*. As described in 3.5.1, this group consisted of a cohort

of 48 patients with a confirmed diagnosis of ET or PMF, but where no demonstrable mutation in *JAK2* V617F or *MPL* had been identified. In the absence of such a mutation, this diagnosis could only be made by morphological assessment of bone marrow and therefore, archived DNA from bone marrow was available for testing. In addition, all 48 cases also had DNA stored from a peripheral blood sample received alongside the bone marrow sample. The patient characteristics are shown in Table 4-1 .

	ET	**PMF**
Number	42	6
Median age (years)	64.6	71
Male	23 (52.3%)	5 (83.3%)

Table 4-1. Summary of patients in sample group 1 - for the validation of *CALR* mutation screening assay.

4.6.2 Identification of *CALR* mutations

DNA from all samples was amplified with both *CALR* and *JAK2* primers in a single PCR reaction. The fluorescent PCR products were analysed using Genemapper fragment analysis software (ABI, ThermoFisher). *CALR* mutations were identified in a total of 24 patients (50%). The characteristics of these mutations are shown in Table 4-2, with corresponding electrophoretograms shown in Figure 4-2. Of note, the most frequently seen mutation was the 52bp deletion, followed by the 5bp insertion – which is in keeping with published data (Chen et al., 2014; Klampfl et al., 2013b; Nangalia et al., 2013).

Complete concordance was seen between peripheral blood and bone marrow samples for the same patient. The majority of referrals for the investigation of a suspected MPN are peripheral blood samples and *JAK2* V617F screening has been routinely performed on peripheral blood since its introduction. The ability to perform *CALR*

mutation screening on the same specimen was a key aspect of incorporating the assay into routine laboratory use.

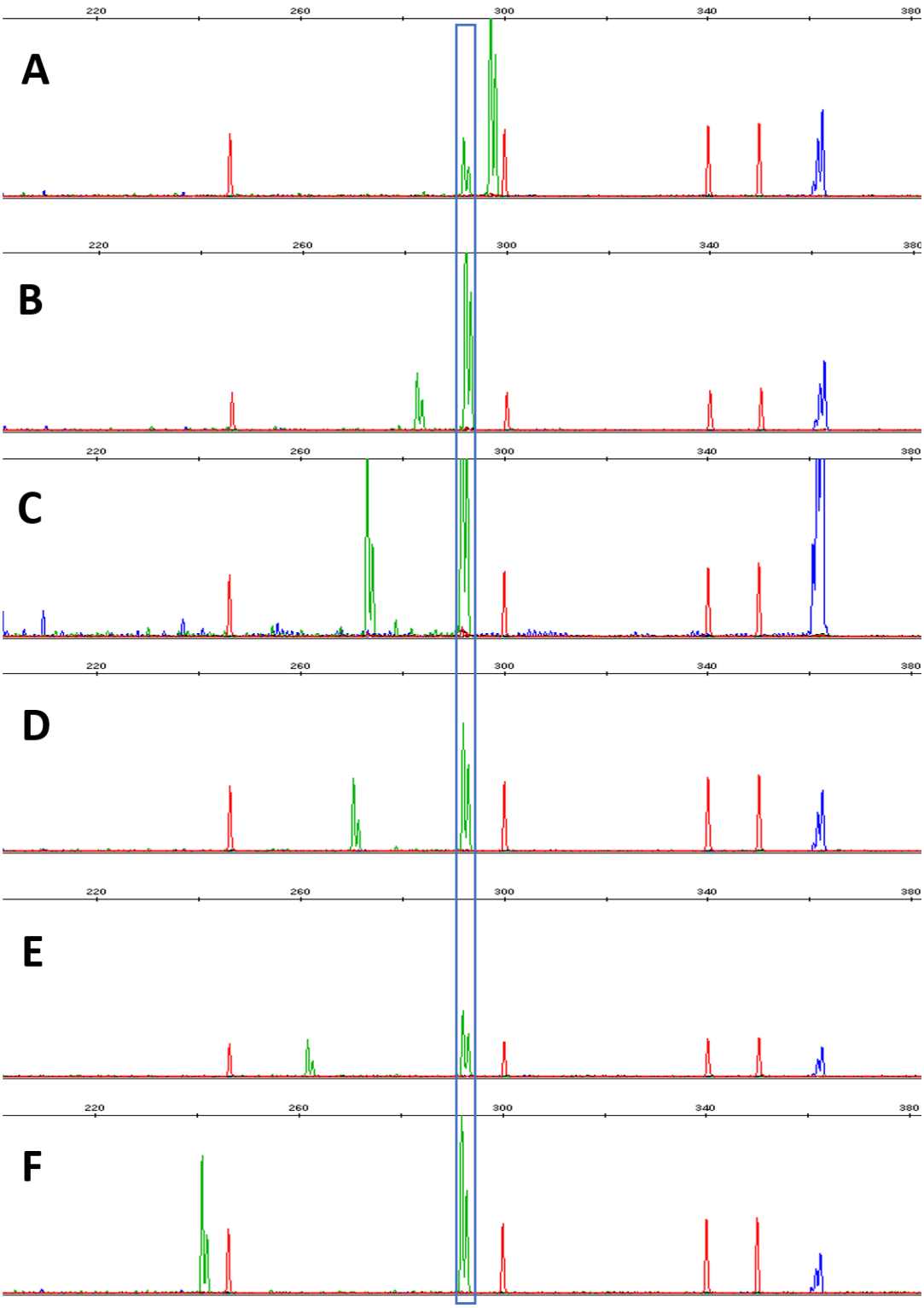

Figure 4-2 Results of fluorescent fragment analysis obtained during assay validation. Each panel showing a distinct mutation, the wild-type CALR peak is shown in the highlighted box. The second green peak shows the mutated CALR PCR product and its relative size (number identified shown in brackets).

(A) 5bp insertion (n=8), (B) 10bp deletion (n=1), (C) 20bp deletion (n=1), (D) 22bp deletion (n=1), (E) 32bp deletion (n=1), (F) 52bp deletion (n=12).

Mutation	Number of cases	Diagnosis		Gender		Median age (years)
		ET	PMF	M	F	
5bp insertion	8	7	1	4	4	64.54
10 bp deletion	1	1	0	1	0	64.30
20 bp deletion	1	1	0	1	0	62.06
22 bp deletion	1	0	1	1	0	63.02
32 bp deletion	1	0	1	1	0	47.63
52 bp deletion	12	11	0	8	4	67.78

Table 4-2. Characteristics of *CALR* mutations in the validation group (sample group 1).

4.6.3 Increased sensitivity of *JAK2* V617F assay

This innovative approach showed a higher level of sensitivity over the existing assay. All samples used in the validation assay had been originally reported as *JAK2* wild type. Following analysis using the fluorescent fragment analysis approach, one case showed the presence of a *JAK2* V617F mutation in the DNA from both the peripheral blood and bone marrow sample. The sensitivity of agarose gel electrophoresis had been previously estimated to be 2-3% allelic burden in our hands, whereas fluorescent fragment analysis is estimated to be capable of identifying a much lower level of mutated DNA, ~1% allelic burden. The increased sensitivity of the fluorescent assay would account for this discordant result and as it was expected that clinicians would want existing patients to be investigated for *CALR* mutations, this would be taken into consideration when analysing and reporting assay results.

4.6.4 Inter-laboratory validation

HMDS has long standing collaborations with several groups within the UK, including one with Professor N Cross of the Wessex Regional Genetics Service, Salisbury, UK. A series of 322 DNA samples from *JAK2* V617F wild type MPN patients diagnosed by HMDS had been provided by for Genome Wide Association Studies (GWAS) in Salisbury. As the group performing this analysis had also characterised the *CALR* mutation (Nangalia et al., 2013) the samples provided by HMDS had also been analysed for the presence of *CALR* mutations. This included 30/48 patients selected for test validation. The results of mutational screening in HMDS was 100% concordant with the findings from the GWAS.

A sample exchange was requested by Dr I Carter from Nottingham City Hospital who was in the process of setting up an assay using the same *CALR* primers and a similar approach. DNA samples from 20 patients were exchanged between our two centres. Concordance was 97.5% (39/40 cases), with Dr Carter detecting a 5bp insertion in one sample which was not detected using the assay developed in this work. Following discussions, it was found that there was a small difference between our techniques. In Nottingham, the *JAK2* V617F and *CALR* amplifications were being performed in two separate PCR reactions, whereas I had amplified both targets in a single multiplex reaction. It was hypothesised that a multiplex approach may have a reduced sensitivity compared with single target amplification. The mutation was detectable when the two targets were reamplified separately and mixed prior to fragment analysis and therefore, this approach was adopted for all future testing.

4.7 Discussion

Following their identification, the importance of being able to screen for *CALR* exon 9 mutations was considered a priority from a diagnostic point of view. The varied mutations found within the *CALR* exon 9 gene region, restricted the type of approach that could be used to identify them in routine laboratory practice. However, the consistent feature of altered PCR product size made it a model candidate for a fluorescent fragment analysis approach. As the existing assay used to detect *JAK2* V617F mutation was also associated with a size change, there was opportunity to develop a technique for the simultaneous detection of both abnormalities. This approach was appealing as it would maximise productivity, whilst minimising bench work and processing times/costs.

Internal laboratory validation demonstrated that *CALR* exon 9 mutations are consistently detectable in both peripheral blood and bone marrow samples from affected individuals, this offers a minimally invasive approach for patients undergoing screening for suspected MPNs. The fluorescent fragment analysis assay also offers an increased sensitivity over the existing assay for detecting *JAK2* V617F. This presents the possibility that patients previously diagnosed as being *JAK2* V617F wild type could be found to be mutated if rescreened using this approach.

External validation demonstrated almost complete concordance between results from two independent sample exchanges. DNA from one sample was found to have a demonstrable *CALR* exon 9 mutation when analysed in Nottingham, which was not identified in Leeds. A comparison of assay protocol identified a single difference in approach, with the laboratory in Nottingham performing separate PCR reactions to amplify the *JAK2* V617F and *CALR* exon 9 regions rather than the multiplex approach I had used. When the assay was repeated using 2 separate PCR reactions as per the approach used in Nottingham, the mutation was detectable.

This discrepancy was thought likely to be due to a reduction in the sensitivity of the two assays by amplifying them together and as a result, separate assays were used from this point onwards. This highlights the importance of performing thorough validation of an assay prior to its introduction into routine laboratory use.

5. Assessing current practice in the referral and diagnosis of suspected classical MPNs

5.1 Rationale and overview

The diagnosis of MPN can be made difficult by the significant overlap in clinical and laboratory features with other reactive conditions. Current guidelines indicate that the clinical features should be confirmed as being persistent and reactive causes (such as infection or inflammation) excluded prior to investigation (Harrison et al., 2010; McMullin et al., 2007; Reilly et al., 2012a). There is no published data to indicate how strictly clinical guidelines are adhered to, nor the efficiency of the referral process as it occurs in current practice. The high frequency of genetic aberrations in classical MPNs is a valuable indicator of disease where detected; however, the absence of a recognised mutation is not sufficient to exclude the presence of disease. The impact that the availability of molecular screening assays has had on referral behaviour is currently unknown.

The purpose of this chapter is to evaluate the investigation of suspected classical MPNs in a regional diagnostic setting. Extensive SQL searches were performed using the HILIS database to investigate how efficient current practice is in the identification of patients with MPNs. The laboratory parameters of each referral were compared against current guidelines to identify the proportion of referrals which met the criteria for investigation. These data were used to analyse the affect that the introduction of molecular screening assays have had on both referral behaviour and frequency of diagnosis.

5.2 How many patients are referred to HMDS for the investigation of a suspected classical MPN per annum?

The number of referrals for suspected classical MPN has increased annually over the past decade. The *JAK2* V617F mutation was identified in 2005 with the first publications available in July of that year. Routine mutational screening was introduced by HMDS in late October 2005. The total number of referrals for suspected classical MPNs are shown for a 10-year period commencing from 2005 (Figure 5-1). In 2005, a total of 356 referrals were received for the investigation of a suspected classical MPN, equating to 2.4% of the laboratory's total workload. This has shown a continued yearly increase and, in 2014 (at the commencement of this analysis), this figure stood at a total of 2114 referrals, accounting for 7.2% of samples.

This increased number of referrals has not been accompanied by a comparable increase in diagnoses. Whilst, the number of referrals has increased almost 6-fold over the past decade, the number of confirmed diagnoses of classical MPN has remained consistent with an average of 314 new cases per annum. This figure is in keeping with the estimate published by HMRN (Smith et al., 2010). Prior to the introduction of *JAK2* V61F screening, the proportion of referrals which resulted in a confirmed diagnosis of a classical MPN was 56% (201/356 cases received in 2005). Following the introduction of mutational screening for *JAK2* V617F, this figure has declined with an average of 26.3% of referrals resulting in a confirmed diagnosis of a classical MPN per annum (range 20.2-37.9% per annum) in the period from 2006-2013 inclusive.

The identification of *CALR* mutations and subsequent introduction of *CALR* mutational screening has led to a further increase in the number of referrals received. Following the routine implementation of *CALR* mutation screening in early 2014, there was a significant rise in the number of referrals for investigation, with a 32.8% increase over the previous year. However, the total proportion of referrals which resulted in a confirmed diagnosis of MPN declined further, to 19.0%, the lowest value to date. (Figure 5-1).

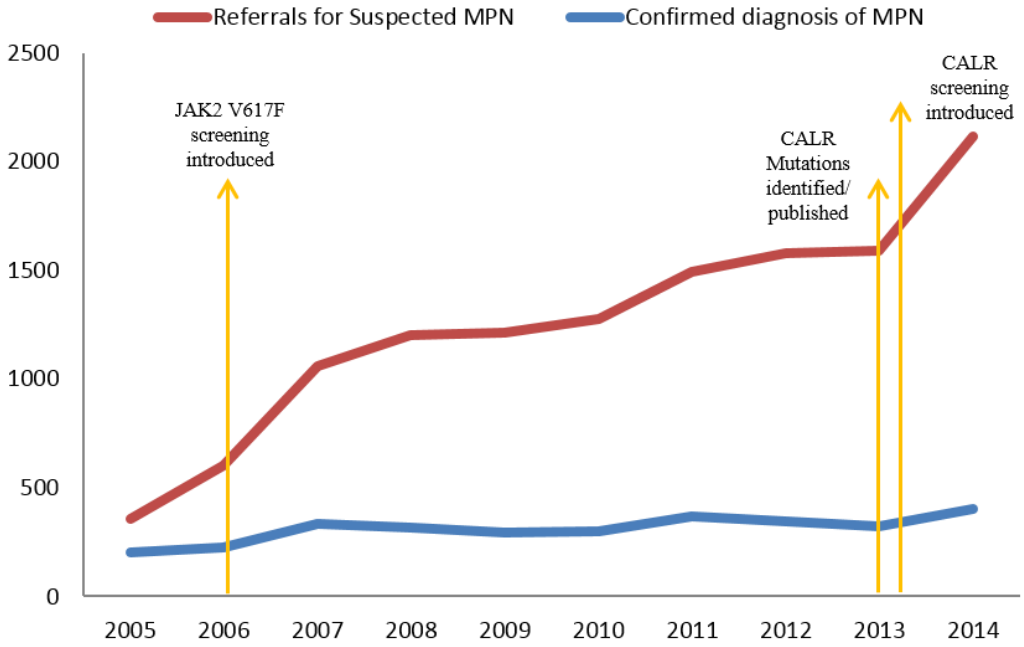

Figure 5-1. Annual number of referrals for suspect MPN received by HMDS 2005-2014.

5.3 How frequently are bone marrow investigations performed on suspected classical MPNs when a *JAK2* V617F or *CALR* mutation is not been detected?

JAK2 V617F and *CALR* mutations are shown to be present in a significant proportion of patients diagnosed with classical MPNs (Chen et al., 2014; Klampfl et al., 2013b; Nangalia et al., 2013). However, the absence of these mutations cannot be used to exclude the presence of disease. In both the WHO and BCSH recommendations, morphological assessment of a bone marrow aspirate is required to conclusively diagnose or exclude disease in patients where clonal markers are not detected (Arber et al., 2016; Harrison et al., 2010; McMullin et al., 2007; Reilly et al., 2012a). Using sample group 2, which consisted of all referrals of suspected classical MPNs received by HMDS between 2005 and the end of 2014 (see 3.5.2.1 for further details), it was possible to identify patients who had undergone bone marrow assessment following receipt of a blood sample which did not have a detectable *JAK2* V617F or *CALR* mutation. On average 178 bone marrow samples are received annually on patients with suspected classical MPNs, where a peripheral blood sample did not show a mutation (range 119-239).

As a proportion of the peripheral bloods received which do not have a detectable *JAK2* V617F or CALR mutation reported each year this represents an average of 19.01%. However, as Figure 5-2 illustrates there is ongoing divergence of the two groups and this proportion has declined annually from 31.82% in 2006, to 10.86% by the end of 2014.

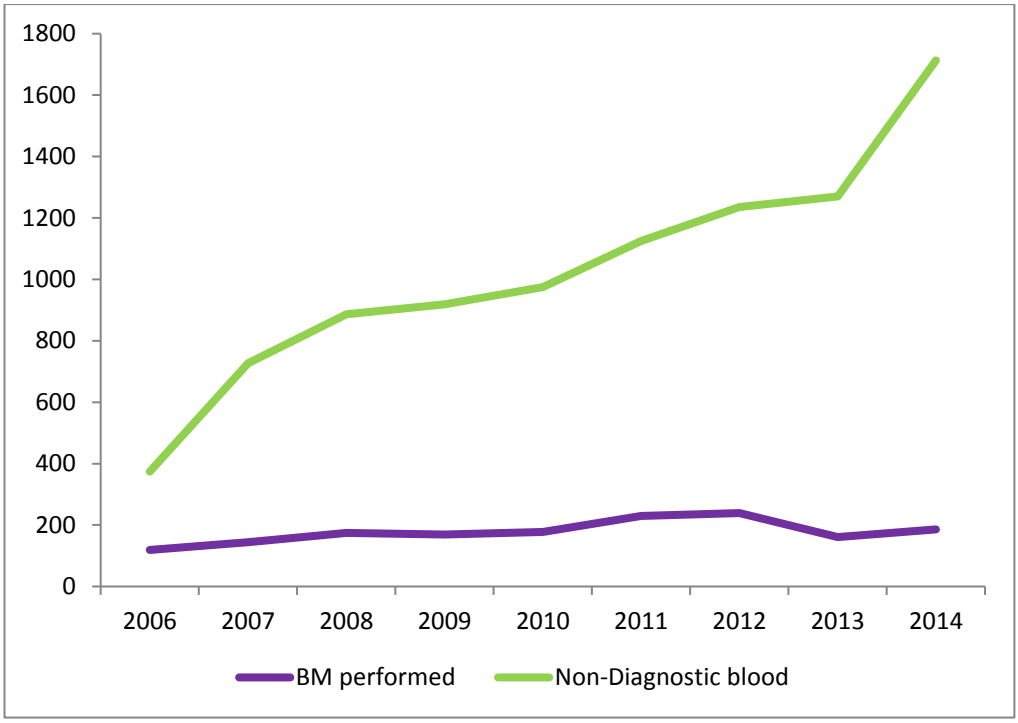

Figure 5-2 Number of bone marrow aspirates received following a peripheral blood sample with no detectable *JAK2* V617F or *CALR* mutation, compared with the total number of non-diagnostic blood samples referred annually from patients with suspected classical MPNs.

BM sampling was performed most frequently in cases of suspected ET (69.39%) with suspected PV and PMF accounting for 8.16% and 3.06% of BM referrals respectively. The suspected diagnosis was not clear in the remaining 19.39% of BM samples received. Haemoglobin and platelet counts were compared between *JAK2* V617F/*CALR* exon 9 wild-type referrals, separated into 2 groups on the basis of whether a follow-up BM sample was received. There were statistically significant differences in both parameters between the two groups (boxplots are shown in Figure 5-3), with patients who underwent subsequent bone marrow sampling have a lower haemoglobin (mean 134g/L in vs 155g/L in PB only group) and higher platelet counts (mean 666 $x10^9$/L vs 329 $x10^9$/L in PB only group). *p* values were reported as <0.001 in both analyses.

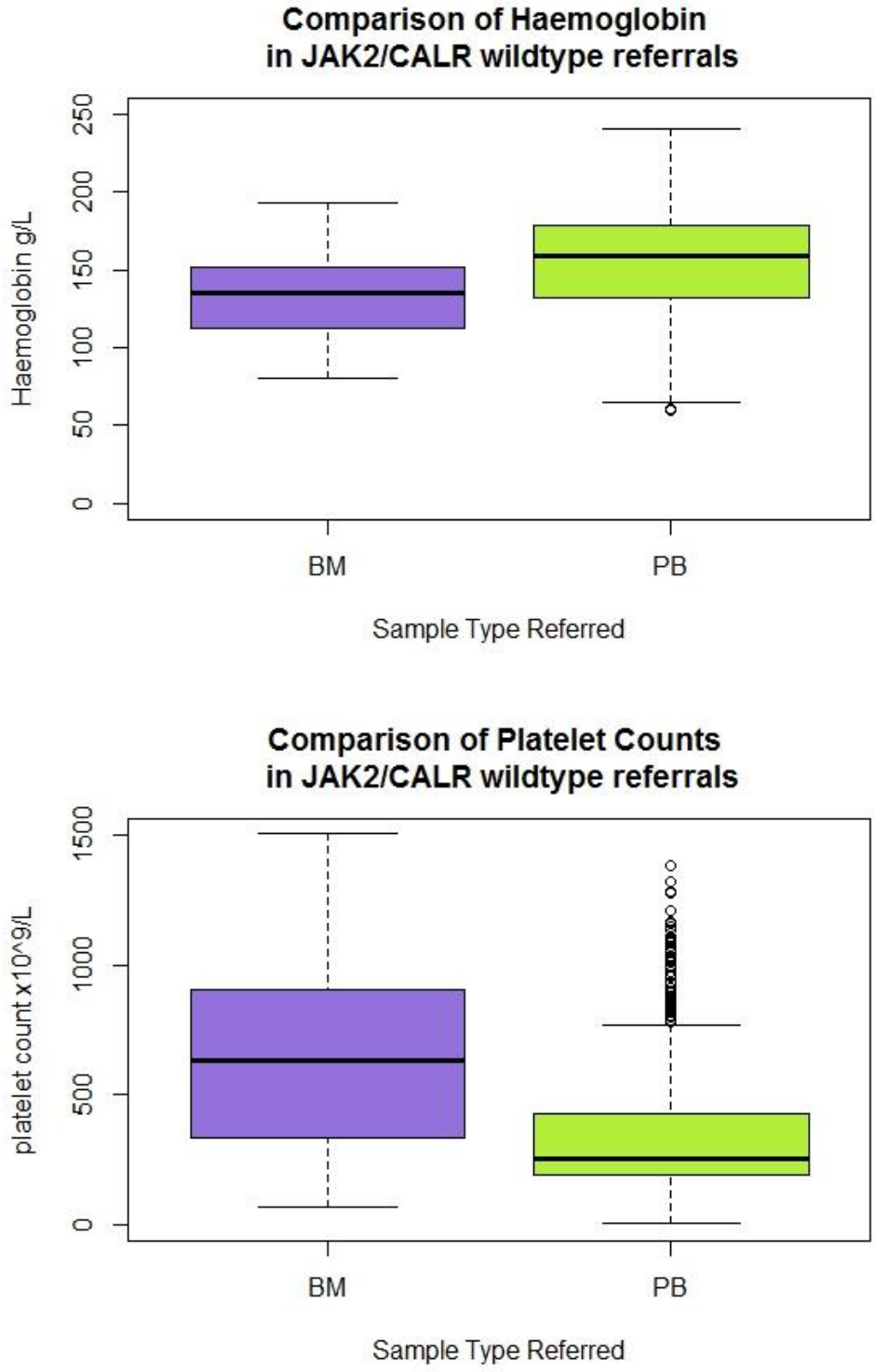

Figure 5-3. Significant differences in haemoglobin and platelet count were observed between *JAK2* V617F/*CALR* exon 9 wild type referrals which underwent subsequent bone marrow investigations versus those with peripheral blood screening only.

5.4 What proportion of referrals for investigation meet laboratory criteria set out in clinical guidelines for investigation?

The low proportion of referrals which result in a confirmed diagnosis of MPN would indicate that there are inefficiencies somewhere within the referral process, although this has not been formally audited to identify where or why. One potential cause could be that referrals are made which do not meet the recommendations set out in clinical guidelines. To ascertain whether this is the case, full blood count parameters and clinical details were compared to the recommendations set out both in the BCSH guidelines as well as in the WHO recommendations in place at the time of referral. This analysis was performed on 4576 referrals received during three consecutive years (2012-2014 inclusive) selected from sample group 2. This time-frame reflects the introduction of routine recording of full blood count data into the HILIS database. Prior to this time, electronic records were not available for the full blood count performed in HMDS.

5.4.1 BCSH guidelines

Each sample was categorised as either meeting or failing to meet the laboratory elements of the BCSH recommendations (Harrison et al., 2014; McMullin et al., 2007; Reilly et al., 2012) on the basis of meeting one of the following 3 criteria:

- Haematocrit >48 in females, >52 in males
- Platelet count >450x10^9/L
- Clinical suspicion of PMF

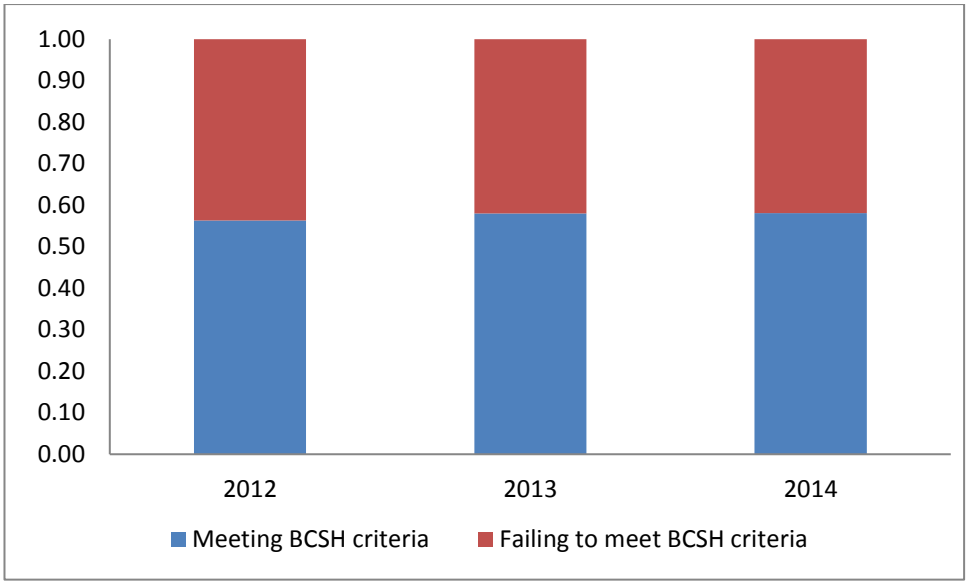

Figure 5-4. Proportion of referrals for suspected MPN received by HMDS during 2012-2014 which meet / fail to meet BCSH recommended laboratory criteria for investigation.

The proportion of referrals with full blood count indices and/or clinical details which would support investigation according to BCSH published guidelines averages at 57.48% of annual referrals (range 56.34-58.08%) over the 3-year period 2012-2014 (Figure 5-4).

5.4.2 WHO criteria

Full blood count indices and clinical details were then compared to the 2008 WHO recommendations (Barbui et al., 2015; Vardiman et al., 2009b) and each referral categorised as either meeting or failing to meet the criteria on the basis of the following parameters:

- Haemoglobin >165g/L
- Haemoglobin >185g/L
- Platelet count $\geq 450 \times 10^9$/L
- Clinical suspicion of PMF

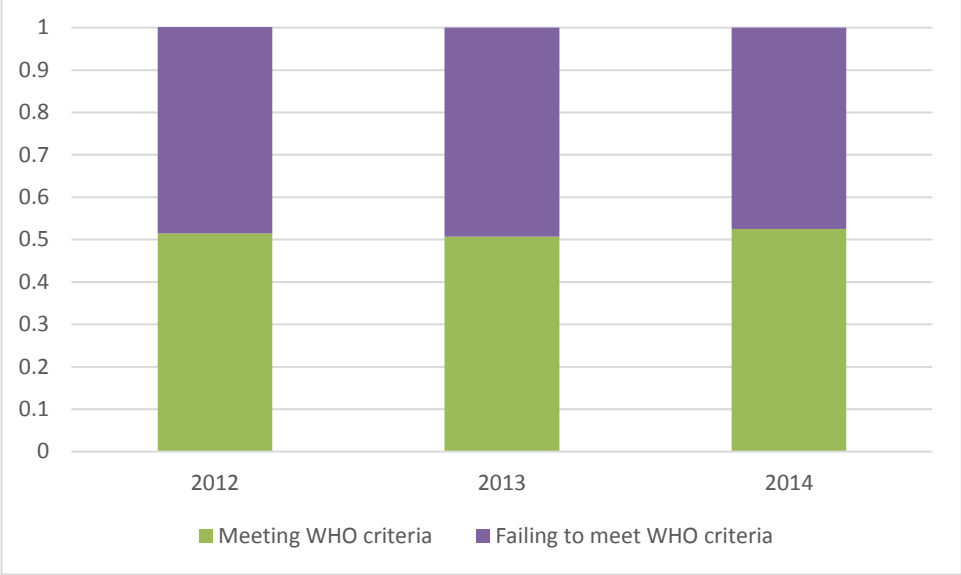

Figure 5-5. Proportion of referrals for suspected MPN received by HMDS during 2012-2014 which meet / fail to meet WHO recommended laboratory criteria for investigation.

The proportion of referrals meeting the laboratory criteria published by the WHO was consistently lower than that meeting the BCSH criteria. On average 51.15% of referrals had features which would support investigation for a suspected classical MPN (range 50.67.-52.52%) as show in Figure 5-5.

5.5 Are reactive causes excluded and symptoms demonstrated to be persistent at the time of referral?

A cohort of the first 50 consecutive referrals from Leeds Teaching Hospitals NHS Trust (LTHT) made in 2014 were selected from sample group 2 and clinical records audited to ascertain whether reactive and secondary causes of symptoms had been excluded prior to molecular investigation, as per the recommendations set out by the BCSH and WHO.

Twenty three of these 50 referrals had abnormal blood count results on at least 2 consecutive samples prior to referral, thus meeting the recommended criteria for investigation. In the remaining 27 referrals, abnormal results were only recorded on 1 sample prior to investigation. Eight out of the 50 referrals were found to have an MPN (6/8 with a *JAK2* mutation, 1/8 with a *CALR* mutation and 1/8 with no identifiable genetic abnormality) of which 7/8 had abnormal blood count results prior to referral.

A reactive or secondary cause for their laboratory features was noted in the clinical history in 35/40 patients who did not have a detectable mutation. These included a pre-existing medical condition (cardiovascular disease (n=6), liver disease (n=3), surgery (n=3), carcinoma (n=1) and rheumatoid disorder (n=1)) or elevated inflammatory markers (n=21). Of these 40 patients only one subsequently had a bone marrow which excluded the diagnosis of an MPN.

In patients with raised inflammatory markers, this test had been performed prior to being seen in haematology in 13/21 patients, indicating that this information was available at the time of clinical assessment. In the remaining 8 patients, tests for the presence of inflammatory markers were requested alongside or after the request for molecular screening.

5.6 How sensitive and specific are the current diagnostic guidelines in the identification of classical MPNs?

A significant proportion of referrals did not meet any of the laboratory criteria set out in the published guidelines at the time of referral, however, it is important to establish how effective the current guidelines themselves are in positively identifying patients with classical MPNs. To ascertain this, 4576 specimens received between January 2012 and December 2014 (taken from sample group 2) were divided into four groups based on whether they a) met the criteria set for investigation/diagnosis and b) whether they were diagnosed with a classical MPN on the basis of peripheral blood analysis (that is, having a JAK2 V617F or CALR mutation). The data is presented in 2 confusion matrices shown below (Table 5-1 and Table 5-2).

The platelet count used in the diagnosis of ET has remained unchanged between the 2008 and 2016 revisions of the WHO diagnostic criteria, however, the thresholds used to diagnose PV have been modified. The upper threshold for haemoglobin levels has been reduced to 160g/dL in females and 165g/dL in males. An upper limit has also been introduced for haematocrit levels, 48 in females and 49 in males. In this analysis the thresholds in use at the time of referral have been used (WHO 2008).

Using the calculations detailed in Table 3-4, the performance of each set of guidelines was calculated. When the laboratory thresholds detailed in the BCSH guidelines are applied to the data set, the criteria have an average sensitivity of 86.96% compared with 88.07% when the WHO guidelines are applied. The specificity of the BCSH guidelines is 49.55% compared with 33.65% with the WHO guidelines.

	Met BCSH criteria	Did not meet BCSH criteria
MPN	780	117
No mutation	1856	1823

Table 5-1 Confusion matrix of referrals according to BCSH guidelines.

	Met WHO criteria	Did not meet WHO criteria
MPN	790	107
No mutation	2441	1238

Table 5-2. Confusion matrix of referrals according to WHO criteria.

5.7 Can we identify other markers that discriminate patients with classical MPN from those without?

5.7.1 Basic Demographic features

It is well documented that the frequency of the classical MPNs increases with age and shows a slight bias toward female subjects (Roman et al., 2016; Smith et al., 2010; Visser et al., 2012). However, neither age nor gender are included in published guidelines. Age is widely used to inform malignancy screening in a number of diseases including breast, bowel and cervical cancer (England, 2016). These two features were recorded for all referrals and compared between those with a confirmed diagnosis of MPN and those with non-diagnostic results.

5.7.1.1 Age of patients at time of referral

Sample group 2 contains a total of 12187 specimens referred to HMDS for the investigation of a suspected classical MPN, received between January 2005 and December 2014. The sample group was separated into two categories on the basis of diagnostic outcome – those in which a mutation of *JAK2* V617F or *CALR* was found, and those in which no mutation of *JAK2* V617F or *CALR* was detected. Age distribution was plotted as a histogram with the two groups highlighted in different colours in order to visualise any overlap (Figure 5-6). Full details of R script used is shown in appendix section 10.2 . A difference in the referral ages was evident between the two categories. A T-test was performed to confirm that this difference was statistically significant (Table 5-3).

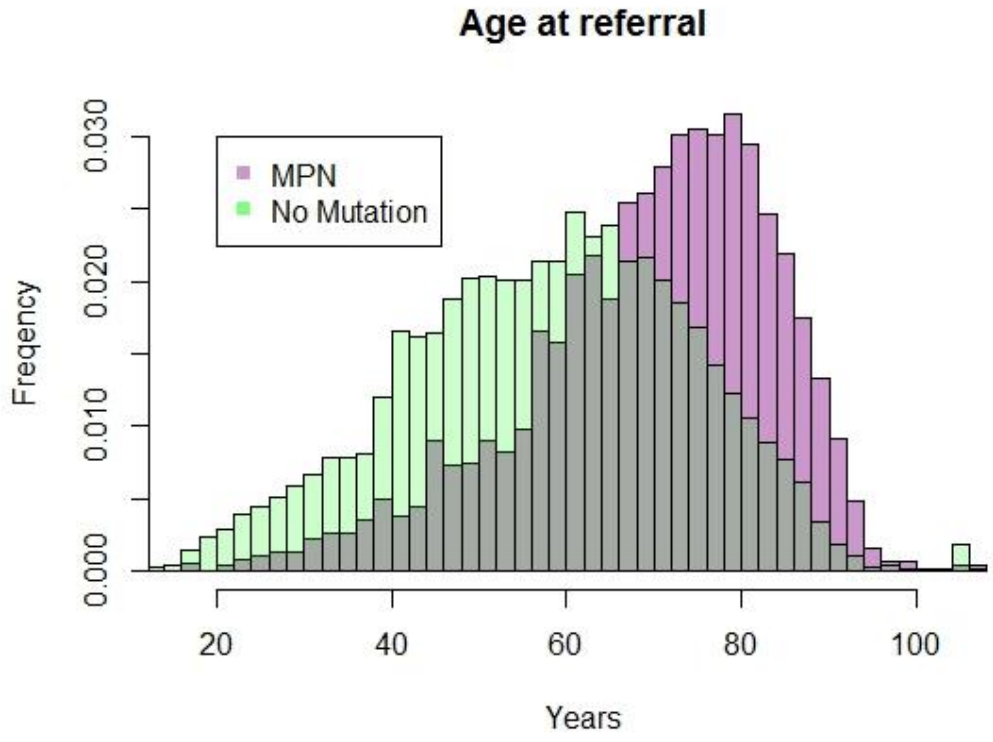

Figure 5-6 Comparison of age distribution between referrals with a diagnostic outcome of MPN and those without.

Category	N	Mean age (years)	T	DF	Significance (p)

MPN	2976	69.49	36.857	5776.3	<0.001
No mutation	9211	58.05			

Table 5-3. Statistical comparison of Age between diagnostic categories.

5.7.1.2 Red cell indices

In addition to haemoglobin (Hb) and haematocrit (HCT), full blood count analysers also commonly measure the total red cell count (RBC) and mean red cell volume (MCV) amongst others. Using the cohort of referrals described in section 5.4, these indices were compared between samples with a mutation in either *JAK2* V617F or *CALR* (referred to as "MPN" in the following analyses) and those with no detectable mutation in these two targets (referred to as "No mutation" for analytical purposes)

The box plots in Figure 5-7 show that there is overlap in the range of values seen between the MPN and No mutation groups. Significance testing by Wilcoxon rank sum test demonstrated a statistically significant difference between the two groups for all but one parameter. In our data set haematocrit was not shown to be significantly different between MPN and No mutation (Table 5-5).

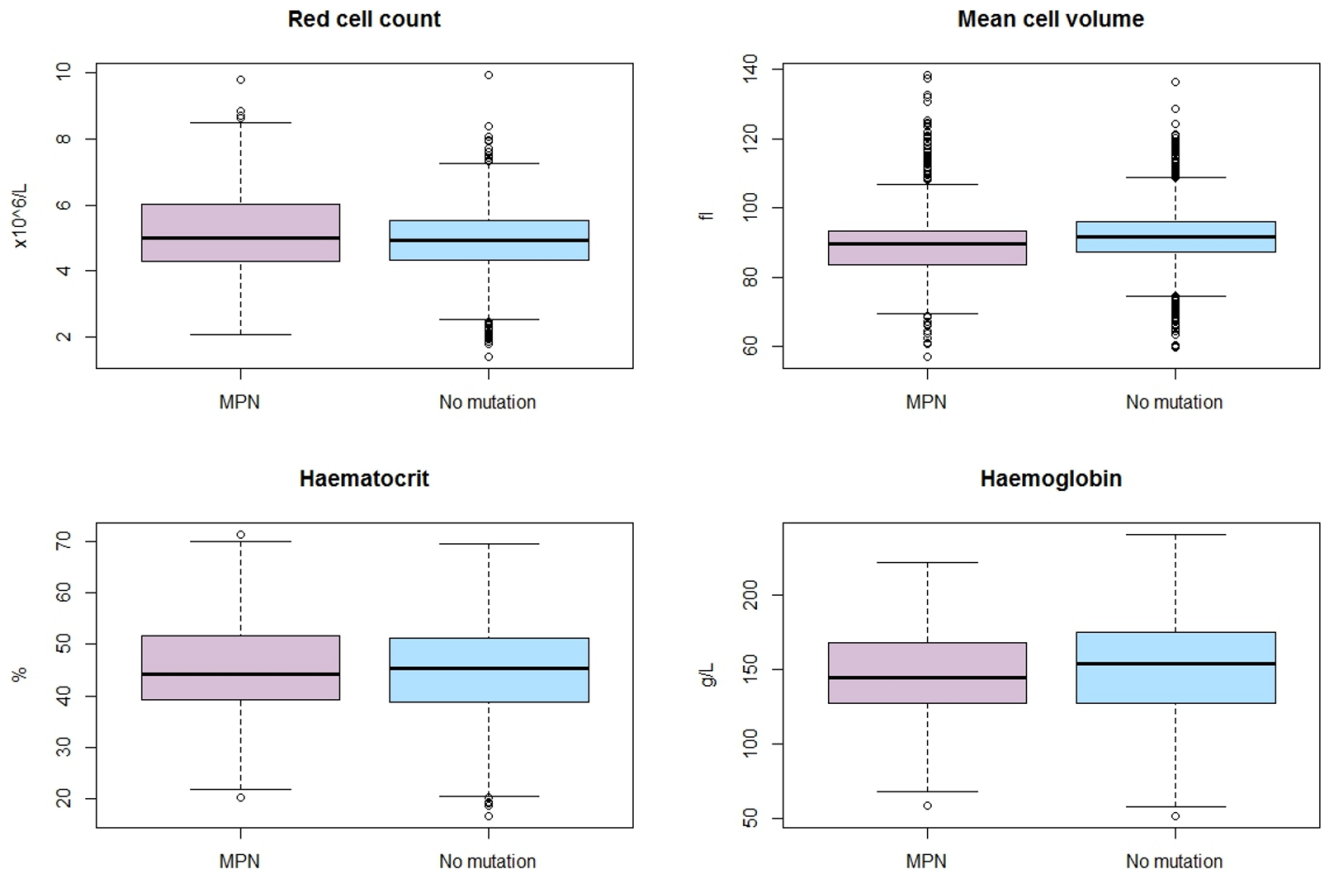

Figure 5-7. Comparison of red cell indices between referrals with a mutation of *JAK2* V617F or *CALR* (MPN) and those without (No mutation).

5.7.1.3 Gender of referred patients

A total of 6468 females and 5719 males were investigated for suspected classical MPN. The gender distribution of cases is shown in Table 5-4.

Diagnostic group	Males (n)	Females (n)	Ratio
MPN	1423	1583	0.90
No mutation	4296	4885	0.88
Combined	5719	6468	0.89

Table 5-4. Gender distribution of referred cases according to diagnostic outcome.

The gender distribution both overall and for referrals resulting in a diagnosis of a classical MPN are comparable and in keeping with published values (Roman et al., 2016).

5.7.2 Other blood count features

Data analysed in section 5.4 shows that the majority of referrals meet the minimum blood count criteria set out by either the BCSH and WHO (Barbui et al., 2015; Harrison et al., 2014; McMullin et al., 2007; Reilly et al., 2012b). However, only three laboratory parameters are included in the criteria – Platelet count, Haematocrit and Haemoglobin. Even the most basic full blood analysers produce a much wider range of indices, which may conceivably be of value in discriminating between specimens from patients with classical MPNs and those without.

5.7.2.1 Leucocyte parameters

White blood cell features are not used as a primary criterion in the diagnosis of classical MPN according to either BCSH or WHO publications. The full blood count analyser used in HMDS measures the total white cell count (WBC), lymphocyte count (Lymph) and neutrophil count (Neutr) and monocyte count (Mono). In this dataset (sample group 2), all leucocyte parameters were shown to have a statistically significant difference between the two groups, with white cell count, neutrophil and monocyte counts having a higher mean value in patients with a classical MPN compared with those who do not. For lymphocyte counts, the inverse pattern was seen. A comparison can be seen in Figure 5-9.

5.7.2.2 Platelet count

A comparison of platelet count between groups showed a significant increase in the mean platelet count associated with classical MPNs versus those with no mutation detected (Figure 5-8).

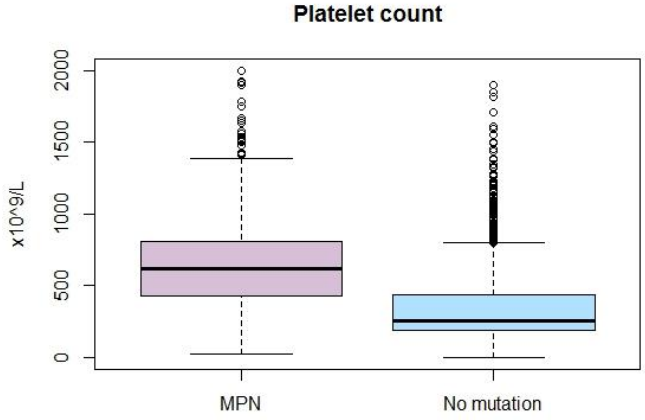

Figure 5-8. Comparison of platelet count between referrals with a mutation detected (MPN) and those without (No mutation).

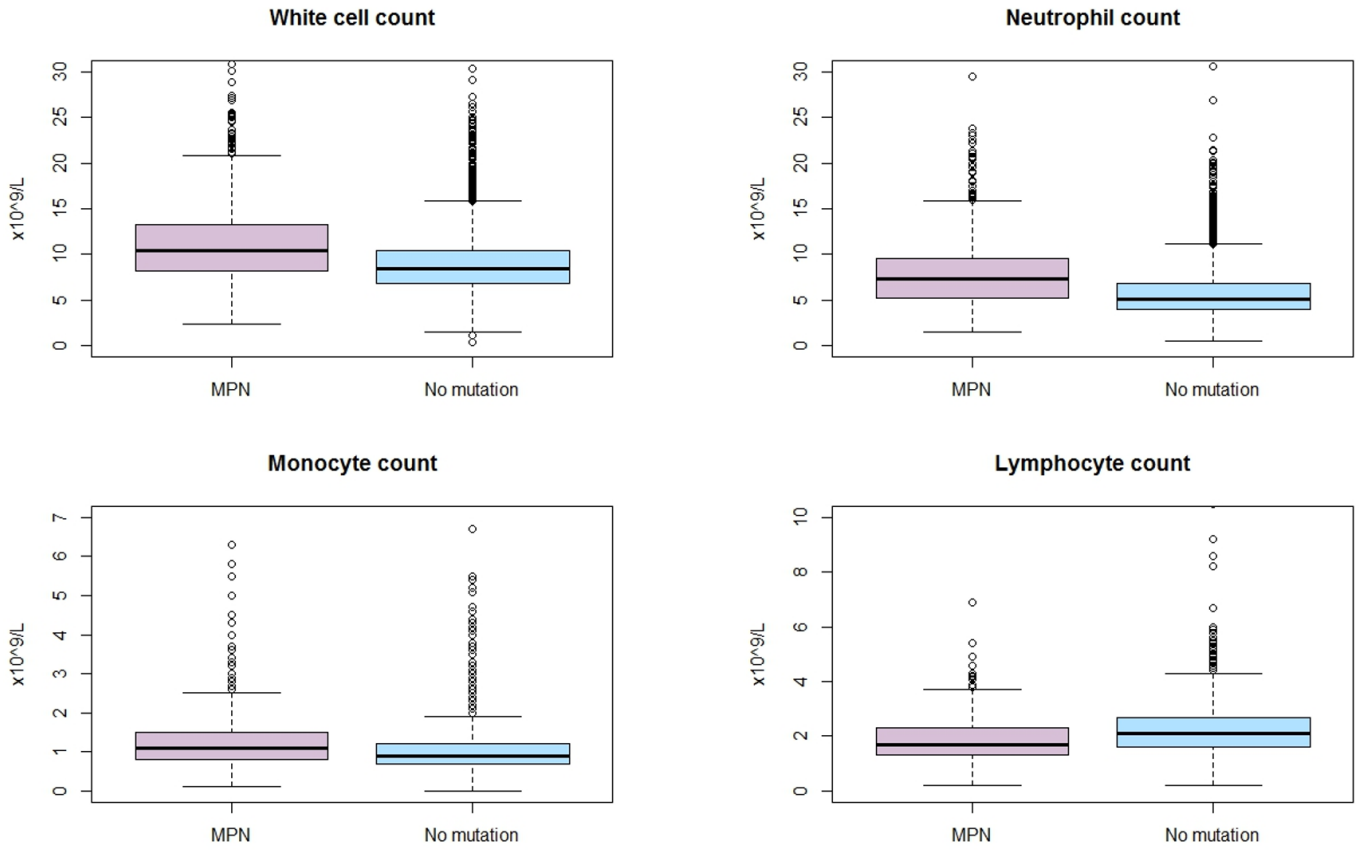

Figure 5-9. Comparison of leucocyte indices between referrals with a detectable mutation (MPN) and those without (No mutation).

Parameter	MPN	No mutation	Significance (*p*)
Hb (g/l)	**147.71**	**151.10**	**<0.001**
HCT (%)	**45.36**	**44.80**	**0.7267**
RBC ($X10^{12}$/l)	**5.16**	**4.88**	**<0.001**
MCV (fl)	**89.45**	**92.02**	**<0.001**
WBC ($x10^9$/l)	**11.49**	**9.04**	**<0.001**
Lymph ($x10^9$/l)	**1.85**	**2.21**	**<0.001**
Neutr ($x10^9$/l)	**8.27**	**5.79**	**<0.001**
Mono ($x10^9$/l)	**1.29**	**1.01**	**<0.001**
Plt ($x10^9$/l)	**649.62**	**338.50**	**<0.001**

Table 5-5. Comparison of mean red cell indices values between MPN and No mutation groups with significance level.

5.8 Discussion

The data presented in this chapter highlights the significant burden that molecular screening for suspected MPNs places upon diagnostic laboratory resources. The disproportionate increase in requests over the past decade shows no indication of slowing which, when paired with the modest increase in the number of diagnoses of MPN, suggests that referrals are not being made in a systematic or efficient way. The small proportion of patients who undergo bone marrow examination following a non-diagnostic result from peripheral blood assessment would also indicate over-requesting, which may suggest that there is either a lack of understanding of the limitations of *JAK2* and *CALR* screening in the diagnosis of MPNs or that the test is being requested in patients with low clinical suspicion of disease.

Between 30-40% of referrals received for investigation of suspected MPN do not meet the diagnostic criteria outlined by the either WHO or BCSH recommendations. In both sets of guidelines, the presence of abnormally high blood parameters is a central feature of the diagnostic criteria. Therefore, screening patients who do not meet these criteria could not result in the diagnosis of an MPN without the assessment of bone marrow morphology.

The exclusion of reactive/secondary causes for symptoms and laboratory features is also a key feature of the diagnostic guidelines. The audit of local referrals undertaken in section 5.5 demonstrates that alternative reasons for the abnormalities seen were identified by the clinician in 70% (35/50) referrals. However, molecular investigations were requested regardless of patient history. In those where a reactive or secondary cause was not identified from patient history, and inflammatory markers had not been assessed prior to clinical assessment, molecular screening was not delayed until a later time point. Molecular screening was deferred in only one of the audited cases and was requested at the patient's second clinic appointment when the blood counts had normalised. The patient notes indicated that it was been performed for completeness not because a clinical suspicion remained.

Part of this can be accounted for by the clinical pathway process described in Figure 3-4. During the first clinic visit, patients are seen by a haematologist who takes a detailed medical history and performs a physical examination. The clinician also reviews the most recent peripheral blood counts available on the hospital result service. Following this consultation, the clinician requests the appropriate investigations to be undertaken and the patient is sent to the phlebotomist for venepuncture. In all the cases audited, it is at this point that a *JAK2* mutation screen is requested, along with a full blood count and other basic blood tests for chemical pathology.

The main problem with this approach is that the full blood count results, on which the clinician bases their further investigations, can be several weeks old. Platelet counts, for example, can be increased during acute phase responses (during infection,

following surgery or blood-loss) and it is not uncommon to see an anomalous raised result. A sustained increase in platelet count would be indicative of an underlying pathology, however, there was often no history/evidence of this.

Ideally, a repeat full blood count would be performed prior to consultation to assess whether the results continue to be abnormal. Performing venepuncture in clinic prior to seeing the consultant as well as afterwards, would be impractical. It would cause additional time delays and discomfort to the patient. Therefore, requesting the molecular screening assays upfront is the most practical approach for the patient/clinician.

In addition to establishing current referral practice, this chapter also aimed to assess how useful the current clinical guidelines were in positively identifying patients with MPNs. Despite evidence that many referrals for *JAK2*/*CALR* screening are not clinically indicated, there is still a large proportion of referrals which do meet the laboratory criteria set out by the WHO and BCSH. However, the low number of diagnoses would suggest that the criteria are not specific in identifying potential MPNs.

Comparisons of diagnostic outcomes in patients according to whether they met one of the two guidelines showed that whilst the sensitivity of these criteria is high (86.96% for BCSH and 88.07% for WHO guidelines), the specificity is much lower (49.55% for the BCSH guidelines and 33.65% for the WHO). The data demonstrates that the guidelines have a very high negative predictive value. This clearly shows that there is minimal benefit to investigating patients who do meet the guidelines in terms of laboratory features. However, the low specificity of the guidelines does mean that a significant proportion of patients in whom testing would be indicated (according to blood parameters) would not be diagnosed with an MPN.

Both sets of guidelines use a minimal number of parameters in their diagnostic criteria. This does have the benefit of being easy to interpret by clinicians however;

the guidelines use arbitrary thresholds for these variables. Research in other haematological malignancies has shown that risk of disease increases on a linear scale and the use of such thresholds is of little significance in patients around the cut-off value ((Lichtman, 2013)). For example, a patient with a platelet count of $449x10^9$/L would not be investigated whereas a patient with a platelet count of $451x10^9$/L would be although the actual difference in risk of disease would be negligible.

There are also many other demographic and laboratory variables which are readily available to clinicians at the point of consultation. Statistical analyses can be performed to assess whether other information could be of value in the positive identification of patients with MPNs. Patient age and gender are well reported in MPNs and it is known that the risk of MPN increases with age. There was a statistically significant difference in the age of patients diagnosed with an MPN versus those with non-diagnostic results in our referral population. It would also be possible to attach a higher risk to male referrals as they too are known to have an increased incidence of MPN. Additional laboratory parameters including additional red blood cell indices such as MCV were found to show significant differences between MPN and non-diagnostic referrals, as were white cell parameters.

The wide range of variables that could potentially be used to guide suspected MPN referrals may be too complex to include in a simple algorithm, especially when there may be interaction between variables that can affect the overall risk. This suggests a potential role for predictive statistical model which can calculate an individual's risk of disease on the basis of a wide range of variables. It is anticipated that such a model would be capable of identifying MPN patients with at least the same level of sensitivity as the current guidelines but with an improved level of specificity. Implementation of such a model could reduce the total number of referrals for investigation, without impacting on patient safety.

6. Features of *CALR* mutations in prospective cohort

6.1 Rationale and context

Clinical data suggests that *CALR* mutated ET has distinct clinical and laboratory features compared to their *JAK2* V617F mutated counterparts, with *JAK2* V617F and *CALR* mutated ET presenting at a younger median age, and having a higher platelet count and lower haemoglobin than *JAK2* V617F ET (Chen et al., 2014; Rotunno et al., 2013b; Rumi et al., 2013). However, these studies have focussed on cases which had an existing diagnosis of ET according to the WHO criteria. In the absence of a *JAK2* V617F or *MPL* mutation, this would have required a bone marrow to be assessed. Data shown in 5.3 shows that in our setting BM sampling is performed infrequently, compared with the number of blood samples received for *JAK2* V617F and *CALR* mutational screening. Furthermore, bone marrow sampling was found to be biased towards younger patients with higher platelet counts and lower haemoglobin levels. If a similar bias was present in the cohort of patients examined in the published studies, this could contribute towards the documented differences between the *JAK2* V617F and *CALR* mutated subgroups.

At the time of analysis, no published data had assessed the features of *CALR* mutated ET in the context of a diagnostic setting, nor in prospectively tested patients. The aim of this chapter was to determine if the features of *CALR* mutated ET diagnosed prospectively in the population referred to a clinical laboratory setting are comparable to those identified through the retrospective testing performed in the literature. To achieve this, the clinical and laboratory features associated with *CALR* mutated ET have been compared to their *JAK2* V617F mutated counterparts which have been identified in HMDS through routine investigations.

6.2 Data overview and comparability of selected cohorts

The data set used in this analysis is taken from sample group 3 (shown in Figure 3-7). Briefly, this comprised all the cases of ET with a demonstrable *CALR* mutation identified prospectively in 2014 following assay introduction, alongside a group of *JAK2* V617F mutated ET's identified during the same time period. Briefly, this group consists of a total of 88 patients who were found to have a *CALR* mutation during this period, along with 311 patients with the *JAK2* V617F mutation.

6.3 Do the laboratory or demographic features of *CALR* and *JAK2* V617F mutated MPNs differ?

6.3.1 Laboratory features

The mean values of the full blood count indices were compared between the groups and tested for significance (Table 6-1). Statistically significant differences were seen in nine of the 11 parameters assessed. As reported in the literature (Rumi et al., 2013) *CALR* mutated ET presented with a significantly lower mean hb and wbc compared with *JAK2* V617F mutated cases. The observed mean plt count was higher in the *CALR* mutated cohort but this was not statistically significant as was indicated in the published data.

6.3.2 Demographic features of patients with *CALR* mutated ET

Age at presentation was significantly lower in the *CALR* mutated group with a mean of 63.61 years group compared to 71.66 years in the *JAK2* V617F mutated patients ($p=<0.01$). This is in keeping with the findings of Rumi et al (Rumi et al., 2014).

Parameter	*CALR* mutated	*JAK2* V617F mutated	P
Plt ($x10^9$/l)	716	655	0.38
Hb (g/l)	127.3	152.9	<0.01
RBC ($x10^6$/µl)	3.93	5.43	<0.01
HCT (%)	38.00	46.96	<0.01
MCV (fl)	98.25	88.11	<0.01
WBC ($x10^9$/l)	7.90	11.48	<0.01
Neutr ($x10^9$/l)	5.74	8.67	<0.01
Lymph ($x10^9$/l)	1.81	1.79	0.68
Mono ($x10^9$/l)	1.07	1.32	0.01
MCH (pg)	33.01	28.85	<0.01
MCHC (g/dl)	33.52	32.64	<0.01

Table 6-1. Comparison of mean full blood count indices between *CALR* and *JAK2* mutated referrals.

6.4 Discussion

The identification of *CALR* mutations in patients with *JAK2* wild-type ET and PMF had significant implications from a diagnostic standpoint. However, understanding the clinical relevancy of these mutations is a key consideration for clinicians. *CALR* mutations were first identified in patients with a previously confirmed diagnosis of *JAK2* /*MPL* wild-type ET which had been made according to the criteria set out by the WHO. As discussed in section 1.5.2.1, in the absence of a recognised genetic lesion, a diagnosis of ET cannot be made without the examination of bone marrow morphology.

In our setting, bone marrow assessment is infrequently performed on patients following inconclusive peripheral blood investigations for suspected classical MPNs. Data presented in 5.3 shows that in our experience, bone marrow testing is performed on patients with more abnormal blood features than seen in those who have peripheral blood screening alone (lower haemoglobin and higher platelet

counts). In routine clinical practice, it is not always feasible to perform a bone marrow assessment. Patients may decline invasive tests or may have existing comorbidities which would outweigh the benefits of confirming a diagnosis of a classical MPN. Taking these factors into account it is possible that the group of patients in whom *CALR* mutation were originally identified may not be representative of those who would be identified through prospective peripheral blood screening. At the outset of this work, there were no publications which assessed the impact of *CALR* mutations in a prospective cohort nor in the population of patients referred in routine clinical practice.

An assessment of the laboratory and clinical features associated with *CALR* mutated ET diagnosed in routine clinical practice were found to be in keeping with those described in the literature. In a routine diagnostic setting *CALR* mutations were associated with lower haemoglobin and white cell counts and higher platelet counts than *JAK2* V617F mutated counterparts. Age at presentation was significantly lower in the *CALR* mutated ET cohort compared to the *JAK2* V617F mutated counterparts in keeping with the published data.

Overall, this data would suggest that the data presented in this work should be applicable to diagnostic practise on a wider scale. Routine diagnostic laboratories see a wide range of patients at various stages of disease development and the ability to generalise the findings of this work is encouraging. In addition, this data indicates that prognostic findings from large scale retrospective studies can be applied to patients in routine clinical practise which is of immense value to both patients and clinicians.

7. Selection and design of a predictive statistical model

7.1 Rationale and overview

The limited specificity of diagnostic guidelines in the positive identification of patients with classical MPNs combined with the potential value of additional biological and demographic variables discussed in 5.7 would suggest that classical MPN diagnosis would be an ideal candidate for the development of a predictive statistical model. The aim of this chapter is to identify suitable statistical modelling approaches for the data available and the intended research outcome.

Selecting the most appropriate methodology to apply to the data is of key importance in the development of a statistically sound model. There is a vast range of statistical approaches available for the interpretation of complex data. The research question provides the best indication as to which methodological approaches are suitable. The aim of this thesis is not to further the understanding of the biological basis of classical MPNs, nor to disclose previous unidentified relationships between biological variables. The primary purpose of the statistical model is in its ability to predict an outcome on the basis of diagnostic measurements. For this reason, predictive statistical models are the most appropriate choice.

Predictive modelling approaches can be broadly divided into supervised and unsupervised techniques depending upon whether the class of data is known or unknown. As the statistical model will be developed using a case control data set, with the class (MPN or no Mutation) known from the outset, this work focusses on supervised, predictive modelling techniques.

The aim of the statistical model will be to predict whether an individual referral with suspected classical MPN is more likely to have a *JAK2* V617F or *CALR* mutation (MPN) or not (No mutation) based on the presentation full blood count data in combination with demographical information, such as age and gender. The classification produced by the statistical model is used to inform downstream investigations.

7.2 Selection of supervised modelling approach

A large scale comparison of supervised, predictive modelling approaches was performed using Waikato Environment for Knowledge Analysis (WEKA) version 3.8, an open source suite of software developed specifically for the purpose of data mining (Frank et al., 2016). The analyses were performed on sample group 4, which was comprised of 2970 referrals which underwent both *JAK2* V617F and *CALR* mutational screening in HMDS with full blood count data available at the time of referral (see 3.5.2.3 and Figure 7-2).

7.2.1 Classifier comparison

A set of eight different supervised classifier approaches were assessed during this initial analysis.

- Naïve Bayes
- J48 Decision Trees
- Random Forest
- Logitboost
- AdaboostM1
- SMO (sequential minimal optimisation/support vector classifier)
- Multilayer Perceptron (neural network)
- Logistic Regression

The classifiers were all run using the default settings in WEKA. The following performance characteristics were assessed to compare and evaluate classifier performance: Accuracy, sensitivity, specificity, precision and F-measure. Sensitivity and specificity measure the proportion of true positive classical MPNs identified correctly and the proportion of true negative referrals which were correctly identified, respectively. Precision (also referred to as positive predictive value) assesses the likelihood of a false positive, whilst the F-measure is the mean of both precision and sensitivity (see Table 3-4 for details of their calculation).

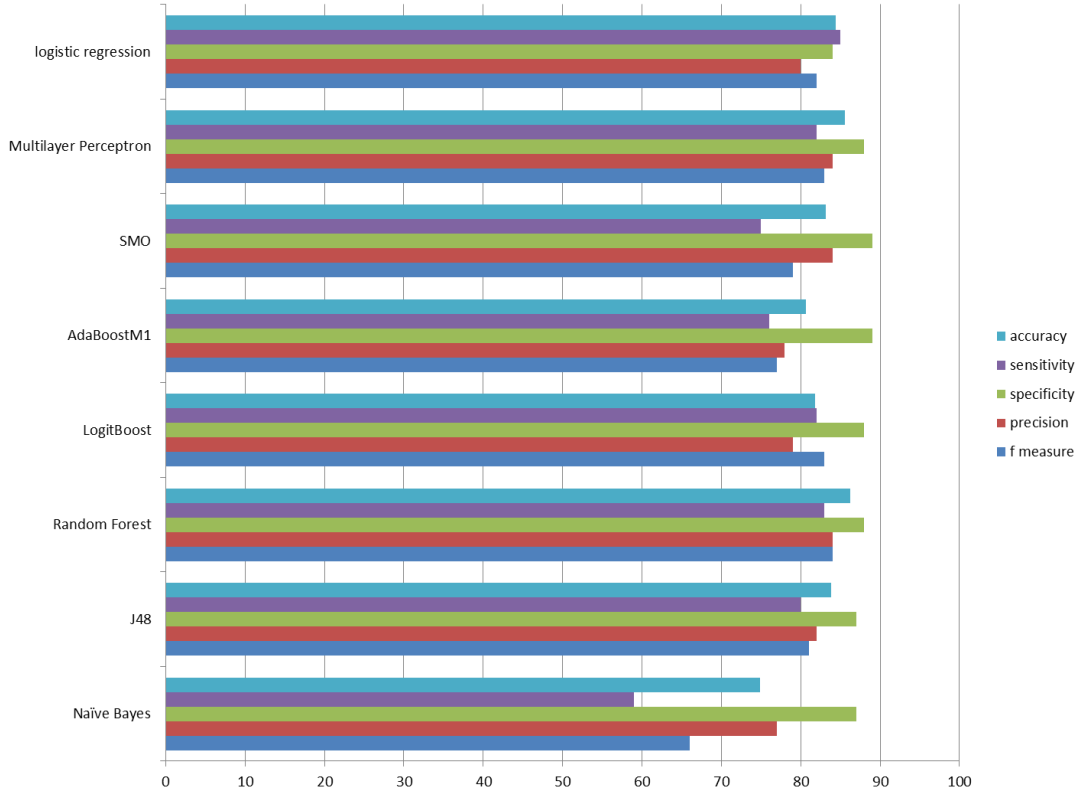

Figure 7-1. Comparison of supervised classifier approaches using WEKA.

7.2.2 Classifier selection

The primary aim of the predictive model was to improve on the specificity achieved using guidelines alone (as described in section 5.6), without reducing the sensitivity, therefore modelling approaches which showed high levels of both characteristics were determined to be most desirable. In this general comparison Logistic

Regression, Random Forest and Boosted analysis showed the best balance between sensitivity and specificity and were selected as the modelling approaches to develop further in this work. Details of the principles of each approach are described in section 3.7.3.

7.3 Development of predictive statistical model in R software

Standard methods from predictive modelling were used to design, optimise and validate the predictive models as outlined in section 3.7.3. This was performed in RStudio using R statistical software version 3.3.2 and the packages listed in Table 3-2. The data scripts used to perform the analyses described within this chapter are available in the appendix (section 10.2).

7.3.1 Data preparation

A data set comprising all referrals of suspected classical MPN received by HMDS was extracted from HILIS as described in section 3.5.2.3. This data set was initially comprised of 12499 individual referrals made between the beginning of 2005 and the end of 2014. Prior to further analysis the data was subjected to a series of 'clean-up' procedures as shown in Figure 7-2.

Missing or empty values were identified and replaced with a missing data indicator ('NA') to prevent interference with statistical analysis. The data was then read into the R software and summary statistics were displayed to identify potentially erroneous data such as typographical or transcription errors or changes in units of measurements. Data points which appeared to be outliers were identified, reviewed and corrected if appropriate.

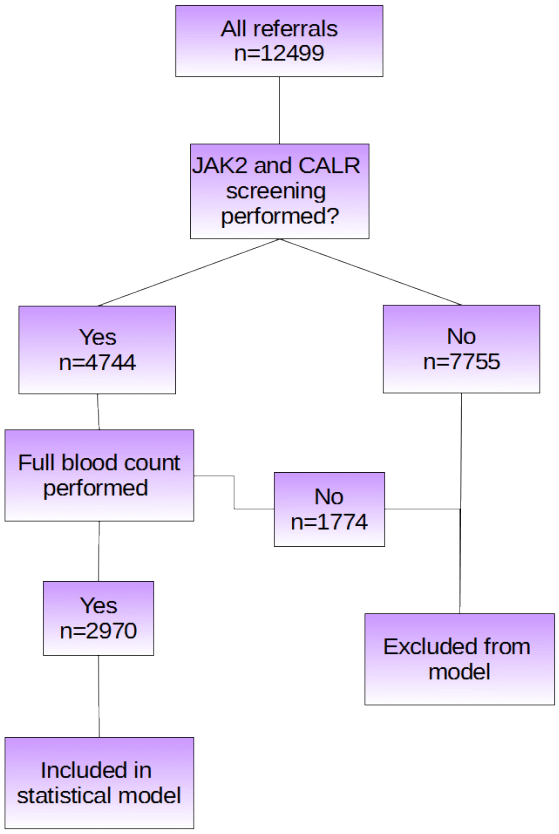

Figure 7-2. Data 'clean up' process undertaken prior to statistical modelling.

Following data clean-up, a data set containing only referrals received during the year 2014 was selected and these data was used for the development of the statistical models. The dataset (sample group 4) contained 2970 referrals, of which 42.63% (n=1266) were confirmed classical MPN (*JAK2* V617F mutated n=1122, *CALR* mutated n=131 and concurrent *JAK2* V617F/*CALR* n=2) and the remaining 57.37% (n=1704) did not have a detectable mutation for comparison.

7.3.2 Data comparison group

The laboratory process for investigating suspected classical MPNs in HMDS has been described in 3.4.1, briefly, following receipt of a peripheral blood sample the laboratory performs a full blood count and morphological assessment followed by screening the specimen for *JAK2* V617F and *CALR* mutations. If no mutation is

detected, the report suggests that a repeat specimen is sent (preferably a bone marrow) to exclude the presence of disease. If received, the bone marrow sample is tested for the presence of *MPL* and/or *JAK2* Exon12 mutations. The proportion of patients with no *JAK2* V617F or *CALR* mutation who are referred with a subsequent sample is low, at approximately 10%, as shown in 5.3. Consequently, the number of specimens in which an MPN is conclusively excluded is insufficient for use in statistical modelling.

Primarily, this issue relates to ET/PMF patients where approximately 25% of cases will not have a mutation in *JAK2* V617F or *CALR*. The number of potentially undiagnosed MPNs (ET/PMF) has been calculated using the published frequencies of gene mutations to give an approximate *CALR*:*MPL* ratio of 6:1 and *CALR*:Triple Negative ratio of 3:1. The data set used for statistical modelling contained 133 *CALR* mutated specimens, therefore the expected number of *MPL* mutated specimens is calculated to be 22 and triple negative ET/PMF to be approximately 44. The comparison group contains 1704 specimens, therefore undiagnosed MPNs may represent 3.9% of this group.

7.3.3 Data normalisation

Prior to logistic regression analysis each variable was normalised. This was performed by visualisation of a quantile-quantile (QQ) plots which compare the data distribution seen in each variable with theoretical values were the data distributed normally. Where significant skewing occurred transformations (logarithmic or square root) were applied and QQ plots redrawn as shown in Figure 7-3 . Red cell parameters (Hb, RBC, HCT, MCV and MCH) were unchanged following QQ plot visualisation. The platelet count required a square-root transformation, whilst the white cell parameters (WBC, Lymph, Neutr and Mono) required a logarithmic transformation.

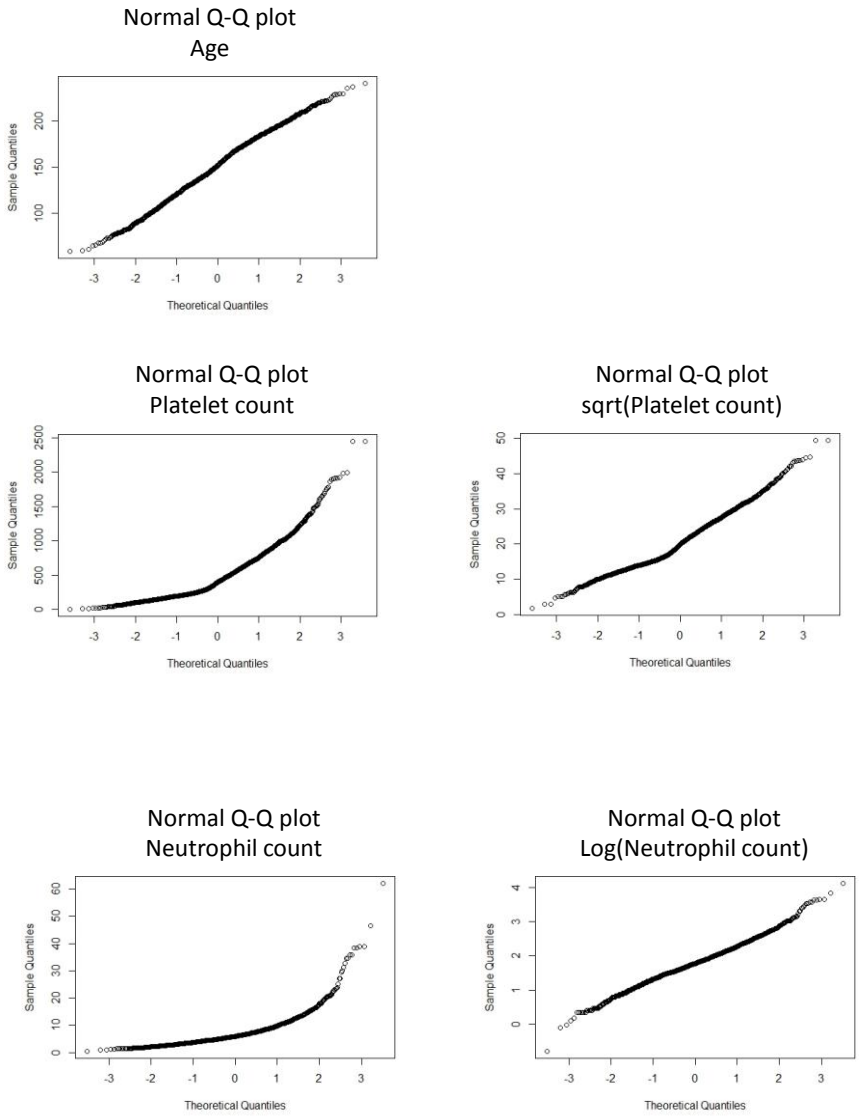

Figure 7-3. Examples of normal Q-Q plots and data transformations required for normalisation and further statistical analysis.

7.3.4 General analysis of data relationships

Using the pairs function, a series of scatter plots was displayed which plotted each of the variables against one another to visualise trends between attributes which may impact on later analysis. Linear relationships were visible between Hb and HCT, WBC and Neutr, MCH and MCHC. A linear relationship is indicated by the data points forming a diagonal line rather than a cloud; an example of each is shown in Figure 7-4. These relationships are easily explained as one of the linear parameters

in each pair is a derivative of the other. For example, neutrophils are the largest subpopulation of white blood cells, and it was expected that in the majority of cases as the white cell count increases, the neutrophil count would follow suit. It was anticipated that it would be necessary to correct for multicollinearity in the models.

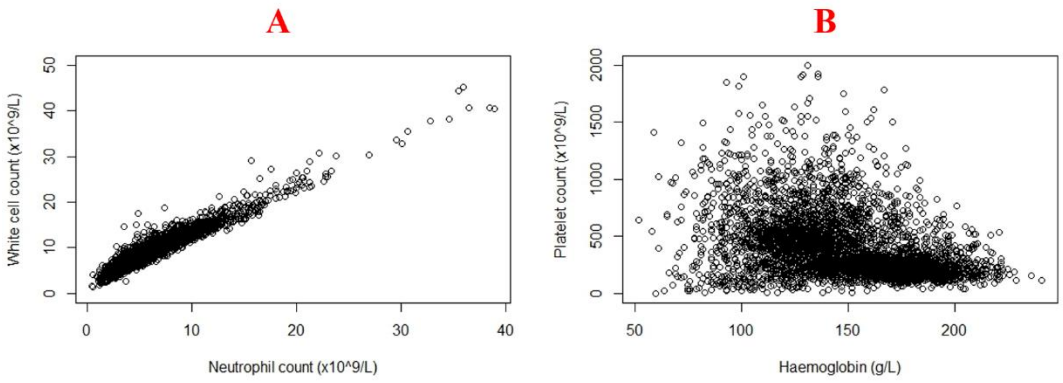

Figure 7-4. Example of full blood count parameters which show collinear (A) and non-linear (B) relationships.

7.3.5 Setting a base performance against which to evaluate model performance

The performance of the current clinical guidelines was calculated in section 5.6 using the data obtained from all suspected MPN referrals received over a 3-year period (2012-2014). This sets the base performance level against which the statistical models developed in this chapter will be assessed. For the model to be deemed successful, it should outperform the guidelines in terms of these metrics as shown in Table 7-1.

	Sensitivity (%)	**Specificity (%)**
BCSH	86.96	49.55
WHO	88.07	33.65

Table 7-1. Performance metrics associated with current clinical guidelines.

7.4 Logistic regression model development

A basic overview of the premise of logistic regression is described in 3.7.3.1. One possible advantage of logistic regression, from the point of the end user, is that the weighting of each variable is quantified, and it is possible to present the user with a formula that can be interpreted. The ability to understand and compare how different biological and demographic variables contribute to the probability of disease may make this approach more readily acceptable in clinical practice.

7.4.1 The logistic regression equation

Logistic regression gives each explanatory variable a coefficient that measures the weight of its contribution to the outcome (dependent variable). Overall the output of the logistic regression equation is the odds of belonging to one of the categories of dependent variable – in this case the probability of having an MPN.

The equation takes the following form:

$$logit(p(x)) = \log\left(\frac{p}{1-p}\right) = a + b_1x_1 + b_2x_2 + \ldots$$

Where $logit(p(x))$ is the log-odds of patient x having an MPN and is equal to the sum of the intercept 'a' plus the value of each explanatory variable 'x_n' multiplied by its regression coefficient 'b_n'. A larger regression coefficient indicates a greater the contribution of the associated explanatory variable to the overall probability of disease. A positive value indicates that the variable increases the probability, whereas a negative coefficient subtracts from it.

7.4.2 Initial variable selection and model design

The first step in the development of the logistic regression model was the selection of the explanatory variables that would be used in the analysis. Initially the model included all variables collected – Age, Gender, Hb, RBC, HCT, MCV, MCH, MCHC, WBC, Neutr, Lymph, Mono and Plt.

Analysis was performed using the core R function 'glm' with data family set to "binomial" in order to link to logistic regression analysis. The output of the initial regression analysis (shown in Table 7-2) identifies highly significant weightings for Age, Hb, RBC, Lymph and Plt. With statistically significant weightings also seen in MCV, Gender, MCHC and HCT.

	Estimate	**PR(>|Z|)**
(Intercept)	**112.86**	**<0.01**
Age	**-0.03**	**<0.01**
Gender (male)	**-0.31**	**0.02**
Hb	**0.18**	**<0.01**
RBC	**-4.90**	**<0.01**
MCV	**-0.27**	**0.01**
MCH	**2.33**	**0.82**
MCHC	**-19.93**	**0.05**
WBC	**-0.23**	**0.84**
Lymph	**1.12**	**<0.01**
Neutr	**-0.62**	**0.43**
Mono	**-0.02**	**0.93**
Plt	**-0.23**	**<0.01**
HCT	**-5.34**	**0.02**

Table 7-2. Results of logistic regression analysis showing weighting and significance of each explanatory variable.

7.4.3 Assessment of multicollinearity within model

Paired plots indicated linear relationships between some of the explanatory variables. Collinearity can occur when explanatory variables are not independent, for example where two variables are closely related and are effectively measuring the same effect. This can result in a statistical model which is unstable and can reduce the statistical power of the model and make it difficult to interpret. To assess the level of collinearity within this logistic regression model, the variance inflation factor (VIF) for each explanatory variable was calculated.

Variance inflation factors are computationally defined as the reciprocal of tolerance (Marquaridt, 1970) and can be denoted as:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Where 'i' is the explanatory variable and 'R_i^2' is the coefficient of determination (the square of the correlation between predicted and actual scores).

	VIF
Age	1.07
Gender	1.21
Hb	121.51
HCT	56.04
MCV	168.36
MCH	358.83
MCHC	62.43
WBC	50.28
Neutr	38.12
Lymph	4.06
Mono	2.56
Plt	1.53

Table 7-3. Size of VIF associated with each explanatory variable from logistic regression analysis.

The size of VIF is an estimate of the magnitude by which standard errors are inflated than would be case if the variables were unrelated. Acceptable levels of VIF have been recommended in the literature, with a maximum of 10, which corresponds to a tolerance of 0.1, commonly indicated (Mansfield and Helms, 1982; Marquaridt, 1970). The size of the VIF associated with each of the explanatory variables is shown in Table 7-3.

It was evident that a high level of multicollinearity exists within the model in its current form. This level of multicollinearity was likely to affect the validity of our statistical model and needed to be addressed.

7.4.4 Rationale behind removal of attributes causing collinearity

Prior to further analysis it was necessary to reduce the level of multicollinearity within the model. To achieve this, a number of explanatory variables needed to be removed from the model. Linear relationships were present between WBC, Neutr and Mono variables. As neutrophils and monocytes are both subpopulations of leucocytes (WBC) this was expected. The decision was made to remove WBC from the analysis and it did not represent a specific cellular population. Examination of the data showed that mono counts were not always reported by the full blood count analyser and for this reason, it was also removed from the analysis.

Linear relationships were also identified between MCV and MCH as well as Hb, RBC and HCT. The parameters are all measures relating to erythrocytes. Of these parameters, HCT and MCH were not directly measured by our instrument and were instead derived from other measured parameters. The decision was made to remove these from the analysis as they represented duplicated information. The logistic regression model was rerun with the more restricted range of explanatory variables. All variables now show statistically significant weightings (Table 7-4).

	Estimate	PR(>|Z|)	VIF
(Intercept)	81.14	<0.01	-
Age	-0.03	<0.01	1.06
Gender(male)	-0.30	0.02	1.14
Hb	0.15	<0.01	66.58
RBC	-5.12	<0.01	68.46
MCV	-0.26	<0.01	11.21
MCHC	-13.28	<0.01	4.62
Neutr	-0.74	<0.01	1.14
Lymph	1.01	<0.01	1.15
Plt	-0.23	<0.01	1.48

Table 7-4. Results of revised logistic regression analysis showing weighting and significance of refined set of explanatory variables and corresponding VIF values.

Re-evaluation of the VIF shows that there was still significant collinearity. The highest VIF values were associated with Hb, RBC and MCV. These three variables are all measurements associated with the red blood cells and therefore biologically related.

Using a series of regression analyses, excluding each of these variables in turn the VIF and √VIF were calculated. The lowest VIF values were obtained when RBC was removed from the analysis (Table 7-5). These variables were selected for inclusion in the final logistic regression analysis.

	VIF
Age	1.04
Gender	1.15
Hb	1.62
MCV	1.23
MCHC	1.39
Neutr	1.13
Lymph	1.15
Plt	1.51

Table 7-5. Size of VIF associated with final selection of explanatory variables chosen for logistic regression model.

7.4.5 Attribute selection

Following the removal of variables contributing to the multicollinearity in the model, attribute selection was performed using the least absolute shrinkage and selection operator (LASSO) method. LASSO regression performs attribute selection to enhance the predictive accuracy of the model by shrinkage. This involves penalising the absolute size of the regression coefficients, and in doing so, some of the coefficients will be shrunk to zero, at which point they are removed.

When the variables selected in section 7.4.4 were analysed using LASSO regression, and the 'best' model selected, none of the variables were 'shrunk' away by the analysis, confirming that these variables share similar weightings within the model and should be included in the logistic regression analysis.

7.4.6 Cross validation of logistic regression model

The explanatory variables selected were incorporated into a final logistic regression model. For this analysis the "lrm" function from the package 'rms' was used as this provides a series of performance metrics that can be used to assess the model (Table 7-6).

		Discrimination indexes		Rank discrimination indexes	
Observations	2298	**R2**	0.55	**C**	0.89
Case	1033	**Brier**	0.13	**Dxy**	0.78
Control	1265				

	Coefficient	Pr(>|Z|)
Intercept	-29.99	<0.01
Age	-0.03	<0.01
Gender (male)	-0.29	0.02
Hb	-0.02	<0.01
MCV	-0.01	0.16
MCHC	11.89	<0.01
Neutr	-0.60	<0.01
Lymph	1.09	<0.01
Plt	-0.24	<0.01

Table 7-6. Performance metrics of final logistic regression model.

The highlighted 'C' denotes the 'c' index', which represent the area under the receiver operating characteristic (ROC) curve. 'Dxy' denotes the Somers' Dxy rank correlation between the predicted probabilities and the observed responses. There is a simple relationship between 'C' and 'Dxy' (Somers, 1962):

$$D_{xy} = 2(C - 0.5)$$

Again, a Dxy of 1 denotes perfect discrimination of observations by the model. In this case the 'C' index is 0.887, indicating a strong predictive model.

Following analysis, cross validation of the model was performed using the "validation" function from within the "rms" package selecting the cross-validation method.

	Original	Training	Test	Optimism	Corrected
DXY	0.7747	0.7748	0.7704	0.0044	0.7703
R2	0.5470	0.5471	0.5582	-0.111	0.5581
C index	0.88735	0.8874	0.8852	0.0022	0.8852

Table 7-7. Cross-validation performance metrics of logistic regression analysis.

The cross-validation process described in 3.7.3.1 is intended to detect overfitting and estimate the extent to which it is affecting the model. Whilst it is desirable to maximise the performance of the model on the training data set, the performance of the model on unseen data will determine its success. Cross validation of this logistic regression model results in slight reductions in the performance metrics (Table 7-7).

7.4.7 Model evaluation

To determine the performance of the model on the training set, the metrics described in Table 3-4 were assessed. As the model returns the likelihood of disease as a probability, it was necessary to choose a threshold level of probability at which the model would classify each sample as either MPN or No mutation.

7.4.7.1 Identifying the threshold giving best model performance

Optimisation of the threshold was performed in two ways; the first was to identify the threshold of probability which yielded the highest sensitivity plus specificity. This was achieved using features of the ROCR package, Figure 7-5 shows sensitivity vs specificity with increasing thresholds of probability in ROCR in the form of a graph. A threshold of 0.56 was calculated as returning the greatest sensitivity plus specificity, achieving a combined value of 1.65.

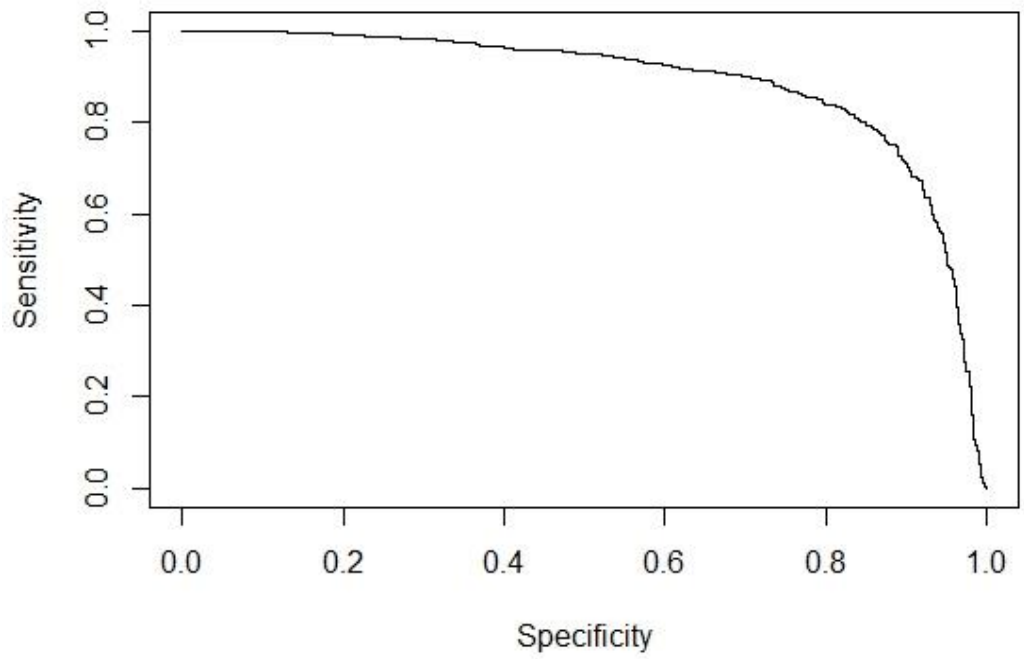

Figure 7-5. Graphical representation of yielded sensitivity vs specificity with increasing threshold of probability in ROCR.

This threshold was applied to the model and performance measured using the confusionMatrix function from the CARET statistical package.

	Actual MPN	Actual no mutation
Predicted MPN	848	215
Predicted no mutation	185	1050

Table 7-8. Performance of logistic regression model at optimal prediction threshold calculated by ROCR analysis.

Using a threshold of 0.56 to determine class returned the predictions shown in Table 7-8. This gave a sensitivity of 0.82 and specificity of 0.83. Whilst this represents a substantial increase in specificity over that achieved using the clinical guidelines (which has been determined to be 49.55% at best), the sensitivity of the model falls below the 86.96% minimum the guidelines achieved with 185/1033 actual MPNs predicted to have no mutation.

Maintaining the level of sensitivity achieved by the guidelines and minimising the number of false negatives produced by the model was more important that achieving maximum levels of specificity. Therefore, a second approach was taken to optimise the threshold of probability within the model.

7.4.7.2 Optimising the threshold to give the best model for purpose

To determine the threshold which would give maximum specificity whilst maintaining sensitivity above that achieved by the clinical guidelines was determined by assessing the sensitivity and specificity of the model across a range of thresholds (Table 7-9).

Using a threshold of 0.7 to determine class membership gave a sensitivity marginally higher that that achieved using the guidelines alone, alongside an increase in specificity (72.73% compared with 49.55% achieved by guidelines). When a threshold of 0.7 was applied to the dataset, the confusion matrix gave the results shown in Table 7-10. This threshold has been applied to all further analyses.

Threshold	Sensitivity	Specificity
0.05	0.17	1.00
0.10	0.29	0.98
0.15	0.38	0.97
0.20	0.47	0.96
0.25	0.54	0.94
0.30	0.61	0.92
0.35	0.66	0.91
0.40	0.72	0.90
0.45	0.74	0.88
0.50	0.77	0.86
0.55	0.82	0.84
0.60	0.84	0.81
0.65	0.87	0.77
0.70	0.89	0.73
0.75	0.92	0.66
0.80	0.94	0.58
0.85	0.94	0.48
0.90	0.97	0.33
0.95	0.98	0.12

Table 7-9. Logistic regression model performance over a range of increasing thresholds of probability.

	Actual MPN	Actual no mutation
Predicted MPN	920	345
Predicted no mutation	113	920

Table 7-10 Predictive performance of logistic regression model with a threshold of 0.7.

7.4.8 Validating the logistic regression model on a test dataset

The logistic regression model was built using data cohort containing a ratio of MPNs:No mutation of 0.74. Data presented in 5.2 shows that the actual frequency of MPNs is much lower than this (in the region of 0.2:1). In addition to this difference in MPN representation, there is the inherent issue of overfitting, whereby the model has been optimised for its performance on a specific data set which may not be applicable to 'real world' data. To accurately assess the suitability of the model in clinical practice it needs to be validated using a 'test' data set – this is a set of data that has not been 'seen' by the model.

7.4.8.1 The test data set

A data set of 515 referrals received for the investigation of a suspected MPN was identified through the SQL searches performed on the HILIS database. These referrals were received during the first calendar quarter of the year 2015 and comprised 112 *JAK2* V617F or *CALR* mutated classical MPNs and 403 samples in which no mutation in *JAK2* V617F or *CALR* was identified. This data set was completely independent of that used to build to model and therefore suitable for model validation.

7.4.8.2 Model performance on test data

The logistic regression model described in 7.4.6 was applied to the test data set and performance characteristics determined using the probability threshold of 0.7 as determined to be optimal in 7.4.7.2. The model performance is shown in Table 7-11.

	Actual MPN	**Actual no mutation**
Predicted MPN	70	82

Predicted no mutation	6	203

Table 7-11. Performance of logistic regression model on test dataset.

The model gave a sensitivity of 92% and specificity of 71%, compared with 89% sensitivity and 72% specificity achieved using the model-build data-set.

7.4.9 Analysis of misclassified data

7.4.9.1 Analysis of false negative classifications

When the logistic regression model was applied to the test data set, 6 referrals which were found to carry either a *JAK2* V617F or *CALR* mutation ('MPN') were incorrectly predicted to have 'No mutation' by the model, these can be referred to as false negatives. Each of these referrals was reviewed to establish whether they had additional clinical features that would have indicated the presence of disease. In two instances, the patient had previously been diagnosed with an MPN and was on treatment at the time of investigation and consequently had normal full blood count indices on the referred sample. One of the referrals was from a patient who had a previously diagnosed, co-existing plasma cell neoplasm, this may have resulted in impaired haematopoiesis which would account for the normal full blood count at the time of referral.

A further sample was from a patient who had been recently received chemotherapy for a non-haematological malignancy and one referral had normal full blood count indices. In these two specimens, the *JAK2* V617F mutations was reported as being present at a low level (estimated to be <10% allelic burden based on area under peak). The final misclassified 'MPN' did have an isolated increased platelet count

of $569x10^9$/L, at the time of writing the patient had not undergone any further investigations and no further clinical information was available.

Of the 6 false negatives, only 4 were determined to be genuine referral for the investigation of a suspected MPN. Three of the 4 may have been misclassified due to the lack of abnormal blood count parameters. In 2/3 of these, impaired haematopoiesis attributable to another malignancy may have masked the underlying MPN and this information was not recorded on the request form, highlighting the importance of providing relevant clinical details to enable appropriate processing of specimens. Only one sample that had a clearly abnormal full blood count would have been misclassified by this model.

7.4.9.2 Analysis of false positive classifications

Analysis of the test data set also identified 82 samples with 'No mutation' that were misclassified as being 'MPN' by the model, herein referred to as false positives. Due to limitations in access to clinical data, only those referrals from within the Leeds Teaching Hospitals could be reviewed further (n=12). Of the patients reviewed, the majority (n=10) were described as having an underlying reactive cause which accounted for deranged full blood count results. In 3/10 referrals, an acute episode of illness coincided with investigation and the full blood count parameters returned to normal following recovery. A chronic reactive process was identified as the probable cause in 7/10 referrals; with chronic iron deficiency stemming from chronic blood loss in 4 of these, active rheumatoid arthritis present in 2 and anaemia of chronic disease in one other.

Hypoxia secondary to smoking and pulmonary disease was determined to be the cause of polycythaemia in one of the 2 remaining specimens which did not have reactive conditions. The final 'false positive' referral was followed up by bone marrow investigations which confirmed a diagnosis of ET.

If this data is representative of the false positives classified, it would indicate that an alternative cause for the abnormal blood features can be identified in the majority of referrals with no mutation detected and that, in a large proportion, the features begin to normalise shortly after investigation. If reactive causes were excluded prior to referral for *JAK2* V617F and *CALR* mutational screening, these specimens would not have been received for investigations.

7.5 Random forest model development

The premise of Random Forest analysis is described in 3.7.3.1. This modelling approach is sometimes referred to as 'black box' method as it is not visible to the user how the outcome is determined and the weighting of the variables in relation to the outcome is difficult to visualise. Random Forest analysis, by its nature, applies arbitrary thresholds to the data in order to form the branches of decision trees. This approach is similar to the way in which clinical guidelines have been set out, albeit on a much larger and more complex scale.

7.5.1 Random forest analysis

Using the data set created in 7.3.1, a Random Forest model was built using the randomForest statistical package, selecting to perform 10-fold cross validation on a set of 500 distinct decision trees as part of the model training process. The initial step of the model training process calculated the optimal number of branches per tree, in this analysis the model accuracy was highest when decision trees of no more than 7 branches were used (Table 7-12).

	Actual MPN	**Actual no mutation**
Predicted MPN	873	158
Predicted no mutation	165	1100

Table 7-12. Performance of optimised Random Forest model on training data.

When this optimised model was applied to the data (Table 7-12), it produced a sensitivity of 84.10% and specificity of 87.44% on the training data set.

7.5.2 Testing the model on an unseen data set

Using the same test data set as described in 7.4.8.1, the Random Forest model was reapplied (Table 7-13) and performance characteristics calculated. The model sensitivity was 97.37% and specificity was 15.09%.

	Actual MPN	**Actual no mutation**
Predicted MPN	74	242
Predicted no mutation	2	43

Table 7-13. Performance of optimised Random Forest model on test dataset.

7.5.3 Analysis of misclassified data

7.5.3.1 False negatives

In this analysis only 2 MPNs were predicted to have no mutation when the random forest model was applied to the test data. One of these had also been misclassified by the Logistic Regression model and was found to be from a patient with an existing diagnosis of ET, who was on treatment at the time of referral. The other false negative produced by this model was a referral with an isolated thrombocytosis of $1087x10^9$/L; unfortunately, no additional clinical information or follow up was available.

7.5.3.2 False positives

The random forest analysis incorrectly classified 242 samples with no mutation as an MPNs. Thirty-three of these false positives had been referred from within Leeds Teaching Hospitals and clinical records were reviewed.

Sixteen of these 33 referrals were due to features of suspected polycythaemia. Ten were attributable to smoking and/or alcohol intake – with full blood count indices resolving to within reference ranges at following reduced intake. Therapy related polycythaemia was diagnosed in a further 2 - both of which were due to testosterone replacement. The remaining 4/16 were without obvious cause, but all resolved shortly after investigation. Of the remaining 17 referrals, 1 referral had pancytopenia and was found to have a pre-exiting diagnosis of Refractory Cytopenia with Multi-lineage Dysplasia and Fibrosis. Four referrals were from patients suffering from thrombotic events and 12 from patients with an isolated thrombocytosis. In the 12 patients with thrombocytosis, 4 were attributed to iron deficiency, a further 4 to chronic reactive conditions and one was following splenectomy, no underlying cause could be identified in the final 3.

Bone marrow investigations were performed in the 3 instances where an underlying cause was not documented, a diagnosis of ET was made in one specimen, systemic mastocytosis in the second and the final bone marrow showed a reactive picture.

7.6 Gradient-boosted modelling

The final statistical modelling approach applied to the data set was gradient boosting. The basic premise of this technique is described in section 3.7.3.3. Like Random Forest modelling, this is a 'black box' analysis and is also an ensemble technique combining multiple different decision tress classifiers. Gradient-boosting differs from Random Forest analysis, in that the vote produced by each classifier within the ensemble is given a weighting depending upon the overall performance of that classifier. The final classification is based upon 'proportional representation' rather than each classifier making an equal contribution to the overall decision.

7.6.1 Development of gradient boosted model

The gradient boosted model was built using features from the 'caret' and 'gbm' packages in RStudio and the data-set described in section 7.3.1. Repeated cross-validation was performed with 10 samples, repeating the process 10 times. The process was run in triplicate, with increasing interaction depths (set at 1, 3 and 5). Each iteration was also run in duplicate, with increasing numbers of decision trees (ranging from 50 to 1500, in increasing increments of 30), over 3 different shrinkage thresholds (0.1, 0.01 and 0.001) with a minimum of 20 observations per node.

The 'best' model was determined according to the overall model accuracy, this was achieved when the 1300 trees were used, with an interaction depth of 5 and shrinkage of 0.01. A comparison of the models produced during the training process is shown in

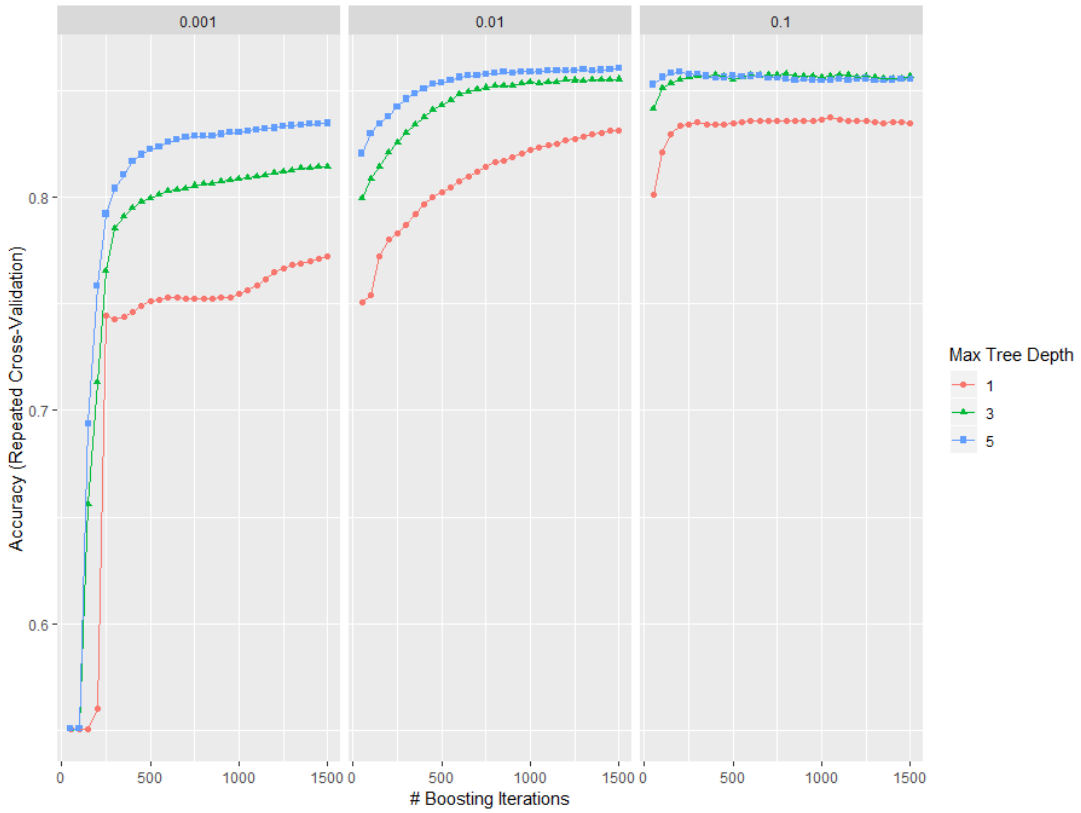

Figure **7-6**.

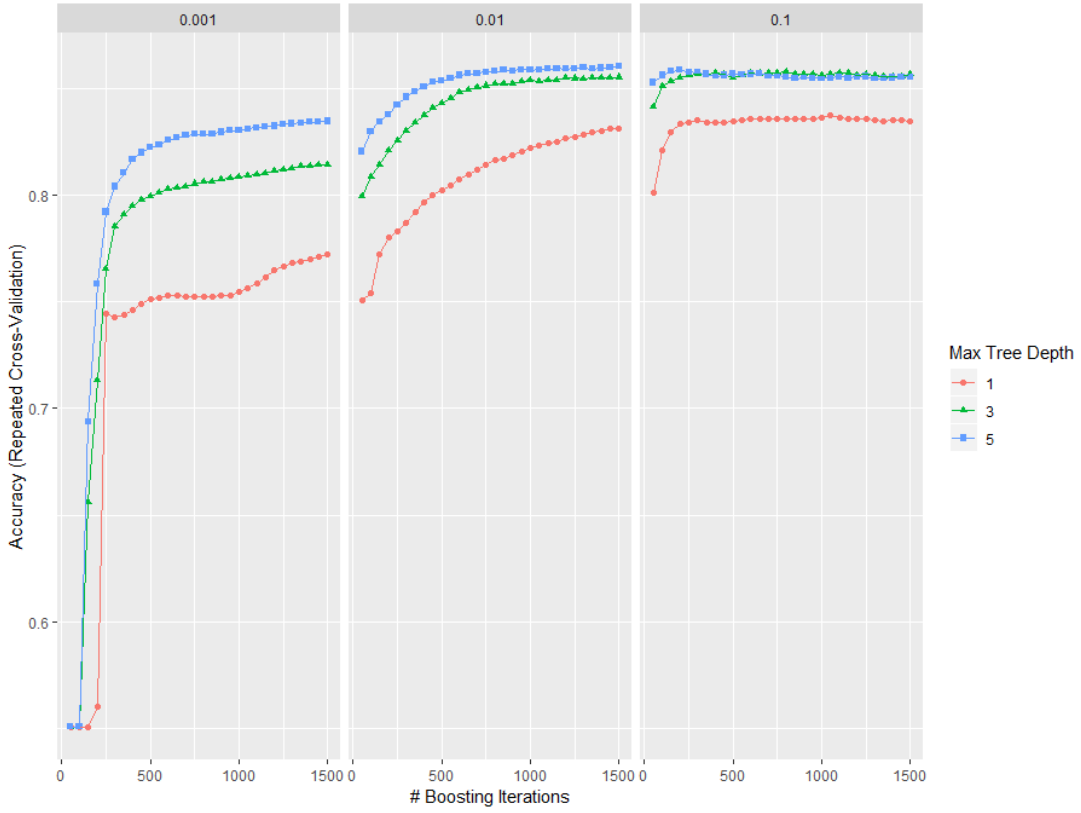

Figure 7-6. Graphical representation showing Gradient boosted model performance during training.

The relative influence of each of the explanatory variables can also be visualised (see Figure 7-7), and shows that Platelet count (relative influence = 43.52), Haemoglobin (relative influence = 16.19) and MCHC (relative influence = 12.06) are the most influential measurements when using this modelling approach.

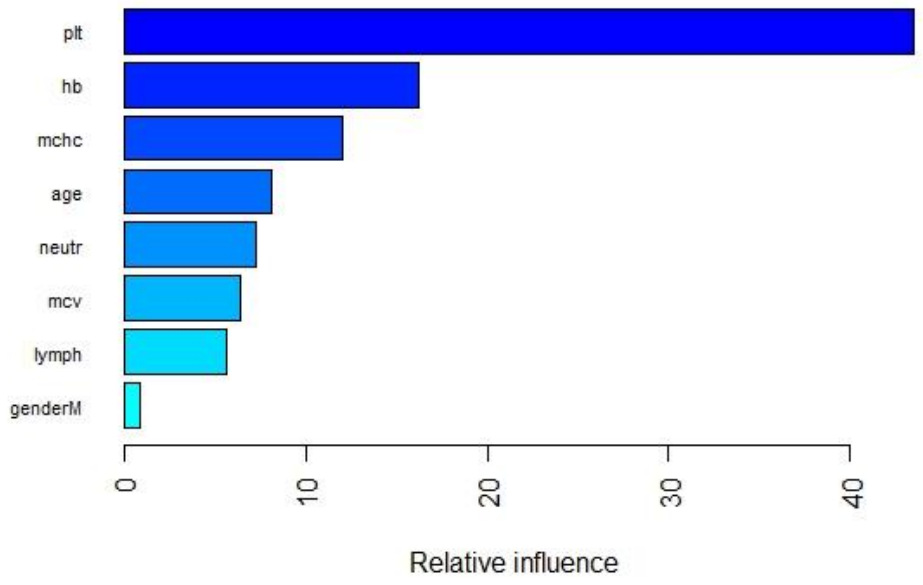

Figure 7-7 Relative influence of each explanatory variable determined by gradient boosted regression analysis.

The performance of the optimised gradient boosted model is shown in Table 7-14. The 'best' gradient boosted model achieved a sensitivity of 90.81% and specificity of 91.99%.

	Actual MPN	**Actual no mutation**
Predicted MPN	929	102
Predicted no mutation	94	1171

Table 7-14. Performance of optimised Gradient Boosted model on training dataset.

7.6.2 Application of gradient boosted model to test data

The optimised gradient boosted model was applied to the test data set described in 7.4.8.1, and performance characteristics calculated. The results of this analysis are shown in Table 7-15.

	Actual MPN	**Actual no mutation**
Predicted MPN	63	37
Predicted no mutation	13	248

Table 7-15. Performance of optimised Gradient Boosted model on test dataset.

The model sensitivity was 63.00% and specificity was 95.02%.

7.6.3 Misclassified Data

Gradient boosted modelling produced the highest number of false negatives of all three approaches when applied to the test data with 13 misclassifications versus 6 produced by the logistic regression model and 2 by the random forest analysis. This is the equivalent of 37.00% of referrals with a *JAK2* V617F or *CALR* mutation failing to be predicted as such by the model. This is a significantly lower level of sensitivity than achieved by the previous two modelling approaches and suboptimal when compared to the use of clinical guidelines alone. For this reason, no further investigation of the misclassified data was performed.

7.7 Discussion

Within the field of haematological malignancies, the reporting of statistical analyses in publications is commonplace. The majority of clinical and biological journal articles contain some form of statistical analysis to indicate the significance of the data within. However, the application of a predictive statistical model in clinical or diagnostic practice has not been previously reported in the diagnosis of MPNs. The aim of this chapter was to investigate the utility of predictive statistical modelling as an alternative or adjunct to existing clinical and diagnostic guidelines.

The current clinical guidelines are skewed towards a high sensitivity for the detection of MPN's. However; this comes at the expense of specificity, with over half of referrals that meet the criteria not having a diagnosable disease. In addition to this, data presented in 5.2, shows that, in practice, the frequency of diagnosis of classical MPNs in the group of patients referred for investigation is very low, at 19%. Follow-up investigations are performed infrequently, with only 10.86% of patients who test wild type for *JAK2* V617F and *CALR* 9 mutations undergoing a bone marrow biopsy. For these reasons, it was felt that this was an area where improved efficiency in diagnostic processes could have significant benefits for both health care providers and patients.

Selecting a suitable classifier approach for the data is a subjective process. There are a vast range of predictive statistical modelling methods available that could, in theory, be applied to the data. Initially, a comparison of eight different supervised classifiers was performed using WEKA. Whilst this approach has not been reported in the context of either classical MPN diagnosis or indeed clinical referral practice in general, it has been applied within the diagnosis of other haematological malignancies, including the analysis of gene expression data in Diffuse Large B-cell Lymphoma (Care et al., 2013; Sha et al., 2015)

The concept of choosing the 'best classifier' is entirely dependent upon the overall goal of the model. In this case, success would be measured against the metrics of the existing clinical/diagnostic guidelines, aiming to maintain sensitivity at the level achieved by the guidelines as a minimum, whilst achieving the maximum specificity possible. This would ensure that the model 'misses' a minimal number of MPNs, whilst minimising the number of patients that would undergo investigation. The general comparison performed in WEKA indicated that logistic regression, random forest and gradient boosted analysis were selected for further development.

The first step of model development was the selection of a suitable data set. The identification of 'cases' for inclusion did not pose any issues and was comprised of data from 1266 referrals which were received by HMDS for the investigation of a suspected MPN and which had a diagnosis confirming this. The selection of a suitable comparison population was less simple. The statistical model was intended to be used primarily within the clinical laboratory setting, with a view to acting as a form of triage to determine which referrals would undergo molecular screening. It was not intended to discriminate between 'normal' individuals from the general population and those with an MPN, but rather to try and identify true MPN's from within the population of patients referred with suspected disease. On this basis, it was decided that a comparison group of 'normal' individuals would not be suitable and instead the control group should consist of patients referred for the investigation of a suspected MPN, who were not confirmed as having the disease.

It was recognised that the use of this comparison population has some drawbacks. Samples referred for the investigation of a suspected MPN currently undergo screening for *JAK2* V617F and *CALR* exon 9 mutations, alongside morphological assessment. However, it is well reported that these two mutations are not present in all cases of MPN, a small proportion will carry a mutation in *JAK2* exon 12 (1-2% of PV cases) or *MPL* exon 10 (3-5% of ET and 5-10% of PMF) and a proportion of MPN patients will not have a demonstrable mutation requiring diagnosis by bone marrow morphology. The low frequency of Exon 12 and *MPL* mutations makes it

economically impractical to perform screening for these mutations on all *JAK2* V617F and *CALR* exon 9 wild type samples in routine practice. Unfortunately, as the data shown in section 5.3 demonstrates, following initial molecular screening, the majority of patients with a non-diagnostic peripheral blood screen do not receive a follow-up bone marrow referral (89.14% of cases) and therefore the presence of an MPN cannot be fully excluded in the majority of cases. The number of patients in which disease has been fully excluded is both small, and as the data demonstrates, the group of patients in which this is achieved are not representative of the referred population as a whole, having significantly different blood count parameters compared to those who were not followed up (Figure 5-3). It is estimated that undiagnosed MPNs account for less than 5% of the comparison group, as this was the most representative sample group available, the decision was made to use it in model development.

A total of 13 explanatory variables were selected for potential inclusion in a statistical model. These were age, gender, haemoglobin, red cell count, haematocrit, mean cell volume, mean cell haemoglobin, mean corpuscular haemoglobin concentration, white cell count, neutrophil count, lymphocyte count, monocyte count and platelet count. The first modelling approach used was logistic regression analysis, prior to model development, was to ensure that the data was in the appropriate form for analysis. Q-Q plots were used to identify variables which required transformation and where necessary this was performed. In addition to these data transformations, a matrix of plots showing each of the variables paired against each other to look for evidence of collinearity between variables. Collinearity occurs where two (or more) variables are measuring the same effect and are therefore not independent of one another.

The initial logistic regression analysis included all the explanatory variables, the model showed statistically significant weightings for all but 4 of the variables. However, the presence of linear relationships as indicated by the paired data plots, would suggest that using all the variables would not produce a robust statistical model. To confirm the presence and magnitude of collinearity in the model, the size

of variance inflation factors associated with each attribute were calculated. This showed prominent levels of multicollinearity between red cell associated variables and between those deriving from white blood cell populations. As stated, multicollinearity occurs when variables are not independent of one another, and as such it was expected as the majority of variables included in the model are related – Hb, RBC, MCV, MCH, MCHC and HCT are all measurements related to the same cell population (erythrocytes) and WBC is a cumulative measure of the neutrophil, lymphocyte and monocyte populations.

To address the collinearity involving white cell parameters, it was decided to remove WBC from the analysis. The rationale for this was that it did not represent a single cellular population and was to some extent a duplication of data contributed by the other WBC related parameters. The decision was also made to remove monocyte count from the analysis as a review of the full blood count data showed that it was not consistently reported by the analyser and was therefore a less reliable variable. The linear relationships seen between the red blood cell related parameters was addressed by removing MCH and HCT from the analysis – the rationale for this was that these parameters were not directly measured by our full blood count analyser and were instead derived from other measured parameters.

The logistic regression analysis was reperformed using the smaller selection of variables and the variance inflation factors were recalculated. Multicollinearity was still present involving Haemoglobin, RBC and MCV. To determine which of these variables should be excluded from the analysis, each was removed in turn and logistic regression analysis performed. Removing RBC from the model resulted in the sufficiently low VIF values and the best predictive power.

Once the issues of collinearity had been addressed, attribute selection was performed using LASSO regression. This technique restricts the absolute size of regression coefficients, with the effect of shrinking those with the lowest values towards zero. In doing so, the variables with the least influence can be removed with the benefit of

creating a simplified model. The 'best' model as determined by LASSO regression did not result in the shrinkage of any of the included variables.

The final model selection would include; Age, Gender, Haemoglobin, MCV, MCHC, Neutrophil, Lymphocyte and Platelet counts as its explanatory variables. Performance metrics were determined for the model and a c index of 0.887 indicated that the model had strong predictive value.

The logistic regression model calculates the likelihood of the outcome (in this model, the likelihood of having a classical MPN or not) as log odds, set on a scale between 0 and 1. The threshold at which the model categorises a sample as a probable classical MPN or not can make a significant difference to the performance of the model in practice. To optimise the performance of the model, the most suitable threshold needed to be determined. There are several considerations to take into account when selecting the optimal threshold for class determination. Firstly, the threshold which gives the 'best' model performance was determined, that is, the threshold which resulted in the highest level of sensitivity and specificity combined – thus resulting in the most 'accurate' model overall. A threshold of 0.56 was calculated as achieving this, and further analysis of the performance metrics showed that using this threshold would achieve a sensitivity of 0.82 and specificity of 0.83.

However, this model was intended to improve upon the performance of the current clinical guidelines that have a sensitivity of 0.87, which is superior to that achieved by the logistic regression model using the 0.56 threshold. The reduced sensitivity of the model would result in an increase in the number of 'cases' missed which would have negative implications for patients. In this case the 'best' model in statistical terms was not the 'best' model for the intended purpose. Instead, a threshold which would give the maximum specificity whilst achieving a minimum sensitivity of 0.87 would be optimal. Performance characteristics were calculated across a wide range of thresholds to manually determine if this was possible. A threshold of 0.7 was able to achieve a higher sensitivity than the guidelines (0.89) whilst achieving an improved specificity (0.73 vs 0.50 achieved by guidelines).

The performance of the model on the case control data set does not necessarily reflect the performance of the model in 'real life'. The data set was selected to have approximately even proportions of MPNs:No mutation samples with a ratio of 0.74:1, this was not reflective of the true incidence of classical MPNs in our referral population where the ratio is approximately 0.2:1. To determine how effective the model would be in practice, it was applied to a separate cohort of 361 suspected MPN referrals was used as a validation data set and performance characteristics measured. The model performance was comparable on the validation data to the training data, achieving sensitivity of 0.92 and specificity of 0.71.

Referrals which were misclassified by the model were identified. Only 6 samples with a classical MPN were misclassified by the logistic regression model, of which 2 were actually known classical MPN patients who were on treatment at the time of investigation and a further 2 had low allelic burden of *JAK2* V617F (<10%), a further case had another coexisting haematological neoplasm which may affect the predictive capabilities of the model. The final classical MPN missed by this model had an isolated thrombocytosis with a platelet count of $569x10^9$/L, well above the upper limits of the normal range. Excluding the two known classical MPN patients would increase the sensitivity of the model slightly to 0.95.

As expected by the lower specificity of the model compared to its sensitivity, the model misclassified a larger number of referrals as MPNs, with a total of 82. Clinical records were reviewed in the 12/82 cases referred from Leeds Teaching Hospitals. The majority (10/12) of these had an underlying reactive condition which would account for the abnormal blood count parameters, and in all cases blood counts did return to normal levels following investigation. Two of the referrals did not have underlying reactive conditions, 1 was deemed to be due to smoking related hypoxia and the final referral had a bone marrow performed which confirmed a diagnosis of triple negative ET.

The development of a random forest model was in many ways easier than logistic regression. The software package used incorporated cross-validation processes as well as optimisation of tree size. In addition to this, random forest analysis is a 'black box' technique, in which the processes taking place between data input and output are hidden from the user. The 'best' model produced using random forest analysis produced a sensitivity of 0.84 and specificity of 0.87. Application of the model to the validation data set produced a higher level of sensitivity (0.97) however, but the specificity was very low at 0.15. Analysis of the MPNs missed by the model (n=2) revealed that whilst 1 of the referrals was from a known ET on treatment (with normal full blood count parameters), however, the second was a referral with an isolated thrombocytosis of $1087x10^9$/L.

A larger number of samples with no mutations were misclassified as MPNs by the random forest analysis (n=242). Thirty-three of these were from Leeds Teaching Hospitals and clinical documents were reviewed in all cases. Suspected polycythaemia was indicated in 16/33, 10 of these were attributed to smoking and/or alcohol use, and 2 were thought to be a side effect of testosterone therapy. The remaining 4 referrals with suspected polycythaemia were without identifiable cause, all of which resolved following investigation. In the 17 non-polycythaemic false positives, 4 were thought to be due to iron deficiency related thrombocytosis, 4 were due to chronic reactive conditions including rheumatoid arthritis, 4 following thrombotic events, 1 due to prior splenectomy and 1 was in a patient who had been previously diagnosed with refractory cytopenia with multi-lineage dysplasia and fibrosis. The remaining 3 had no identifiable cause for investigation and underwent bone marrow analysis. ET was diagnosed in one case, systemic mastocytosis was found in a further individual and the final case showed reactive changes only.

Gradient boosted modelling initially appeared to offer the best levels of specificity (89.82%) and sensitivity (91.21%) with both exceeding that achieved by the use of current clinical guidelines alone (best sensitivity 88.07% (WHO), best specificity of 49.55%(BCSH)). Model optimisation was the most complex of the three techniques and required the longest computational processing time. However, when applied to

the test data set, the model performance deteriorated significantly and had much poorer predictive power than either logistic regression or random forest analysis.

In summary, the logistic regression model performed better than both the random forest analysis and gradient boosted model in the classification of suspected MPN referrals. The overall performance was superior to that achieved using the clinical guidelines alone and if used, may reduce significantly the number of patients undergoing investigation for suspect MPNs. In addition, the use of logistic regression has the added benefit of weighting each variable allowing for some evaluation of how different biological parameters contribute to the probability of disease. The data within this chapter would indicate that MCHC and lymphocyte count have the biggest weighting within the model, both in an additive way. In real terms, higher MCHC and lymphocyte counts increase the likelihood that a referral will not have a mutation of *JAK2* or *CALR* detected. There are no publications that specifically identify lymphocyte counts as being of significance in MPN diagnosis. However, reduced B cell progenitors are a frequent finding in myelodysplastic syndrome (Jacobs et al.) and it could be hypothesised that lymphocyte count may be an indicator of overall haematopoietic function. Likewise, MCHC may be a general indicator of effective haematopoiesis. Increased values for the other scalar explanatory variables or male gender act in a subtractive manner and increase the probability that a referral will be classified as a case.

The comparable performance of the model on both the test and validation cohorts is a good indicator that the model is robust and fit for purpose. To assess the impact that the implementation of this model would have in practice, a further data set of referrals of suspected MPNs will be analysed using the model in the next chapter.

8. Assessing the impact of utilising a predictive statistical model in routine diagnostic practice

The development and validation of the predictive logistic regression model for the diagnosis of MPNs demonstrates a clear statistical improvement over the use of clinical guidelines alone. However, the use of such models in clinical and/or laboratory practice has not been reported in this field and the acceptance of such an approach by clinicians and patients may be challenging. It requires a change in approach and will contradict both current diagnostic guidelines as well as modifying current clinical practice.

Demonstrating that the model performance is superior to current methods goes partway towards justifying its use. In addition to this, it must also be shown that it does not compromise patient safety, and that implementation would not have a detrimental effect on workload, resources and in the case of HMDS, income generation. In this chapter, the impact of utilising the predictive logistic regression model will be examined in more detail.

8.1 The data used to assess impact

As detailed in 3.5.2.5, all referrals received by HMDS for the investigation of suspected MPNs during the 2016 calendar year were selected for assessment of the model in 'real life'. Sample group 6 was initially composed of 2174 referrals, with 438 confirmed diagnoses of MPN and 1740 with no mutation detected. Following removal of duplicate patient referrals and exclusion of those in which clinical details did not indicate a suspected MPN (therefore were not screened for *JAK2* V617F or *CALR* mutations) or where peripheral blood was not received for full blood count analysis, the data set was reduced to 1581 referrals (see Figure 8-1).

The data set contained; 801 male and 780 female referrals (ratio 1.03:1) aged between 16 and 94 years of age (median age – 61 years). A diagnosis of MPN was confirmed in 305/1581 referrals with the remaining 1276/1581 referrals having no mutation in *JAK2* or *CALR* detected.

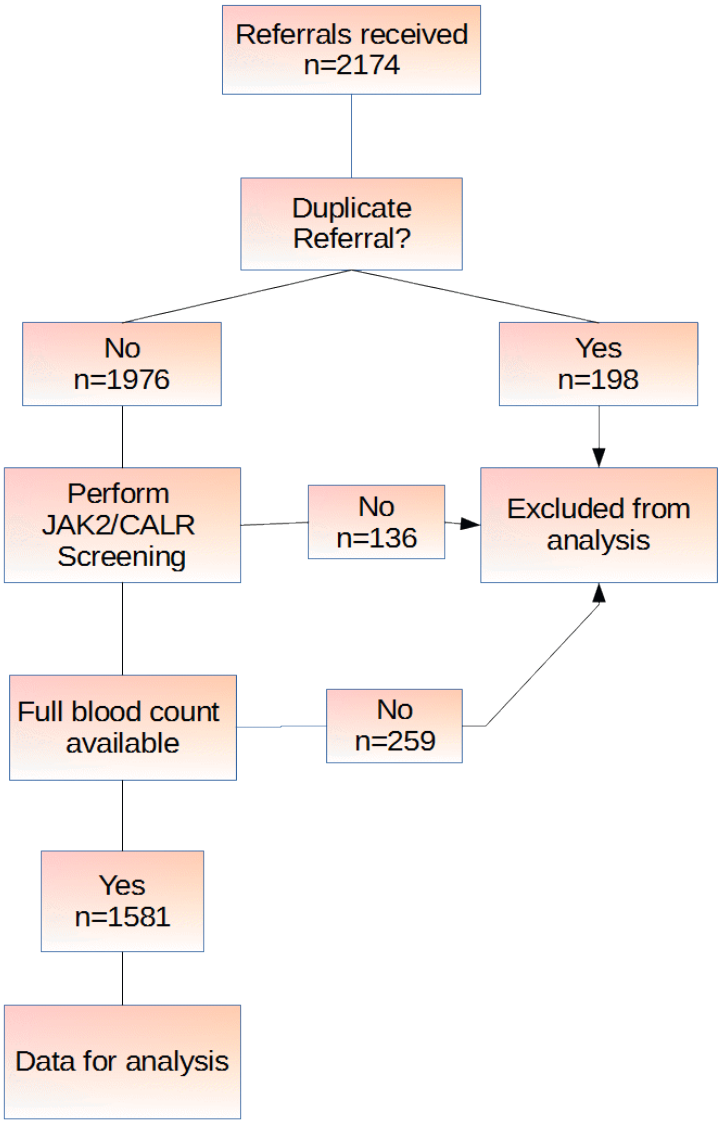

Figure 8-1. Selection of referrals for analysis of model impact in routine diagnostic practice.

8.2 How would the model be used in practice?

This model relies upon accurate full blood count data, and as described in 3.3, it is not practical to perform full blood count analysis prior to consultation at the patient's initial visit to clinic. Therefore, in the first instance, the predictive model has been designed to be used within HMDS (or a similar laboratory) as a form of sample triage, although it could also be used in a clinical setting if desired. The statistical model is intended to guide and support existing practice rather than replace it and so it was envisioned that the prediction produced by the model would be used alongside morphological assessment to indicate whether a specimen underwent further investigations following full blood count assessment.

Using the optimised threshold of 0.7 (as described in 7.4.7.2) referrals which were classified as an MPN by the model (predicted risk <0.7) would undergo molecular screening for *JAK2* V617F and *CALR* exon 9 mutations. Those which returned a predicted risk of >=0.7 would be predicted to not be an MPN and would not be investigated any further. An exception to this, would be where there were clinical features recorded (such as splenomegaly or unexplained thrombotic events) or morphological features (for example, the presence of tear drop poikylocytes) were indicative of an underlying pathological process.

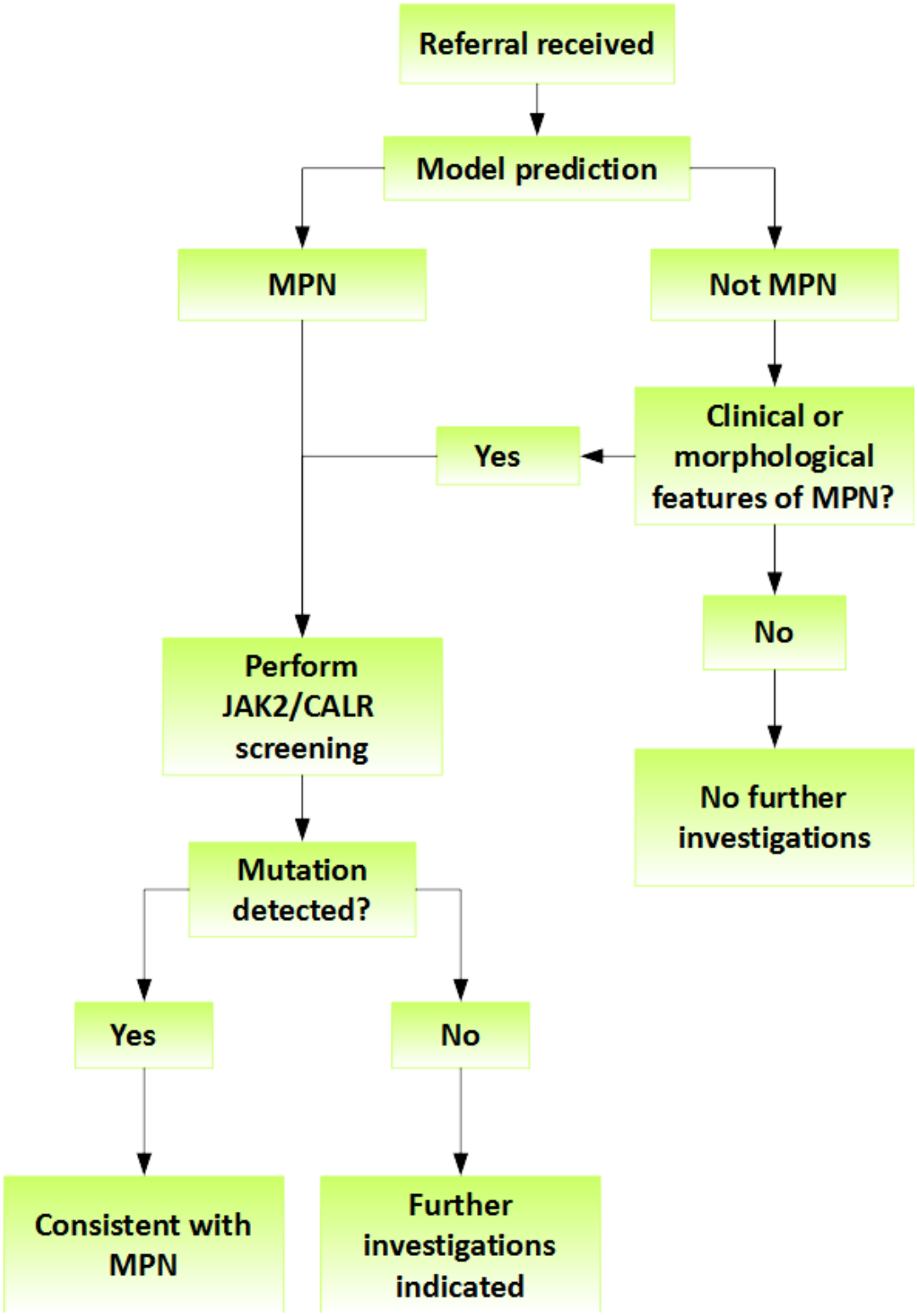

Figure 8-2. Proposed workflow diagram incorporating predictive modelling.

8.3 How would the application of predictive logistic regression model have impacted on workload in 2016?

The logistic regression model developed in chapter 7 was applied to the data set described in 8.1. Referrals were categorised as 'MPN' and 'not MPN' based on the model's prediction. This resulted in 37.13% (n=587) referrals being categorised as 'MPN' and the remaining 62.87% (n=994) referrals being classed as 'not MPN' by the model.

8.3.1 Model performance

Using the performance metrics detailed in previous chapters, the logistic regression model showed a sensitivity of 0.88 and specificity of 0.75.

	Actual MPN	**Actual no mutation**
Predicted MPN	267	320
Predicted no mutation	38	956

Figure 8-3. Performance of predictive model on suspected MPN referrals received during 2016.

8.3.2 Would the genuine MPN referrals have been missed in practice?

When the data was analysed further, 38 of the referrals which were predicted as not being MPNs, which were subsequently found to have an MPN. As in section 7.4.9.1, the request data, and where possible, clinical records were reviewed. Upon further investigation 14/38 false negatives were found to have a previously

diagnosed MPN and were on treatment at the time of referral, accounting for the normal full blood count data.

Of the remaining 24, 5 had been referred due to unexplained thrombotic events, clinical features which have been prompted investigation according to the workflow diagram outlined (Figure 8-2). Two referrals were concurrently undergoing investigation for other possible haematological malignancies. In one case, a patient was suspected as having MDS or PMF due to worsening cytopenia and the clinical details supported investigation. In the second case, the patient was referred with clinical details of suspected PMF and/or a lymphoproliferative disorder and morphological features supported the suspicion of PMF which would have resulted in investigations proceeding, despite the prediction of the statistical model.

One of the remaining false negatives did not demonstrate *JAK2* V617F or *CALR* exon 9 mutations in the specimen used in this analysis and was initially reported as non-diagnostic, however, morphological examination of the peripheral blood had evidence of tear drop poikylocytes which would have prompted investigation. A subsequent bone marrow received on this patient confirmed the diagnosis of an MPN.

A further 3, presented with elevated haemoglobin and haematocrit but no additional clinical information was available to indicate whether this was persistent or whether other causes had been excluded. As *JAK2* mutations are found in >95% of cases of PV, it is likely that these would have been investigated despite the model classification as the exclusion value of an unmutated result is very high.

An isolated mild thrombocytosis was present in a further 6 referrals, this was documented as being persistent by the referring clinician in 2/6 of these, the other 4 referrals would not have been investigated any further.

The final 7 referrals in this group had normal full blood count indices and, without exception, these referrals were received with no additional clinical information and would not have undergone further investigation as a result of the model prediction.

Overall, a total of 11 individuals with a demonstrable mutation would have been rejected by the laboratory on the basis of the prediction provided by the statistical model combined with the absence of clinical and morphological evidence to support further investigation.

Removal of the known MPNs on treatment and reclassification of those which would have undergone investigation on the basis of clinical and morphological assessment adjusted the performance to a sensitivity of 0.96 and specificity of 0.75.

8.3.3 False positives

The predictive logistic regression model misclassified 320 non-diagnostic referrals as an MPN. Request data and clinical records were examined in the 24/320 referrals received from Leeds Teaching Hospitals. Subsequent bone marrow assessment was performed in only 5/24 of these referrals; an MPN was not confirmed in any of these instances. Another haematological malignancy was found in 2/24 individuals (multiple myeloma n=1 and hairy cell leukaemia n=1). Reactive causes were diagnosed in 11/24 referrals with a further 3/24 thought to be caused by lifestyle choices (smoking and alcohol related). In the remaining 3 referrals, no identifiable cause was found and at the time of writing, no further investigations had been performed.

8.4 What would the practical impact of using the model have been?

8.4.1 Cost burden analysis

To establish the impact that adopting a predictive model would have had financially over this time period a cost-burden analysis was performed. This is achieved by considering each of the steps undertaken in the laboratory during the process of investigating a suspected MPN and calculating the associated costs. Full details of the cost breakdown are included in the appendix (Figure 10-1 through Figure 10-6). The total cost of processing a suspected MPN referral in the laboratory was calculated to be £33.97p.

During 2016, the 1581 referrals included in this analysis were processed at a total cost of £53,706.57p to the laboratory. The implementation of the predictive model developed in this work would have resulted in only 601 referrals of these referrals being investigated, at a total cost of £20,415.97p to the laboratory. This would have resulted in an overall cost saving of £33,290.60p. With estimated savings of around £170,000 over a 5-year period.

8.4.2 Impact on income generation

The cost savings described in 8.4.1 represent the expenditure within the laboratory, however, this is only part of the picture. Every referral processed by HMDS for investigation of a suspected MPN currently incurs a charge for the referring hospital of £200. It is proposed that, following implementation of the statistical model, referrals that did not meet criteria for full investigation would incur a reduced charge of £100. Based on the 2016 referral figures, 960 referrals would have been invoiced at the lower charge, which would result in a reduction in income of £96000.

8.4.3 Impact on staff time

Reducing the number of specimens undergoing laboratory investigations would decrease the amount of staff time required to provide this service. Cost burden analyses estimate that each referral requires approximately 1hr of staff time to

process and report. A reduction of 960 specimens would free 18.85 hours of staff time on average per week in HMDS.

8.4.4 Equipment availability

In addition to the staff time that could be released, a reduction in workload would also decrease the amount of equipment time requirement. As shown in Table 8-1, this would free over 290 hours of equipment time annually.

Equipment	Requirement/sample(s)	Saving/annum (based on 960 fewer referrals)
Automated DNA extraction	40 minutes / 12 samples	53 hours and 20 minutes
PCR protocol	2 hours / 24 samples	80 hours
Fragment analysis	40 minutes / 4 samples	160 hours

Table 8-1. Potential equipment usage savings associated with reduced referral numbers.

8.4.5 Clinical savings

In addition to the financial and time savings that can be made in the laboratory, there are also potentially significant savings for clinical departments. Implementing the use of predictive modelling tools in the haematology clinic setting, prior to referral to

HMDS, would have allowed 960 patients to be discharged to primary care following their initial visit. At a cost of £109 per appointment (2017/18 NHS National Tariff), this would have saved £104,640.

8.5 Developing a user interface

The wider implementation of this predictive model requires some form of user interface. To this end, I have developed a web-based application into which a clinician or scientist can input the relevant values from patient details and full blood count results. The application then uses the model developed in this work to calculate the likelihood of the patient having an MPN and displays this to the user alongside a statement indicating whether the result meets the criteria for further laboratory testing. The model can be access through the following web address:

https://classicalmpnpredictor.shinyapps.io/MPNpredictor/

Screenshots follow which show the user interface prior to the input of new data (Figure 8-4), along with examples of the output when the model predicts a high likelihood of the patient having an MPN (Figure 8-5) as well as a low likelihood prediction (Figure 8-6).

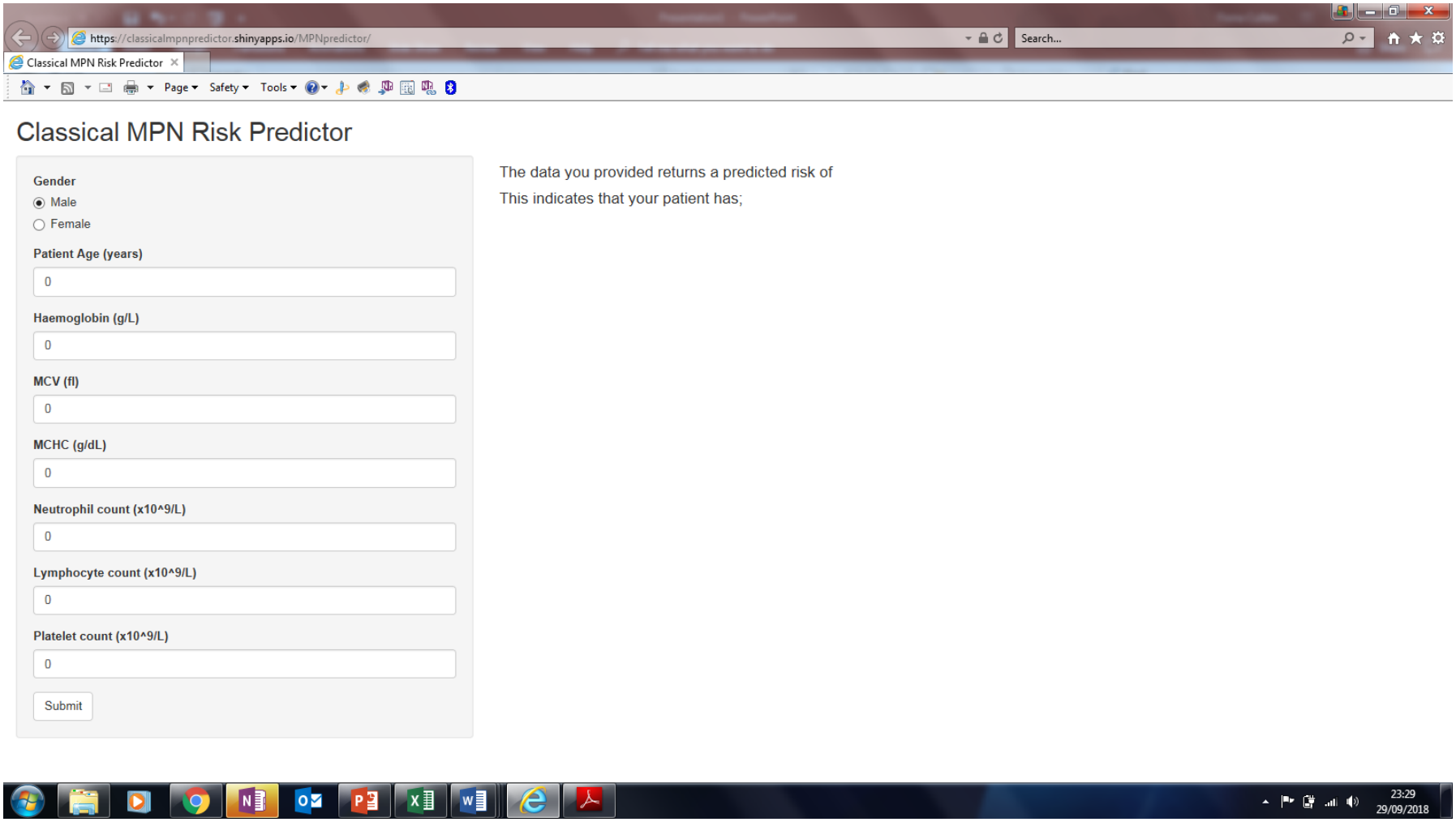

Figure 8-4 Screenshot (1) User interface prior to input of new patient data

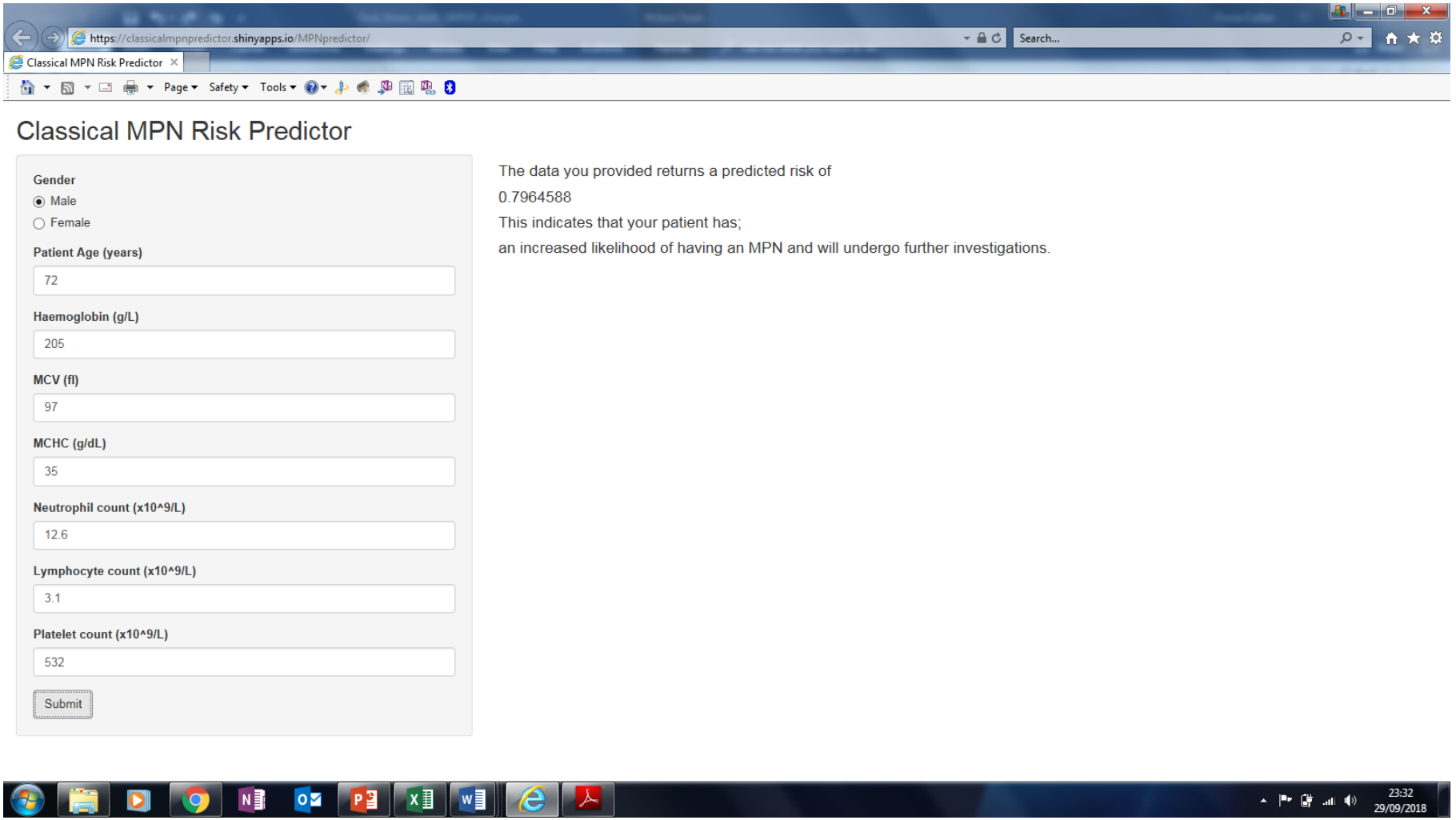

Figure 8-5 Screenshot (2) User interface output when prediction is high likelihood of MPN

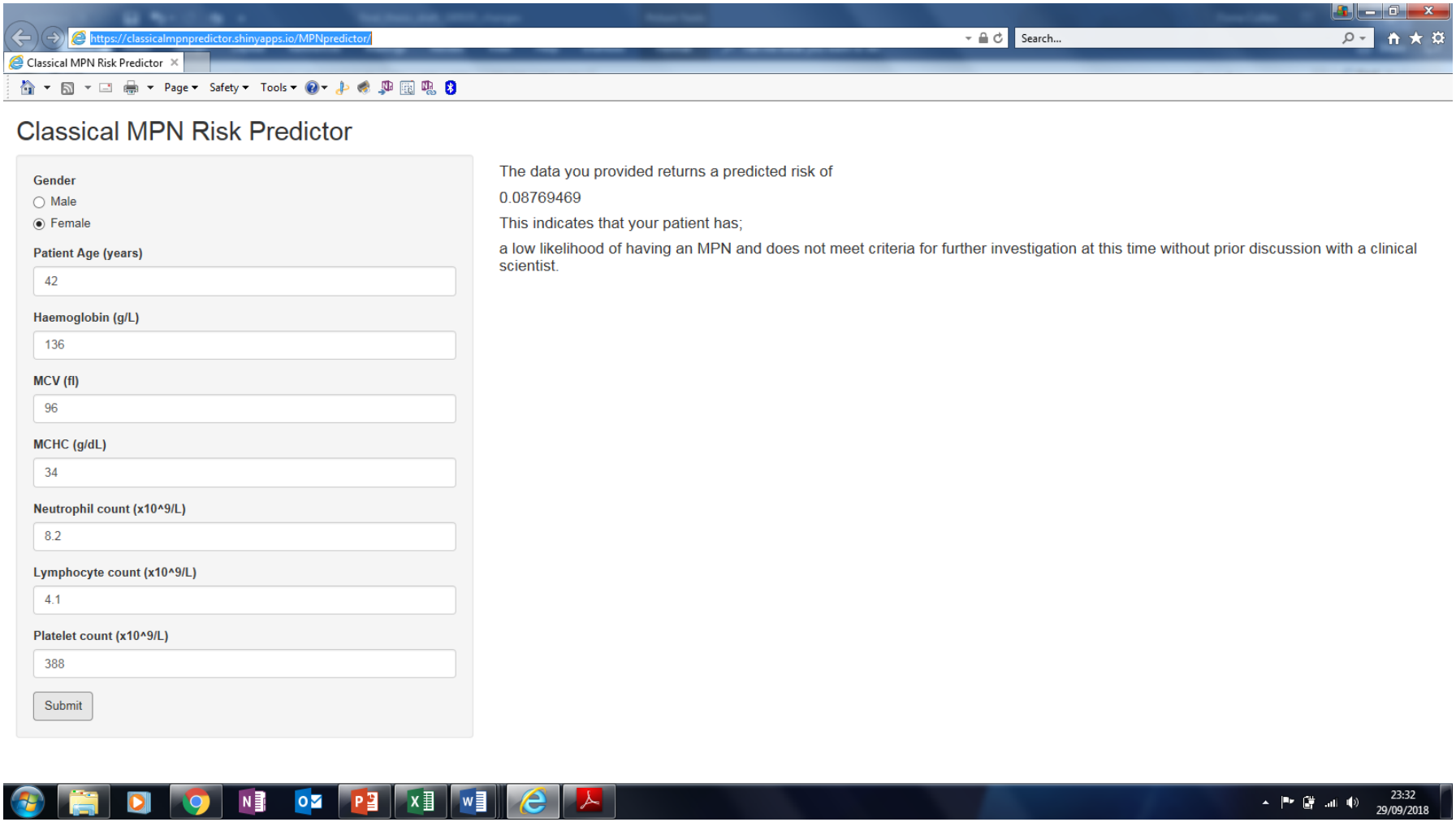

Figure 8-6 Screenshot (3) User interface when prediction is low likelihood of MPN

8.6 Discussion/summary

During the 2016 calendar year, 1518 referrals were made which were suitable for analysis by the model. Whilst that does not represent the total number of referrals that underwent *JAK2* V617F and *CALR* exon 9 mutation screening, it does comprise the total of referrals which were recorded as being made solely for the investigation of a suspected MPN.

The initial model performance is comparable with that reported during both its development and validation. With both sensitivity and specificity exceeding the level achieved using the criteria set out by current published guidelines alone. Furthermore, during the investigation of false negative predictions, it was found that a significant proportion of misclassified referrals were either known MPN's on active therapy or had clinical/morphological features which would have indicated that mutational screening was justified. Removal or reclassification of these referrals from the analysis further improved the sensitivity of the model.

In analysing the false negative predictions, 7/11 of the cases would have been rejected under the proposed system had normal full blood count parameters. The finding of a mutation alone, would not have been sufficient to meet current diagnostic criteria and therefore the diagnosis of an MPN could not have been made. It could also be argued that these referrals would not have been made in the first instance, were the referring clinician following current recommended guidelines. It is also recognised that mutations including *JAK2* V617F and those found in *CALR* exon 9 are found in both normal individuals (Jaiswal et al., 2014), and that whilst uncommon, can also be found in disorders outside of the classical MPNs (Linda M. Scott, 2014; Scott et al., 2005). This would suggest that in isolation, the presence of a mutation should be treated with caution.

Statistically, the predictive model has consistently demonstrated a higher level of sensitivity than the use of the clinical guidelines, which indicates that the number of false negatives produced by the model is lower than expected using the guidelines alone. Whilst, there may be differences in the specific individuals who would have been missed by each approach, the overall impact is lower when the statistical model is used. As such, there is no evidence that there would be a negative impact on patient safety.

The utilisation of the proposed predictive statistical model would have had a significant financial impact. Considering laboratory and clinical costings, the use of predictive modelling would have had an estimated overall cost saving of £52,490.60p during the 2016 period, with projected savings in excess of £262,000 over a 5-year period. In addition to the financial savings, 960 clinic appointments could have been made available and HMDS could have reallocated 290 hours of equipment resources and a 0.5 whole time equivalent member of staff to alternative service requirements.

9. General Discussion

9.1 Overview

This thesis is the first reported example of predictive statistical modelling being used to determine the appropriateness of investigations in suspected myeloproliferative neoplasms. The clinical laboratory setting in which this work has taken place allowed access to an unparalleled dataset reflecting a wide population of patients undergoing such investigations. The methods used in this work demonstrate that current referral behaviour could potentially be improved and that adherence to published guidelines is often impractical in a clinic setting. Furthermore, this work has objectively quantified the efficiency of clinical guidelines and highlights the potential opportunity to re-evaluating the current scheme. The data presented proposes an alternative approach to the identification of patients with probable classical MPNs by applying statistical modelling to biological variables, some of whose utility is previously unreported in the context of classical MPNs.

9.2 Molecular screening for classical MPNs

The MPNs are a heterogenous group of disorders classified together under a single term. They exhibit a range of molecular defects, but none of these are disease defining. Whilst upwards of 99% of confirmed cases of PV have a demonstrable *JAK2* (V617F or Exon 12) mutations, prior to the identification of mutations in the *Calreticulin* gene, mutations of *JAK2* V617F and *MPL* were identifiable in little more than half of those individuals with confirmed ET or PMF. It is estimated that *CALR* mutations increase the proportion of ET and PMF diagnoses with a demonstrable mutation to over two thirds, this still leaves a significant proportion of

cases without an identified molecular aberration. These patients would require bone marrow assessment to confirm the presence of disease.

At the outset of this work, mutations in the *CALR* gene had not been identified and the discovery of mutations in this gene in 2012 altered the course of this research. The omission of such a high frequency mutation from statistical analysis would have significantly limited the impact of this work therefore the development and validation of a diagnostic assay for identifying mutations in this gene was critical, not only for the meaningful completion of this thesis, but also for routine clinical laboratory use. Data presented within this thesis highlights the significant work that is required for the development and validation of a new assay for use in a diagnostic laboratory. Mutations as significant, or with as high frequency, as *CALR* are rarely identified, and this work also offers a unique insight into the impact of introducing new molecular assays on referral behaviour

There was very limited information about the clinical and laboratory features of *CALR* mutations at the time that this work was undertaken. The only published data available on the features of CALR mutations had been derived from patients with a WHO confirmed diagnoses of ET in whom bone marrow morphology was a prerequisite. As, from our data, bone marrow aspiration is performed on younger patients with higher platelet counts than seen in the overall population referred for investigation of suspected MPN, it was hypothesised that the published data may have been naturally biased towards this cohort of patients. This could potentially account for the published features associated with *CALR* mutations (younger age, higher platelet count).

This work set out to determine whether the clinical and laboratory findings from the publications were also seen in the population of patients referred to a diagnostic laboratory. Data presented in this work demonstrates that both the clinical and laboratory features of *CALR* mutated patients described in the literature were consistent with those found in a routine diagnostic laboratory. Having confirmed

that our population of MPN patients are comparable to those identified in other centres, I am confident that conclusions drawn from this work will be applicable on a wider scale.

9.3 Clinical practice - guidelines vs reality

The identification of molecular markers of disease should, in theory, improve the efficiency of diagnostic practice. This does not, however, appear to be the case in the diagnosis of the myeloproliferative neoplasms. The data presented in this thesis show that the introduction of molecular screening tests has been accompanied by significant increases in referral numbers but has had negligible impact on the number of confirmed diagnoses.

It was hypothesised that the increasing disparity between referrals and diagnoses may be reflective of high numbers of inappropriate referrals. The auditing of clinical practice showed that, for logistical reasons, *JAK2* V617F and *CALR* mutational screening is being requested without first confirming the reported abnormal blood counts. In patients where the clinical suspicion is low, the results appear to be treated as a test of exclusion, rather than as a support tool to restrict investigations to those patients in whom true clinical suspicion is present and who demonstrate persistently abnormal blood counts. Indeed, published data indicates that a third of ET and PMF patients will not demonstrate a mutation of *JAK2* V617F or *CALR*, however in our cohort less than 10% of referrals with no evidence of a *JAK2* V617F or *CALR* mutation in the peripheral blood screens were investigated further.

9.4 Predictive modelling

Clinical guidelines for the investigation of suspected myeloproliferative disorders have evolved over the past 50 years. However, during this time there has not been a critical appraisal of their performance. Amendments to the guidelines have largely been driven by either the addition of genetic markers or adjustments in FBC parameter thresholds and the basic structure of the criteria have remained unchanged. Furthermore, it is unclear how the original diagnostic variables and thresholds used in the criteria were originally identified.

HMRN has demonstrated that the socio-demographic profile of the population HMDS serves is comparable to the national average. Furthermore, the HILIS database contains biological and demographic data on each referral. Given that, from 2005 to 2017, HMDS investigated over 18,000 cases for suspected classical MPNs, the data used within this thesis offers an unparalleled data set for the development of a predictive statistical model. Indeed, using logistic regression analysis of full blood count parameters in conjunction with demographic information, it is possible to correctly classify over 95% of referrals who will demonstrate the presence of a *JAK2* V617F or *CALR* mutation. In contrast to current referral practice whereby >80% of referrals do not have a *JAK2* V617F or *CALR* mutation, the classifier would remove cases from further referral and reduce this figure to <30%.

9.5 Implementing changes in clinical practice

Initially, it is intended that the predictive model developed in this work, will be used as an adjunct to current screening protocols in the diagnostic laboratory. Full blood count and demographic information from suspected classical MPN referrals will be inputted into the app and the classified as having a either a high or low likelihood of having a classical MPN. Those classified as having a low likelihood will not be investigated unless clinical or morphological evidence indicates otherwise (such as thrombotic complications, hepatosplenomegaly or morphological abnormalities (tear-drop poikylocytes or leucoerythroblastic features). Samples classified as having a high likelihood of being an MPN, will undergo screening for *JAK2* and *CALR* mutations. Those in which a *JAK2* V617F or *CALR* mutation is detected will

be reported as per current protocols. Whereas those in which neither mutation is identified without would be reported as having no mutation in *JAK2* V617F/*CALR* with a comment to indicate that if there is no reactive cause for clinical/laboratory features, a bone marrow would be indicated.

Adopting such an approach during 2016 would have a significant impact in HMDS, saving over 290 hours of equipment time and the equivalent of a 0.5 whole-time equivalent staff member time each year. The reduction in workload associated with adopting this predictive model would reduce the departmental income by approximately £96000 per annum. However, the staff time and equipment made available would allow other areas of the service to expand and recoup lost income. Furthermore, all patients with suspected classical MPNs are seen in clinic on a minimum of 2 occasions. If the use of this model could be incorporated into clinical practice and used by clinicians prior to referral to the laboratory, it would have reduced this to a single clinic appointment in over 60% of referrals received in 2016. Using the figures stated in chapter 7, this could have potentially saved 994 clinic appointments in 2016, at a cost of £120 per appointment (according to the NHS national tariff 2016/17) this could have saved an additional £119,280, as well as decreasing pressure on clinical services and reducing patient anxiety. Taking both laboratory and clinical costings into account there are potential overall savings for the NHS in the region of £52,000 per annum.

The integration of predictive statistical modelling into routine clinical practice is the ultimate endpoint of this work. In doing so, it would offer greater benefits, in terms of financial savings (both laboratory and clinical) and patient pathway efficiency (reduced waiting times and appointment requirements) than restricting it's use to the diagnostic laboratory. The app developed in this work offers a simple, accessible way for clinicians to quickly establish the appropriateness of testing in their patients. However, successful implementation would require some modifications to current working practices to ensure that the data inputted was up to date and accurate. This

would require clinician engagement and a willingness to alter the way in which patients are seen in clinic. Demonstrating the value of using this model in a prospective cohort will build clinician trust and it is hoped that this will lead to adoption of this model in the diagnostic pathway for patients with suspected classical MPNs.

9.6 Impact on patient experience

It is acknowledged that many patients respond positively to an active investigatory approach. However, this does not necessarily mean that *JAK2* V617F and *CALR* exon 9 mutational screening should be seen as a preliminary investigation. Data from the clinical audit presented in section 5.5 showed that in more than half of referrals audited, persistent abnormal blood count parameters were demonstrated prior to requesting *JAK2* V617F and *CALR* exon 9 mutational screening. In these patients, an active testing approach could be employed without the inclusion of mutational screening.

One can appreciate that requesting *JAK2* V617F and *CALR* mutational screening during the initial clinic consultation may save time. Hospital services are under mounting pressure to reduce waiting times and reduce costs. These pressures may contribute to the premature requesting of *JAK2* V617F and *CALR* mutational screening. The proposed implementation of the model during the lab following consultation and for the haematology laboratory to then use the modelling app to determine whether the specimen be referred onwards for *JAK2* V617F and *CALR* exon 9 mutational screening. Alternatively, if point of care testing were available in the haematology outpatient clinic, it may be feasible to perform the full blood count prior to consultation and for the clinician to use the app to determine the appropriateness of further investigation themselves.

9.7 Further work

The data used to develop this model is taken from a restricted geographical population. HMRN publications show that this population is comparable to that of the national population but demonstrating the performance of this model on an independent geographical cohort would be valuable. As such, it is intended that the model will be applied to a series of independent referrals, possibly in a different geographical location to confirm its suitability.

The inclusion of *MPL* and *JAK2* Exon 12 mutations in this work was restricted by the low frequency of bone marrow aspirates received. As a pilot study, a recent HILIS database search identified a total of 27 *MPL* exon 10 or *JAK2* exon 12 mutated MPNs which were suitable for assessment using the predictive model app. The model predicted the correct class membership in 24/27 cases (sensitivity = 88.89%). Whilst encouraging, the sample size is insufficient to draw any meaningful conclusions. Further validation is planned to be performed on a dataset tested for all 4 mutations (*JAK2* V617F, *JAK2* exon 12, *MPL* and *CALR*) thereby helping to establish the value of this model in the identification of patients who would benefit from upfront *JAK2* exon 12 and/or *MPL* mutation analysis.

The investigation of suspected classical MPNs is one of many potential applications of predictive statistical modelling in a haematological setting. This could potentially include a similar approach utilising FBC parameters in conjunction with patient

demographics in a range of conditions. For example, to better target the population referred for screening in suspected Paroxysmal Nocturnal Haemoglobinuria (PNH), or the identification of a reactive eosinophilia or monocytosis from their malignant counterparts. Furthermore, given the diagnostic challenge in the identification of MDS from other non-clonal cytopenias, or clonal haematopoiesis of indeterminate potential, full blood count data could be supplemented with additional laboratory data (for example immunophenotypic characteristics). Following completion of this work, it is anticipated that further research could be undertaken in these areas.

In conclusion, predictive statistical modelling has the potential to beneficially alter the way in which samples are referred for laboratory investigations in suspected malignant conditions. This work provides potential evidence of the positive impact this approach would have in the investigation of suspected classical MPNs. Demonstrating that better targeting the population for investigation would reduce both laboratory and clinical workloads, decrease clinic waiting times and would result in significant financial savings. Furthermore, this would allow resources to be directed towards those clinically appropriate investigations which would ultimately improve patient care.

10. Appendices

10.1 Materials

Description	Catalogue/Product number	Manufacturer
Albumin, Bovine	A-3059	Sigma Chemical Company
Ammonium Chloride	55	Vickers Laboratory Ltd
AmpliTaq Gold DNA polymerase kit	4311816	Applied Biosystems
Ethanol	P20809	VWR International
Facsflow	342003	Becton Dickinson
Genescan 500 Rox dye Size Standard	4310361	Life Technologies
Hi-Di Formamide solution	43311320	Applied Biosystems
QIAamp DNA Mini Kit	5136	Qiagen Ltd
UltraPURE, DNAse, RNAse free Water	10977-035	Invitrogen
Primers	custom product	Sigma-Genosys
100mM dNTP mix	39026	Bioline
Molecular biology grade water	10977-035	Gibco

Table 10-1 Details of reagents used in fluorescent fragment analysis experiments

Target	Sequence
***JAK2* COMMON REVERSE**	56-FAMN/CTG AAT AGT CCT ACA GTG TTT TCA GTT TCA
***JAK2* WILD TYPE SPECIFIC FORWARD**	ATC TAT AGT CAT GCT GAA AGT AGG AGA AAG
***JAK2* MUTANT SPECIFIC FORWARD**	AGC ATT TGG TTT TAA ATT ATG GAG TAT ATT
***CALR* EXON 9 REVERSE**	AGA CAT TAT TTG GCG CGG
***CALR* EXON 9 FORWARD**	5HEX/TGA GGT GTG TGC TCT GCC T

Table 10-2 Primer sequences for amplification and fluorescent fragment analysis of JAK V617F and *CALR* exon 9 mutations

REAGENT	CONCENTRATION	VOLUME/40 TESTS
Ultrapure H_2O		829 µL
ABI 10x gold buffer	1x	100 µL
***JAK2* fluorescent reverse primer**	5pmol final concentration	2 µL
***JAK2* wild type forward primer**	5pmol final concentration	2 µL
***JAK2* mutant forward primer**	5pmol final concentration	2 µL
DNTP	125µM final concentration	5 µL
$MgCl_2$	1.5mM final concentration	60 µL
***Taq* polymerase**	0.5 Units	0.1µL
DNA		

Table 10-3 Reagent list for *JAK2* V617F mutations screen PCR reaction

REAGENT	CONCENTRATION	VOLUME/40 TESTS
Ultrapure H_2O		831 µL
ABI 10x gold buffer	1x	100 µL
***CALR* fluorescent forward primer**	5pmol	2 µL
***CALR* reverse primer**	5pmol	2 µL
DNTP	125µM final concentration	5 µL
$MgCl_2$	1.5mM final concentration	60 µL
***Taq* polymerase**	0.5 Units	0.1µL
DNA		

Table 10-4 Reagent list for *CALR* exon 9 mutation screen PCR reaction

STEP	TEMPERATURE	TIME
Pre-activation	95˚C	10 minutes
35 cycles as follows:		
Denaturation	95°C	30 seconds
Annealing	58°C	30 seconds
Extension	72°C	30 seconds
Followed by:		
Final step		
Extension	72°C	7 minutes
Hold	4°C	∞

Table 10-5 Thermal cycle settings for combined *JAK2*/*CALR* PCR.

10.2 R Scripts for Statistical Model Development, Testing and Validation

10.2.1 Data preparation

```
# LIBRARIES REQUIRED TO PERFORM ANALYSIS #
library(MASS)
library(faraway)
library(rgl)
library(caret)
library(gmodels)
library(ROCR)
library(rms)

#READ IN DATA#
dfr <- read.csv("modelling_data_v1.csv", header=T)

#CHECK DATA STRUCTURE, LEVELS OF FACTORS#
str(dfr)

#OVERVIEW SUMMARY OF ALL DATA#
summary(dfr)
```

```
#SELECT COHORT OF SAMPLES WITH FULL MOLECULAR ANALYSIS AND COMPLETE FULL BLOOOD COUNT DATA#

complete <- dfr[dfr$molecular!='JAK2_NEG',]

summary(complete)

idx1 <- is.na(complete$hct)

complete <- complete[idx1==FALSE,]

idx2 <- is.na(complete$plt)

complete <- complete[idx2==FALSE,]

summary(complete)

str(complete)


#CLASS PERFORMANCE

#IF WE PREDICTED THAT EACH REFERRAL WAS A CASE IRRESPECTIVE OF THE DATA WHAT WOULD BE #THE SUCCESS OF THE PREDICTOR - THIS GIVES A BASELINE AGAINST WHICH TO EVALUATE THE #PERFORMANCE OF A CLASSIFIER


base_performance <- 1266/2970 #NUMBER OF CASES/TOTAL NUMBER OF REFERRALS#

base_performance


#IN THIS DATASET THE BASE PERFORMANCE IS 0.4263 - A CLASSIFIER WOULD NEED TO EXCEED THIS TO BE MORE SUCCESSFUL THAN THE BASELINE#


#QQ PLOTS OF EACH VARIABLE#
```

```
qqnorm(complete$age)
qqnorm(complete$plt)
qqnorm(log(complete$plt))
qqnorm(sqrt(complete$plt))
qqnorm(complete$hb)
qqnorm(complete$wbc)
qqnorm(log(complete$wbc))
qqnorm(sqrt(complete$wbc))
qqnorm(complete$rbc)
qqnorm(complete$mcv)
qqnorm(complete$lymph)
qqnorm(log(complete$lymph))
qqnorm(complete$neutr)
qqnorm(log(complete$neutr))
qqnorm(complete$mono)
qqnorm(log(complete$mono))
qqnorm(complete$hct)
qqnorm(complete$mch)
qqnorm(log(complete$mch))
qqnorm(complete$mchc)
qqnorm(log(complete$mchc))
qqnorm(sqrt(complete$mchc))
```

#ALTER FUNCTIONAL FORM OF VARIABLES AS INDICATED BY QQ PLOTS#

```
complete$wbc <- (log(complete$wbc))
complete$plt <- (sqrt(complete$plt))
complete$hct <- (log(complete$hct))
complete$lymph <- (log(complete$lymph))
complete$neutr <- (log(complete$neutr))
complete$mono <- (log(complete$mono))
complete$mch <- (log(complete$mch))
complete$mchc <- (log(complete$mchc))

str(complete)

###### PLOTTING PAIRS ######
pairs_data <- data.frame(with(complete, cbind(age, hb, lymph, neutr, mono, plt, mchc, mch, hct, wbc)))
pairs(pairs_data)
```

10.2.2 Logistic regression model

```
#INCLUDE ALL VARIABLES IN MODEL TO BEGIN WITH#
model1 <- glm(case_control ~ age + gender + hb + rbc + hct + mcv + mch + mchc + wbc + lymph + neutr + mono + plt, family=binomial, data=complete)
summary(model1)
plot(model1)

#ASSESS COLINEARITY IN MODEL#
```

```
vif(model1)

sqrt(vif(model1))


#THERE IS MULTICOLINEARITY IN THE MODEL AS INDICATED BY THE PAIRS PLOTS#

#REMOVAL OF VARIABLES WITH LEAST STATISTICALLY SIGNIFICANT WEIGHTINGS#

#MCH, WBC, MONO AND HCT REMOVED#


model2 <- glm(case_control ~ age + gender + hb + rbc + mcv + mchc + lymph + neutr + plt, family=binomial, data=complete)

summary(model2)

vif(model2)


#LASSO REGRESSION FOR ATTRIBUTE SELECTION#


#CREATING A DATASET FOR PENALISED MODELLING#


pen_data <- data.frame(with(complete, cbind(case_control,age,gender,hb,rbc,mcv,mchc,lymph,neutr,plt)))

x <- as.matrix(pen_data[,2:10])

y <- as.matrix(pen_data[,1])


grid <- seq(0, 100000, by=10)/100000

mod.lasso <- glmnet(x, y, alpha=1, family="binomial", lambda=grid)

plot(mod.lasso, xvar="lambda", label=T, lwd=1)
```

```
grid <- seq(0, 100000, by=10)/100000
cv.out=cv.glmnet(x,y,alpha=1, lambda=grid, family="binomial", nfolds=10)
plot(cv.out)

bestlam=cv.out$lambda.min
lasso_coef <- as.matrix(coef(mod.lasso, s=bestlam))
chosen_variables_max <- rownames(lasso_coef)[lasso_coef!=0]
chosen_variables_max <- chosen_variables_max[-1]
rownames(lasso_coef)[lasso_coef!=0]

bestlam=cv.out$lambda.1se
lasso_coef <- as.matrix(coef(mod.lasso, s=bestlam))
chosen_variables_min <- rownames(lasso_coef)[lasso_coef!=0]
chosen_variables_min <- chosen_variables_min[-1]
rownames(lasso_coef)[lasso_coef!=0]


# CALCULATE RELAXED LASSO MODELS BASED ON LASSO SELECTION. NOTICE "BEST" LASSO #
# MODEL DOES NOT SHRINK AWAY ANY VARIABLES #


mod_max <- glm(as.formula(paste("case_control ~ ", paste(chosen_variables_max, collapse="+"))), data=pen_data)
summary(mod_max)
```

```
mod_min <- glm(as.formula(paste("case_control ~ ", paste(chosen_variables_min, collapse="+"))), data=pen_data)

summary(mod_min)


#MODEL PERFORMANCE USING USING THE LRM FUNCTION RATHER THAT GLM#


lrm_model <- lrm(case_control ~ age + gender + hb + rbc + mcv + mchc + lymph + neutr + plt, data=complete, x=TRUE, y=TRUE)

lrm_model


#CORRELATIONSTATISTICS SHOWN ARE: #

#C - C INDEX = AUC #

#DXY - SOMERS' D = (NC-ND)/(NC+ND) - DOES NOT INCLUDE 1/2 TIES #

#DXY CAN ALSO BE CALCULATED FROM C USING =2(C-1/2) #

#GAMMA - GOODMAN AND KRUSKAL'S GAMMA = (NC-ND)/(NC+ND) #

#TAU-A - KENDALL RANK COEFFICIENT = NC-ND/N(N-1)/2 - DOES NOT INCLUDE ADJUSTMENTS FOR TIES #


validate(lrm_model, method="cross")


#MODEL EVALUATION AND VALIDATION#


#SETTING CLASS VALUES FOR ACTUAL AND PREDICTED OUTCOMES#
```

```
actual_outcome <- as.factor(complete$case_control)

predicted_probability <- as.numeric(predict(model2, complete, type='response'))

predicted_outcome <- as.factor(ifelse(predicted_probability<='0.70', 'case','control'))



#SHOW A SIMPLE 2X2 TABLE OF THESE TWO CLASSES#

table(actual_outcome, predicted_outcome)

#CROSSTABLE FUNCTION ADDS A FEW EXTRA DETAILS TO THE TABLE (FROM GMODELS PACKAGE) #

CrossTable(actual_outcome, predicted_outcome)

#using caret package to create a confusion matrix with performance measures#

confusionMatrix(predicted_outcome, actual_outcome, positive="case")

#KAPPA STATISTIC - ADJUSTS THE ACCURACY OF THE MODEL BY TAKING INTO ACCOUNT THE POSSIBILITY OF A CORRECT PREDICTION BASED ON CHANCE ALONE #

#<0.2=POOR AGREEMENT, >0.8=VERY GOOD AGREEMENT #

#MODEL SENSITIVITY - THIS IS THE MEASURE OF TRUE POSITIVES/(TRUE POSITIVES + #

#FALSE NEGATIVES) #
```

```
rec <- sensitivity(predicted_outcome, actual_outcome, positive="case")

rec


#MODEL SPECIFICITY - THIS IS THE MEASURE OF TRUE NEGATIVES/(TRUE NEGATIVES + #

# FALSE POSITIVES) #


spec <- specificity(predicted_outcome, actual_outcome, positive="case")

spec


#MODEL PRECISION - LIKELYHOOD OF MODEL CORRECTLY CLASSIFYING A POSITIVE (ALSO CALLED #

# POSITIVE PREDICTIVE VALUE) #


prec <- posPredValue(predicted_outcome, actual_outcome, positive="case")

prec


#F MEASURE - COMBINATION OF PRECISION AND RECALL INTO A SINGLE VALUE #


f <- (2*prec*rec)/(prec+rec)

f



#VISUALISING PERFORMANCE USING ROCR#
```

```
pred <- prediction(predictions=predicted_probability, labels=actual_outcome)

perf <- performance(pred, measure = "sens", x.measure = "spec")
plot(perf)

perf.auc <- performance(pred, measure="auc")
unlist(perf.auc@y.values)

perf@alpha.values[[1]][which.max(perf@x.values[[1]]+perf@y.values[[1]])]
max(perf@x.values[[1]]+perf@y.values[[1]])

#############################################################################
#EXAMPLE SCRIPT FOR MANUAL THRESHOLD SELECTION#
#REPEATED AT MULTIPLE VALUES OF PREDICTED_PROBABILITY#

actual_outcome <- as.factor(complete$case_control)
predicted_probability <- as.numeric(predict(model4, complete, type='response'))
predicted_outcome <- as.factor(ifelse(predicted_probability<='0.05', 'case','control'))
confusionMatrix(predicted_outcome, actual_outcome, positive="case")

#############################################################################
#

#TESTING LOGISTIC REGRESSION MODEL ON NEW DATA DATASET#
```

```
#MODEL VALIDATION ON A NEW DATA SET - SCRIPT IS LARGELY REPEATED#

test_data <- read.csv("2015qmnrequests.csv", header=T) #READ IN NEWDATA#
str(test_data)


#MAKE ADJUSTMENTS TO VARIABLE TYPES AND FUNCTIONAL FORMS TO
MATCH ORIGINAL# #DATASET#

test_data$smoker <- as.factor(test_data$smoker)
test_data$thrombosis <- as.factor(test_data$thrombosis)
test_data$bp <- as.factor(test_data$bp)
test_data$dm <- as.factor(test_data$dm)
test_data$treatment <- as.factor(test_data$treatment)
test_data$dead <- as.factor(test_data$dead)
test_data$status <- as.factor(test_data$status)
test_data$cause.of.death <- as.factor(test_data$cause.of.death)
test_data$time <- as.factor(test_data$time)

str(test_data)

idx1 <- is.na(test_data$hct)
complete_test <- test_data[idx1==FALSE,]
idx2 <- is.na(complete_test$plt)
```

```
complete_test <- complete_test[idx2==FALSE,]
summary(complete_test)

complete_test$wbc <- (log(complete_test$wbc))
complete_test$plt <- (sqrt(complete_test$plt))
complete_test$hct <- (log(complete_test$hct))
complete_test$lymph <- (log(complete_test$lymph))
complete_test$neutr <- (log(complete_test$neutr))
complete_test$mono <- (log(complete_test$mono))
complete_test$mch <- (log(complete_test$mch))
complete_test$mchc <- (log(complete_test$mchc))

#MODEL EVALUATION AND VALIDATION#

#SETTING CLASS VALUES FOR ACTUAL AND PREDICTED OUTCOMES#
test_outcome <- as.factor(complete_test$case_control)
test_probability <- as.numeric(predict(model2, complete_test, type='response'))
test_predicted_outcome <- as.factor(ifelse(test_probability <='0.70', 'case','control'))

print(test_probability)
print(test_predicted_outcome)
print(test_outcome)

#SHOW A SIMPLE 2X2 TABLE OF THESE TWO CLASSES#
```

```
table(test_outcome, test_predicted_outcome)



#CROSSTABLE FUNCTION ADDS A FEW EXTRA DETAILS TO THE TABLE (FROM GMODELS PACKAGE) #

CrossTable(test_outcome, test_predicted_outcome)

#USING CARET PACKAGE TO CREATE A CONFUSION MATRIX WITH PERFORMANCE MEASURES#

confusionMatrix(test_predicted_outcome, test_outcome, positive="case")

rec <- sensitivity(test_predicted_outcome, test_outcome, positive="case")
rec

spec <- specificity(test_predicted_outcome, test_outcome, positive="case")
spec

prec <- posPredValue(test_predicted_outcome, test_outcome, positive="case")
prec

negprec <- negPredValue(test_predicted_outcome, test_outcome, positive="case")
negprec

f <- (2*prec*rec)/(prec+rec)
```

```
f

#VISUALISING PERFORMANCE USING ROCR#


pred2 <- prediction(predictions = test_probability, labels=test_outcome)

perf2 <- performance(pred2, measure = "tpr", x.measure = "fpr")
plot(perf2)

perf.auc <- performance(pred, measure="auc")
unlist(perf.auc@y.values)


#EXTRACTING MISMATCHES FROM LOGISTIC REGRESSION FOR
INVESTIGATION#

#MISMATCHES FROM TRAINING SET#

complete$predicted <- predicted_outcome
lrm_mismatches <- complete[complete$case_control!=complete$predicted,]
lrm_mismatches


#WRITE DATA TO FILE#
```

```
write.csv(lrm_mismatches, "lrm_mismatches.csv")


#MISMATCHES IN TEST SET#


complete_test$probability <- test_probability

complete_test$predicted <- test_predicted_outcome

lrm_test_mismatches <-
complete_test[complete_test$case_control!=complete_test$predicted,]


#WRITE DATA TO FILE#

write.csv(lrm_test_mismatches, "lrm_test_mismatches2.csv")
```

10.2.3 Random forest analysis

```
#RANDOMFOREST CLASSIFICATION#

#TO BE RUN IN CONJUNCTION WITH LOGISTIC REGRESSION SCRIPT AS
VARIABLES AND DATASETS #HAVE BEEN CREATED/MODIFIED USING SCRIPT
FROM PREVIOUS ANALYSIS


library(randomForest)
```

```
#SELECT DATA FOR RANDOM FOREST ANALYSIS - REMOVING ALL CASES WITH NA VALUES IN #

#BOTH TRAINING AND TEST SETS - REQUIRES TEST_DATA FROM LOGISTIC REGRESSION SCRIPT #


rf_data <- complete[,c("request_number", "year","case_control","age","hb", "rbc","wbc","lymph","plt","hct", "gender", "mchc", "mch", "neutr", "mono", "mcv")]

rfidx1 <- is.na(rf_data$lymph)

rf_data <- rf_data[rfidx1==FALSE,]

rfidx2 <- is.na(rf_data$neutr)

rf_data <- rf_data[rfidx2==FALSE,]

rfidx3 <- is.na(rf_data$mono)

rf_data <- rf_data[rfidx3==FALSE,]

summary(rf_data)


rf_test <- test_data[,c("request_number", "year", "source_type" , "case_control","age","hb", "rbc","wbc","lymph","plt","hct", "mcv", "gender", "mchc", "mch", "neutr", "mono")]

rftidx1 <- is.na(rf_test$lymph)

rf_test <- rf_test[rftidx1==FALSE,]

rftidx2 <- is.na(rf_test$neutr)

rf_test <- rf_test[rftidx2==FALSE,]

rftidx3 <- is.na(rf_test$mono)

rf_test <- rf_test[rftidx3==FALSE,]

summary(rf_test)
```

```
#RANDOM FOREST MODEL INCLUDES ALL EXPLANATORY VARIABLES#

rf_model<-train(case_control ~ age + gender + hb + rbc + wbc + neutr + lymph + mono + plt
+ hct + mchc + mch + mcv, data=rf_data, method="rf", na.action=na.exclude,
                trControl=trainControl(method="cv", number=10),
                prox=TRUE,allowParallel=TRUE)



print(rf_model)

print(rf_model$finalModel)

saveRDS(rf_model, "my-fitted-rf.rds")

fit <- readRDS("my-fitted-rf.rds")

fit <- rf_model

#USING RANDOM FOREST MODEL TO PREDICT OUTCOME ON BOTH TEST SET
AND TRAINING SET#

predict(fit, rf_data)
predict(fit, rf_test)

rf_test_outcome <- as.factor(rf_data$case_control)
rf_predicted_outcome <- (predict(rf_model,rf_data))
```

```
summary(rf_test_outcome)

summary(rf_predicted_outcome)


rf_test_outcome2 <- as.factor(rf_test$case_control)

rf_predicted_outcome2 <- (predict(fit,rf_test))

CrossTable(rf_test_outcome, rf_predicted_outcome)

table(rf_test_outcome, rf_predicted_outcome)


CrossTable(rf_test_outcome2, rf_predicted_outcome2)

table(rf_test_outcome2, rf_predicted_outcome2)

ConfusionMatrix(rf_test_outcome2, rf_predicted_outcome2)


#PLOT THE ACCURACY OF THE CROSS VALIDATION WITH INCREASING
NUMBERS OF #

#RANDOMLY SELECT VARIABLES#


plot(rf_model)


#EXTRACTING THE MISMATCHES FROM THE MODEL FOR FURTHER
INVESTIGATION#


rf_data$predicted <- predict(fit, rf_data)

rf_test$predicted <- predict(fit, rf_test)


rf_mismatches <- rf_data[rf_data$case_control!=rf_data$predicted,]
```

```
write.csv(rf_mismatches, "random_forest_mismatches.csv")


rftest_mismatches <- rf_test[rf_test$case_control!=rf_test$predicted,]
write.csv(rftest_mismatches, "random_forest_test_mismatches.csv")
```

10.2.4 Gradient boosted model

```
#GRADIENT BOOSTING#
#USING CARET AND GBM PACKAGES#


#CREATE DATASET FOR ANALYSIS - USING DATASETS DEFINED IN THE
LOGISTIC REGRESSION ANALYSIS SCRIPT#


idxc1 <- is.na(complete$neutr)
gbm_data <- complete[idxc1==FALSE,]
idxc2 <- is.na(gbm_data$lymph)
gbm_data <- gbm_data[idxc2==FALSE,]
idxc3 <- is.na(gbm_data$mono)
gbm_data <- gbm_data[idxc3==FALSE,]


fitControl <- trainControl(
  method = "cv",
  number = 10)


fitControl <- trainControl(
  method = "repeatedcv",
```

```
  number = 10,
  repeats = 10)



gbmGrid <-  expand.grid(interaction.depth = c(1, 3, 5),
                        n.trees = (1:30)*50,
                        shrinkage = c(0.1, 0.01, 0.001),
                        n.minobsinnode = 20
)



gbmmodel<-train(case_control ~ age + gender + hb + mcv + mchc + neutr + lymph + plt,
                data=gbm_data,
                method="gbm",
                distribution="bernoulli",
                trControl=fitControl,
                tuneGrid=gbmGrid,
                verbose=FALSE)



gbm_fit <- gbmmodel
gbm_fit
summary(gbm_fit, las=2, cex.names=0.7)
ggplot(gbm_fit)
```

```
trellis.par.set(caretTheme())
plot(gbm_fit)

#CREATE DATASETS WITH BOTH ACTUAL AND PREDICTED OUTCOMES#

gbm_outcome <- as.factor(gbm_data$case_control)
gbm_predicted_outcome <- (predict(gbm_fit,gbm_data))

#DISPLAY ACTUAL AND PREDICTED OUTCOMES AS MATRIX#

CrossTable(gbm_outcome, gbm_predicted_outcome)
table(gbm_outcome, gbm_predicted_outcome)
confusionMatrix(gbm_outcome, gbm_predicted_outcome, positive='case')

#PLOT MODEL#
plot(gbmmodel)

#SUMMARY OF PERFORMANCE#

gbmrec <- sensitivity(gbm_predicted_outcome, gbm_test_outcome, positive="case")
gbmrec

#MODEL SPECIFICITY - THIS IS THE MEASURE OF TRUE NEGATIVES/(TRUE
NEGATIVES +#
```

```
#FALSE POSITIVES) #

gbmspec <- specificity(gbm_predicted_outcome, gbm_test_outcome, positive="case")
gbmspec

#MODEL PRECISION - LIKELYHOOD OF MODEL CORRECTLY CLASSIFYING A POSITIVE #
# (ALSO CALLED POSITIVE PREDICTIVE VALUE) #

gbmprec <- posPredValue(gbm_predicted_outcome, gbm_test_outcome, positive="case")
gbmprec

gbmnegprec <- negPredValue(gbm_predicted_outcome, gbm_test_outcome, positive="case")
gbmnegprec

#F MEASURE - COMBINATION OF PRECISION AND RECALL INTO A SINGLE VALUE#

gbmf <- (2*gbmprec*gbmrec)/(gbmprec+gbmrec)
gbmf



#READ IN MODEL VALIDATION DATA#

gmb_2015 <- read.csv("2015qmnrequests.csv", header=T)
```

```
#ALTER FUNCTIONAL FORMS OF EACH VARIABLE AS PER TEST MODEL#

gmb_2015$wbc <- (log(gmb_2015$wbc))

gmb_2015$plt <- (sqrt(gmb_2015$plt))

gmb_2015$hct <- (log(gmb_2015$hct))

gmb_2015$lymph <- (log(gmb_2015$lymph))

gmb_2015$neutr <- (log(gmb_2015$neutr))

gmb_2015$mono <- (log(gmb_2015$mixed))

idxc1 <- is.na(gmb_2015$neutr)

gmb_2015a <- gmb_2015[idxc1==FALSE,]

idxc2 <- is.na(gmb_2015a$lymph)

gmb_2015a <- gmb_2015a[idxc2==FALSE,]

idxc3 <- is.na(gmb_2015a$mono)

gmb_2015a <- gmb_2015a[idxc3==FALSE,]

#PREDICT FIT OF MODEL ON VALIDATION DATASET#

gbm_test <- predict(gbm_fit, newdata=gmb_2015a)

gbm_test_outcome <- as.factor(gmb_2015a$case_control)

gbm__test_predicted_outcome <- (predict(gbm_fit, gmb_2015a))

#VIEW ACTUAL AND PREDICTED OUTCOMES IN MATRIX#
```

```
CrossTable(gbm_test_outcome, gbm__test_predicted_outcome)

table(gbm_test_outcome, gbm__test_predicted_outcome)

confusionMatrix(gbm_test_outcome, gbm__test_predicted_outcome, positive='case')


#WRITE MISMATCHED DATA TO FILE FOR INVESTIGATIONS#


gmb_2015a$predicted <- gbm__test_predicted_outcome

gbm_test_mismatches <- gmb_2015a[gmb_2015a$case_control!=gmb_2015a$predicted,]

write.csv(gbm_test_mismatches, "gbm_test_mismatches.csv")
```

10.2.5 Logistic regression model impact assessment

```
#Model impact on 2016 data - script is largely repeated to evaluate model on #

#new dataset #

##############################################################################

#READ IN NEW DATA#


test16_data <- read.csv("2016_qmpns.csv", header=T)

str(test16_data)


#MAKE ADJUSTMENTS TO VARIABLE TYPES AND FUNCTIONAL FORMS TO
MATCH ORIGINAL# #DATASET#
```

```
test_data$smoker <- as.factor(test_data$smoker)
test_data$thrombosis <- as.factor(test_data$thrombosis)
test_data$bp <- as.factor(test_data$bp)
test_data$dm <- as.factor(test_data$dm)
test_data$treatment <- as.factor(test_data$treatment)
test_data$dead <- as.factor(test_data$dead)
test_data$status <- as.factor(test_data$status)
test_data$cause.of.death <- as.factor(test_data$cause.of.death)
test_data$time <- as.factor(test_data$time)

str(test_data)

#REMOVE INCOMPLETE RECORDS#

idx1 <- is.na(test16_data$hct)
complete_test16 <- test16_data[idx1==FALSE,]
idx2 <- is.na(complete_test16$plt)
complete_test16 <- complete_test16[idx2==FALSE,]
summary(complete_test)
idx3 <- is.na(complete_test16$lymph)
complete_test16 <- complete_test16[idx3==FALSE,]
idx4 <- is.na(complete_test16$neutr)
summary(complete_test16)
complete_test16 <- complete_test16[idx4==FALSE,]
```

```
complete_test16$wbc <- (log(complete_test16$wbc))
complete_test16$plt <- (sqrt(complete_test16$plt))
complete_test16$hct <- (log(complete_test16$hct))
complete_test16$lymph <- (log(complete_test16$lymph))
complete_test16$neutr <- (log(complete_test16$neutr))
complete_test16$mixed <- (log(complete_test16$mixed))

str(complete_test16)
complete_test16$mcv <- as.numeric(complete_test16$mcv)

#MODEL EVALUATION AND VALIDATION#

#SETTING CLASS VALUES FOR ACTUAL AND PREDICTED OUTCOMES#

test16_outcome <- as.factor(complete_test16$case_control)
test16_probability <- as.numeric(predict(model2, complete_test16, type='response'))
test16_predicted_outcome <- as.factor(ifelse(test16_probability <= 0.7, 'case','control'))

test16_probability

#SHOW A SIMPLE 2X2 TABLE OF THESE TWO CLASSES#

table(test16_outcome, test16_predicted_outcome)
```

```
#CROSSTABLE FUNCTION ADDS A FEW EXTRA DETAILS TO THE TABLE (FROM GMODELS PACKAGE) #

CrossTable(test16_outcome, test16_predicted_outcome)

#USING CARET PACKAGE TO CREATE A CONFUSION MATRIX WITH PERFORMANCE MEASURES#

confusionMatrix(test16_predicted_outcome, test16_outcome, positive="case")

rec16 <- sensitivity(test16_predicted_outcome, test16_outcome, positive="case")
rec16

spec16 <- specificity(test16_predicted_outcome, test16_outcome, positive="case")
spec16

prec16 <- posPredValue(test16_predicted_outcome, test16_outcome, positive="case")
prec16

negprec16 <- negPredValue(test16_predicted_outcome, test16_outcome, positive="case")
negprec16

f16 <- (2*prec16*rec16)/(prec16+rec16)
f16

#VISUALISING PERFORMANCE USING ROCR#
```

```
pred2 <- prediction(predictions = test_probability, labels=test_outcome)

perf2 <- performance(pred2, measure = "tpr", x.measure = "fpr")
plot(perf2)

perf.auc <- performance(pred, measure="auc")
unlist(perf.auc@y.values)

#EXTRACTING MISMATCHES FROM LOGISTIC REGRESSION FOR INVESTIGATION#

#MISMATCHES IN TEST16 SET#
complete_test16$predicted <- test16_predicted_outcome
lrm_test16_mismatches <- complete_test16[complete_test16$case_control!=complete_test16$predicted,]
write.csv(lrm_test16_mismatches, "lrm_test16_mismatches.csv")

#SUMMARY DATA FOR CHAPTER 7#
summary(test16_data)
summary(complete_test16)
summary(test16_predicted_outcome=='control')
```

10.3 Cost Burden Analysis

	Cost Per Test
Sample receipt (registration, FBC, screening)	
Combined Reagents, Instruments and Consumables	0.93
Staff	3.62
DNA Extraction	
Combined Reagents, Instruments and Consumables	2.35
Staff	1.32
PCR	
Combined Reagents, Instruments and Consumables	5.09
Staff	5.28
Fragment Analysis	
Combined Reagents, Instruments and Consumables	3.16
Staff	0.92
Data transfer, analysis and reporting	
Staff	14.44
Total Cost per Test	37.12

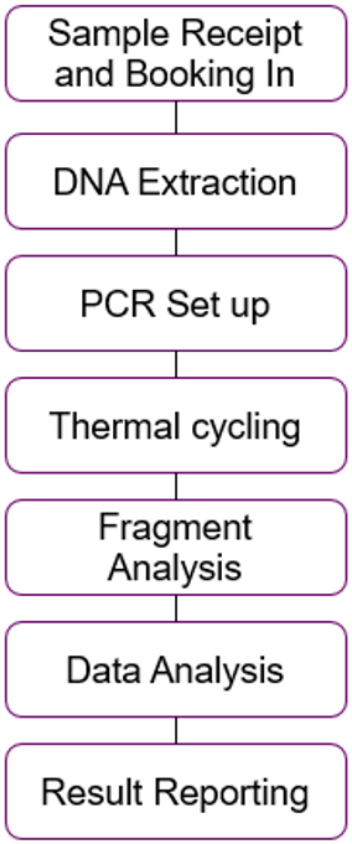

Figure 10-1 Summary of Cost Burden Analysis for Fluorescent PCR detection of *JAK2* V617F and *CALR* mutations.

Reagents	Total cost	Denominator	Cost per test	Comments
Stomatalyser	131.47	1042	0.13	use 24 packs per year, performing 25000 counts in total
Cell Pack	21.06	168.92	0.12	use 12 per month, performing 25000 counts per year
Cell Clean	83.74	6250	0.01	use 4 per annum
Paper Rolls	45.60	96.15	0.47	13 boxes per annum
QC material	259.16	6250	0.04	4 packs per yer
Total			**0.78**	

Instruments and Consumables	Total cost	Denominator	Cost per test	Comments
Sysmex XP-300	2614.00	25000	0.10	maintainance contract per annum assuming 25000 counts performed annually
Glass slides	33.15	833.33	0.04	2500 slides per box, 40 boxes per year, 3 per sample
Microscope?	130.60	18000	0.01	maintainance per microscope, assessing 18000 films per annum
Total			**0.15**	

Staff	Band	Cost per Min	Time	Cost per test	Comments
Sample processing	2	0.20	10	2.00	Total time to perform registration, slide preparation and full blood count per sample
Morphological Assessment	8	0.54	3	1.62	Morphological Assessment plus test selection
Total				**3.62**	

Total Costs	4.55

Figure 10-2. Cost Burden Analysis (sample receipt) - detailed breakdown of costings for sample receipt, full blood count analysis, morphological preparation, registration and screening.

Reagents	Total Cost	Denominator	Cost per test	Comments
Ammonium Chloride	11.49	5000	0.00	
Qiagen DNA extraction kit	504.14	250	2.02	475.60 per kit, each doing 250 extractions
Total			**2.02**	

Instruments and Consumables	Total Cost	Denominator	Cost per test	Comments
QIAcube	1099.22	8000	0.14	£2228.40/year maintenance, 3120 samples per year. Mawell FOC.
Centrifuge	147.80	5200	0.03	Assuming 100 samples per week processed
Pasteur Pippette	4.23	100	0.04	Assuming 5 pasteur pippettes used during process
15mL tube	38.94	500	0.08	1 per sample
1.5 ml tubes	44.52	1000	0.04	1 tube for DNA
Total			**0.33**	

Staff	Band	Cost per Min	Time	Cost per test	Comments
Extraction process	3	0.22	3	0.66	30 min hands on time for 10 samples
DNA quantification and storage	3	0.22	3	0.66	30 min hands on time for 10 samples
				0.00	
Total				**1.32**	

Total Costs	3.67

Figure 10-3. Cost Burden Analysis (DNA extraction) - detailed breakdown of costings for DNA extraction process.

Reagents	Total Cost	Denominator	Cost per test	Comments
AmpliTaq Gold Kit	1443.40	3000	0.48	
Molecular grade water	12.01	50	0.24	
Non-fluorescent primers	6.00	3000	0.00	
Fluorescent primers	65.00	3000	0.02	
dNTPs	135.40	3000	0.05	
Total			**0.79**	

Instruments and consumables	Total Cost	Denominator	Cost per test	Comments
0.2 ml PCR tubes	44.52	500	0.09	1 tube per PCR
200uL tips stacked	27.98	325	0.09	6 per sample
10uL tips filtered	87.88	160	0.55	6 per sample
40 uL filtered	29.68	160	0.19	6 per sample
200 uL filtered	100.70	160	0.63	6 per sample
plate	82.68	1600	0.07	using an average of 64 wells of each microtitre plate
Thermocycler (including service contract)	370.00	2500	0.15	£370 per cycler calibration and maintainance, each running 5000 PCR reactions per annum
Total			**1.76**	

Staff	Band	Cost per Min	Time	Cost per test	Comments
PCR set up	7	0.44	6	2.64	30 min hands on time for 10 samples
Total				**2.64**	

Total Costs	5.19

Figure 10-4. Cost Burden Analysis (PCR) - detailed breakdown of costings for PCR amplification process.

Reagents		Total Cost	Denominator	Cost per test	Comments
Formamide		21.98	100	0.219844	£20.74 per 25 ml. 5ml per week
Size standard		296.32	800	0.37040375	ROX500
10x buffer		90.63	500	0.18126	25 ml per bottle, 5 ml per week, 100 samples per week
POP7 for 3130		332.56	500	0.6651288	lasts 1 week
Molecular grade water for maintenance		53.00	1000	0.053	£50 per litre, 100 ml per week, 100 samples per week
Total				**1.48963655**	
Instruments and Consumables		**Total Cost**	**Denominator**	**Cost per test**	**Comments**
3130 plates		3.31	30	0.11024	£3.12 per plate, running an average of 30 samples per batch.
Capilary electrophoresis instrument		8090	5200	1.55576923	£8090 per annum running 100 samples per week
Total				**1.66600923**	
Staff	**Band**	**Cost per Min**	**Time**	**Cost per test**	**Comments**
Maintenance	7	0.44	0.6	0.264	1 hour per week, 100 samples per week, 0.6 minute per sample
Set-Up	7	0.44	1.5	0.66	40 sample run, 60 minutes
Staff 1				0	
Total				**0.924**	
Total Costs				4.07964578	

Figure 10-5. Cost Burden Analysis (Fragment Analysis) - detailed breakdown of costings for fluorescent fragment analysis of PCR products.

Staff	Band	Cost per Min	Time	Cost per test	Comments
First analysis	7	0.44	10	4.40	
Check	7	0.44	5	2.20	
Reporting results	7	0.44	8	3.52	
Authorisation	8a	0.54	8	4.32	
Total				**14.44**	

Total Costs	14.44

Figure 10-6. Cost Burden Analysis (Data transfer, analysis and reporting) - detailed breakdown of costings for the interpretation and reporting of *JAK2* V617F and *CALR* mutation analysis.

ABBREVIATIONS

A	ALANINE
ABL	ABELSON
AML	ACUTE MYELOID LEUKAEMIA
ASO	ALLELE SPECIFIC OLIGONUCLEOTIDE
ATP	ADENOSINE TRIPHOSPHATE
AUROC	AREA UNDER RECEIVER OPERATOR CURVE
B/NK	B/NK CELL PROGENITOR
BCR	BREAKPOINT CLUSTER REGION
BCSH	BRITISH COMMITTEE FOR STANDARDS IN HAEMATOLOGY
BLAST	BASIC LOCAL ALIGNMENT SEARCH TOOL
BM	BONE MARROW
BP	BASE PAIRS
BSA	BOVINE SERUM ALBUMIN
C/EBPα	CCAAT/ENHANCER-BINDING PROTEIN ALPHA
CAD	COMPUTER AIDED DESIGN
CALR	CALRETICULIN
CD	CLUSTER OF DIFFERENTIATION
CGL	CHRONIC GRANULOCYTIC LEUKAEMIA
CMP	COMMON MYELOID PROGENITOR
DNA	DEOXYRIBONUCLEIC ACID
DOH	DEPARTMENT OF HEALTH
ECMP	EUROPEAN CLINIAL AND MOLECULAR PATHOLOGY
EDTA	ETHYLENEDIAMINETETRAACETIC ACID
EPO	ERYTHROPOIETIN
EPOR	ERYTHROPOIETIN RECEPTOR
ER	ENDOPLASMIC RETICULUM
ESR	ERYTHROCYTE SEDIMENTATION RATE
ET	ESSENTIAL THROMBOCYTHAEMIA
ETP	EARLY THYMIC PROGENITOR
F	PHENYLALANINE
F	FEMALE

FAB	FRENCH-AMERICAN-BRITISH
FAM	FAM FLUOROPHORE
FBC	FULL BLOOD COUNT
FN	FALSE NEGATIVE
FP	FALSE POSITIVE
GATA-1	ERYTHROID TRANSCRIPTION FACTOR (GATA-BINDING FACTOR 1)
GMP	GRANULOCYTE MACROPHAGE PROGENITOR
GRADE	GRADING OF RECOMMENDATIONS, ASSESSMENT, DEVELOPMENT AND EVALUATION
GWAS	GENOME WIDE ASSOCIATION STUDIES
HB	HAEMOGLOBIN
HCT	HAEMATOCRIT
HEX	HEX FLUOROPHORE
HILIS	HMDS INTERGRATED LABORATORY INFORMATION SYSTEM
HMDS	HAEMATOLOGICAL MALIGNANCY DIAGNOSTIC SERVICE
HMRN	HAEMATOLOGICAL MALIGNANCY RESEARCH NETWORK
HSC	HAEMATOPOIETIC STEM CELL
ICD	INTERNATIONAL CLASSIFICATION OF DISEASE
ICD-O	INTERNATIONAL CLASSIFICATION OF DISEASE – ONCOLOGY
IL-6	INTERLEUKIN-6
ITP	IDIOPATHIC THROMBOCYOPENIA PURPURA
JAK2	JANUS KINASE 2
K	LYSINE
KDEL	LYSINE, ASPARTIC ACID, GLUTAMIC, LEUCINE
L	LEUCINE
LASSO	LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR
LYMPH	LYMPHOCYTE
M	MALE
MCH	MEAN CELL HAEMOGLOBIN
MCHC	MEAN CORPUSCULAR HAEMOGLOBIN CONCENTRATION

MCV	MEAN CELL VOLUME
MDS	MYELODYSPLASTIC SYNDROME
MEP	MEGAKARYOCYTIC ERYTHROID PROGENITOR
MEP	MEGAKARYOCYTIC/ERYTHROID PROGENITOR
MLP	MULTI-LYMPHOID PROGENITOR
MONO	MONOCYTE
MPD	MYELOPROLIFERATIVE DISORDER
MPL	MYELOPROLIFERATIVE LEUKAEMIA PROTEIN
MPN	MYELOPROLIFERATIVE NEOPLASM
MPP	MULTI-POTENT STEM CELL
MPS	MYELOPROLIFERATIVE SYNDROME
MPV	MEDICAL PRACTICE VARIATION
NCBI	NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
NEUTR	NEUTROPHIL
NH4CL	AMMONIUM CHLORIDE
NHS	NATIONAL HEALTH SERVICE
NICE	NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE
PCR	POLYMERASE CHAIN REACTION
PNH	PAROXYSMAL NOCTURNAL HAEMOGLOBINURIA
PI3K	PHOSPHOINOSITIDE 3-KINASE
PLT	PLATELET
PMF	PRIMARY MYELOFIBROSIS
PV	POLYCYTHAEMIA VERA
PVSG	POLYCYTHAEMIA VERA STUDY GROUP
R	ARGININE
RBC	RED BLOOD CELL
ROC	RECEIVER OPERATOR CURVE
RCF	RELATIVE CENTRIFUGAL FORCE
SNP	SINGLE NUCLEOTIDE POLYMORPHISM
SQL	STRUCTURED QUERY LANGUAGE
STAT5	SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 5
TM	MELTING TEMPERATURE

TN	TRUE NEGATIVE
TP	TRUE POSITIVE
TPO	THROMBOPOIETIN
TVSG	THROMBOCYTHAEMIA VERA STUDY GROUP
UCSC	UNIVERISTY OF CALIFORNIA, SANTA CRUZ
UKAS	UNITED KINGDOM ACCREDITATION SERVICE
V	VALINE
W	TRYPTOPHAN
WBC	WHITE BLOOD CELL
WHO	WORLD HEALTH ORGANISATION

References

Abdel-Wahab, O., T. Manshouri, J. Patel, K. Harris, J. Yao, and C. Hedvat, 2010, Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias, v. 70, p. 447-452.

Adler, D., and D. Murdoch, 2015, rgl: 3D Visualization Using OpenGL.

Araki, M., Y. Yang, N. Masubuchi, Y. Hironaka, H. Takei, S. Morishita, Y. Mizukami, S. Kan, S. Shirane, Y. Edahiro, Y. Sunami, A. Ohsaka, and N. Komatsu, 2016, Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms: Blood, v. 127, p. 1307-1316.

Arber, D. A., A. Orazi, R. Hasserjian, J. Thiele, M. J. Borowitz, M. M. L. Beau, C. D. Bloomfield, M. Cazzola, and J. W. Vardiman, 2016, The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.

Barbui, T., J. Thiele, A. Carobbio, P. Guglielmelli, A. Rambaldi, A. M. Vannucchi, and A. Tefferi, 2014a, Discriminating between essential thrombocythemia and masked polycythemia vera in JAK2 mutated patients: Am J Hematol, v. 89, p. 588-90.

Barbui, T., J. Thiele, H. Gisslinger, G. Finazzi, A. Carobbio, and E. Rumi, 2014b, Masked polycythemia vera (mPV): results of an international study: Am J Hematol, v. 89, p. 52-54.

Barbui, T., J. Thiele, H. Gisslinger, H. M. Kvasnicka, A. M. Vannucchi, P. Guglielmelli, A. Orazi, and A. Tefferi, 2018, The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion: Blood Cancer J, v. 8, p. 15.

Barbui, T., J. Thiele, F. Passamonti, E. Rumi, E. Boveri, M. Ruggeri, F. Rodeghiero, E. S. G. d'Amore, M. L. Randi, I. Bertozzi, F. Marino, A. M. Vannucchi, E. Antonioli, V. Carrai, H. Gisslinger, V. Buxhofer-Ausch, L. Müllauer, A. Carobbio, A. Gianatti, N. Gangat, C. A. Hanson, and A. Tefferi, 2011a, Survival and Disease Progression in Essential Thrombocythemia Are Significantly Influenced by Accurate Morphologic Diagnosis: An International Study.

Barbui, T., J. Thiele, A. M. Vannucchi, and A. Tefferi, 2015, Rationale for revision and proposed changes of the WHO diagnostic criteria for polycythemia vera, essential thrombocythemia and primary myelofibrosis: Blood Cancer Journal, v. 5.

Baxter, E. J., L. M. Scott, P. J. Campbell, C. East, N. Fourouclas, and S. Swanton, 2005, Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders, v. 365, p. 1054-1061.

Beer, P. A., P. J. Campbell, L. M. Scott, A. J. Bench, W. N. Erber, and D. Bareford, 2008, MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort, v. 112, p. 141-149.

Bejar, R., K. Stevenson, O. Abdel-Wahab, N. Galili, B. Nilsson, G. Garcia-Manero, H. Kantarjian, A. Raza, R. L. Levine, D. Neuberg, and B. L. Ebert, 2011, Clinical Effect of Point Mutations in Myelodysplastic Syndromes: New England Journal of Medicine, v. 364, p. 2496-2506.

Biswas, M., P. K. Prakash, M. Cossburn, K. Myers, and F. Hanna, 2003, Life-threatening thrombotic complications of relative polycythaemia: J Intern Med, v. 253, p. 481-3.

Bose, P., and S. Verstovsek, 2016, The evolution and clinical relevance of prognostic classification systems in myelofibrosis: Cancer, v. 122, p. 681-92.

BSI, 30 November 2012, Medical laboratories. Requirements for quality and competence, BSI.

Burmester, H., E. M. Wolber, P. Freitag, J. Fandrey, and W. Jelkmann, 2005, Thrombopoietin production in wild-type and interleukin-6 knockout mice with acute inflammation: J Interferon Cytokine Res, v. 25, p. 407-13.

Butcher, C. M., U. Hahn, L. B. To, J. Gecz, E. J. Wilkins, and H. S. Scott, 2008, Two novel JAK2 exon 12 mutations in JAK2V617F-negative polycythaemia vera patients, v. 22, p. 870-873.

Byun, J. M., Y. J. Kim, T. Youk, J. J. Yang, J. Yoo, and T. S. Park, 2017, Real world epidemiology of myeloproliferative neoplasms: a population based study in Korea 2004–2013: Annals of Hematology, v. 96, p. 373-381.

Capitani, M., and M. Sallese, 2009, The KDEL receptor: New functions for an old protein: The Golgi Complex, v. 583, p. 3863-3871.

Chachoua, I., C. Pecquet, M. El-Khoury, H. Nivarthi, R.-I. Albu, C. Marty, V. Gryshkova, J.-P. Defour, G. Vertenoeil, A. Ngo, A. Koay, H. Raslova, P. J. Courtoy, M. L. Choong, I. Plo, W. Vainchenker, R. Kralovics, and S. N. Constantinescu, 2016, Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants: Blood, v. 127, p. 1325-1335.

Chaligne, R., C. James, C. Tonetti, R. Besancenot, J. P. Le Couedic, and F. Fava, 2007, Evidence for MPL W515L/K mutations in hematopoietic stem cells in primitive myelofibrosis, v. 110, p. 3735-3743.

Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson, 2018, shiny: Web Application Framework for R. R package version 1.1.0.

Chen, C. C., J. P. Gau, H. J. Chou, J. Y. You, C. E. Huang, Y. Y. Chen, J. Lung, Y. S. Chou, Y. W. Leu, C. H. Lu, K. D. Lee, and Y. H. Tsai, 2014, Frequencies, clinical characteristics, and outcome of somatic CALR mutations in JAK2-unmutated essential thrombocythemia: Ann Hematol.

Chen, E., P. A. Beer, A. L. Godfrey, C. A. Ortmann, J. Li, A. P. Costa-Pereira, C. E. Ingle, E. T. Dermitzakis, P. J. Campbell, and A. R. Green, 2010, Distinct clinical phenotypes associated with JAK2V617F reflect differential STAT1 signaling: Cancer Cell, v. 18, p. 524-35.

Chi, J., K. A. Nicolaou, V. Nicolaidou, L. Kouma, A. Mitsidou, and C. Pierides, 2013, Calreticulin exon 9 frameshift mutations in patients with thrombocytosis.

Cohen, J., 1960, A Coefficient of Agreement for Nominal Scales: Educational and Psychological Measurement, v. 20, p. 37-46.

Cross, N. C. P., 2011, Genetic and Epigenetic Complexity in Myeloproliferative Neoplasms.

Dameshek, W., 1951, Some speculations on the myeloproliferative syndromes: Blood, v. 6, p. 372-5.

Delhommeau, F., S. Dupont, V. Della Valle, C. James, S. Trannoy, and A. Masse, 2009, Mutation in TET2 in myeloid cancers, v. 360, p. 2289-2301.

Doulatov, S., F. Notta, K. Eppert, L. T. Nguyen, P. S. Ohashi, and J. E. Dick, 2010, Revised map of the human progenitor hierarchy shows the origin of

macrophages and dendritic cells in early lymphoid development: Nat Immunol, v. 11, p. 585-93.
Doulatov, S., F. Notta, E. Laurenti, and J. E. Dick, 2012, Hematopoiesis: a human perspective: Cell Stem Cell, v. 10, p. 120-36.
England, P. H., 2016, NHS Cancer Screening Programmes.
Faraway, J., 2014, faraway: Functions and datasets for books by Julian Faraway. .
Fawcett, T., 2006, An introduction to ROC analysis: Pattern Recogn. Lett., v. 27, p. 861-874.
Frank, E., M. A. Hall, and I. H. Witten, 2016, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann.
Fujiwara, H., 2018, Histological evaluation of myeloproliferative neoplasms: J Clin Exp Hematop, v. 58, p. 45-50.
Genovese, G., A. K. Kahler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landen, M. Hoglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Gronberg, C. M. Hultman, and S. A. McCarroll, 2014, Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence: N Engl J Med, v. 371, p. 2477-87.
Giovanni Barosi , V. R., Elisa Bonetti, Rita Campanelli, Adriana Carolei, Paolo Catarsi, Antonina M. Isgrò, Letizia Lupo, Margherita Massa, Valentina Poletto, Gianluca Viarengo, Laura Villani, Umberto Magrini, 2012, Evidence that Prefibrotic Myelofibrosis Is Aligned along a Clinical and Biological Continuum Featuring Primary Myelofibrosis.
Gisslinger, H., G. Jeryczynski, B. Gisslinger, A. Wolfler, S. Burgstaller, V. Buxhofer-Ausch, M. Schalling, M. T. Krauth, A. I. Schiefer, C. Kornauth, I. Simonitsch-Klupp, C. Beham-Schmid, L. Mullauer, and J. Thiele, 2016, Clinical impact of bone marrow morphology for the diagnosis of essential thrombocythemia: comparison between the BCSH and the WHO criteria: Leukemia, v. 30, p. 1126-32.
Gocek, E., A. N. Moulas, and G. P. Studzinski, 2014, Non-receptor protein tyrosine kinases signaling pathways in normal and cancer cells: Critical Reviews in Clinical Laboratory Sciences, v. 51, p. 125-137.
Godfrey, A. L., E. Chen, F. Pagano, C. A. Ortmann, Y. Silber, B. Bellosillo, P. Guglielmelli, C. N. Harrison, J. T. Reilly, F. Stegelmann, F. Bijou, E. Lippert, M. F. McMullin, J.-M. Boiron, K. Döhner, A. M. Vannucchi, C. Besses, P. J. Campbell, and A. R. Green, 2012, JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone: Blood, v. 120, p. 2704.
GRADE, 2016, The Grading of Recommendations Assessment, Development and Evaluation (short GRADE) working group.
Grand, F. H., C. E. Hidalgo-Curtis, T. Ernst, K. Zoi, C. Zoi, and C. McGuire, 2009, Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms, v. 113, p. 6182-6192.
Guglielmelli, P., A. Pacilli, G. Rotunno, E. Rumi, V. Rosti, F. Delaini, M. Maffioli, T. Fanelli, A. Pancrazzi, D. Pietra, S. Salmoiraghi, C. Mannarelli, A. Franci, C. Paoli, A. Rambaldi, F. Passamonti, G. Barosi, T. Barbui, M. Cazzola, and A. M. Vannucchi, 2017, Presentation and outcome of patients with 2016

WHO diagnosis of prefibrotic and overt primary myelofibrosis: Blood, v. 129, p. 3227.

Haase, V. H., 2013, Regulation of erythropoiesis by hypoxia-inducible factors: Blood reviews, v. 27, p. 41-53.

Harrell Jnr, F. E., 2015, rms: Regression Modeling Strategies.

Harrison, C. N., D. Bareford, N. Butt, P. Campbell, E. Conneally, M. Drummond, W. Erber, T. Everington, A. R. Green, G. W. Hall, B. J. Hunt, C. A. Ludlam, R. Murrin, C. Nelson-Piercy, D. H. Radia, J. T. Reilly, J. Van der Walt, B. Wilkins, M. F. McMullin, and B. C. f. S. i. Haematology, 2010, Guideline for investigation and management of adults and children presenting with a thrombocytosis: Br J Haematol, v. 149, p. 352-75.

Harrison, C. N., N. Butt, P. Campbell, E. Conneally, M. Drummond, A. R. Green, R. Murrin, D. H. Radia, A. Mead, J. T. Reilly, N. C. P. Cross, and M. F. McMullin, 2014, Modification of British Committee for Standards in Haematology diagnostic criteria for essential thrombocythaemia: British Journal of Haematology, v. 167, p. 421-423.

Hartley, R. M., A. M. Epstein, C. M. Harris, and B. J. McNeil, 1987, Differences in ambulatory test ordering in England and America. Role of doctors' beliefs and attitudes: Am J Med, v. 82, p. 513-7.

Jacobs, N. L., S. G. Holtan, L. F. Porrata, S. N. Markovic, A. Tefferi, and D. P. Steensma, Host immunity affects survival in myelodysplastic syndromes: Independent prognostic value of the absolute lymphocyte count.

Jaiswal, S., P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler, and B. L. Ebert, 2014, Age-related clonal hematopoiesis associated with adverse outcomes: N Engl J Med, v. 371, p. 2488-98.

James, C., V. Ugo, J. P. Le Couedic, J. Staerk, F. Delhommeau, and C. Lacout, 2005, A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera: Nature, v. 434, p. 1144-1148.

Jamieson, C. H. M., J. Gotlib, J. A. Durocher, M. P. Chao, M. R. Mariappan, M. Lay, C. Jones, J. L. Zehnder, S. L. Lilleberg, and I. L. Weissman, 2006, The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation: Proceedings of the National Academy of Sciences, v. 103, p. 6224.

Jensen, M. K., P. De Nully Brown, O. J. Nielsen, and H. C. Hasselbalch, 2000, Incidence, clinical features and outcome of essential thrombocythaemia in a well defined geographical area: European Journal of Haematology, v. 65, p. 132-139.

Kaser, A., G. Brandacher, W. Steurer, S. Kaser, F. A. Offner, H. Zoller, I. Theurl, W. Widder, C. Molnar, O. Ludwiczek, M. B. Atkins, J. W. Mier, and H. Tilg, 2001, Interleukin-6 stimulates thrombopoiesis through thrombopoietin: role in inflammatory thrombocytosis: Blood, v. 98, p. 2720-5.

Kaushansky, K., 2014, Lineage-Specific Hematopoietic Growth Factors — NEJM.

Klampfl, T., H. Gisslinger, A. S. Harutyunyan, H. Nivarthi, E. Rumi, and J. D. Milosevic, 2013a, Somatic mutations of calreticulin in myeloproliferative neoplasms: New Engl J Med, v. 369, p. 2379-2390.

Klampfl, T., H. Gisslinger, A. S. Harutyunyan, H. Nivarthi, E. Rumi, J. D. Milosevic, N. C. Them, T. Berg, B. Gisslinger, D. Pietra, D. Chen, G. I. Vladimer, K. Bagienski, C. Milanesi, I. C. Casetti, E. Sant'Antonio, V. Ferretti, C. Elena, F. Schischlik, C. Cleary, M. Six, M. Schalling, A. Schonegger, C. Bock, L. Malcovati, C. Pascutto, G. Superti-Furga, M. Cazzola, and R. Kralovics, 2013b, Somatic mutations of calreticulin in myeloproliferative neoplasms: N Engl J Med, v. 369, p. 2379-90.

Koopmans, S. M., A. M. van Marion, and H. C. Schouten, 2012, Myeloproliferative neoplasia: a review of clinical criteria and treatment: Neth J Med, v. 70, p. 159-67.

Kralovics, R., F. Passamonti, A. S. Buser, S. S. Teo, R. Tiedt, and J. R. Passweg, 2005, A gain-of-function mutation of JAK2 in myeloproliferative disorders, v. 352, p. 1779-1790.

Kraus, A., M. Michalak, and M. Joachim Krebs and Marek, 2007, Endoplasmic reticulum dynamics and calcium signaling, CalciumA Matter of Life or Death, v. Volume 41, Elsevier, p. 199-218.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, and L. Scrucca, 2015 caret: Classification and Regression Training.

Kurita, S., 1974, [Epidemiological studies of polycythemia vera in Japan (author's transl)]: Nihon Ketsueki Gakkai Zasshi, v. 37, p. 793-5.

Laszlo, J., 1975, Myeloproliferative disorders (MPD): Myelofibrosis, myelosclerosis, extramedullary hematopoiesis, undifferentiated MPD, and hemorrhagic thrombocythemia: Seminars in Hematology, v. 12, p. 409-432.

Lee, F. S., and M. J. Percy, 2011, The HIF pathway and erythrocytosis: Annu Rev Pathol, v. 6, p. 165-92.

Lee, T. S., H. Kantarjian, W. Ma, C. H. Yeh, F. Giles, and M. Albitar, 2011, Effects of clinically relevant MPL mutations in the transmembrane domain revealed at the atomic level through computational modeling: PLoS One, v. 6, p. e23396.

Levine, R. L., M. Wadleigh, J. Cools, B. L. Ebert, G. Wernig, and B. J. Huntly, 2005, Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis, v. 7, p. 387-397.

Liaw, A., and M. Wiener, 2002, Classification and Regression by randomForest, R News, p. 18-22.

Lichtman, M. A., 2013, Does a Diagnosis of Myelogenous Leukemia Require 20% Marrow Myeloblasts, and Does <5% Marrow Myeloblasts Represent a Remission? The History and Ambiguity of Arbitrary Diagnostic Boundaries in the Understanding of Myelodysplasia, Oncologist, v. 18: Durham, NC, USA, p. 973-80.

Marchioli, R., G. Finazzi, R. Landolfi, J. Kutti, H. Gisslinger, C. Patrono, R. Marilus, A. Villegas, G. Tognoni, and T. Barbui, 2005, Vascular and neoplastic risk in a large cohort of patients with polycythemia vera: J Clin Oncol, v. 23, p. 2224-32.

Marquaridt, D. W., 1970, Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation: Technometrics, v. 12, p. 591-612.

Martinez-Aviles, L., C. Besses, A. Alvarez-Larran, F. Cervantes, J. C. Hernandez-Boluda, and B. Bellosillo, 2007, JAK2 exon 12 mutations in polycythemia vera or idiopathic erythrocytosis, v. 92, p. 1717-1718.

Marty, C., C. Pecquet, H. Nivarthi, M. El-Khoury, I. Chachoua, M. Tulliez, J.-L. Villeval, H. Raslova, R. Kralovics, S. N. Constantinescu, I. Plo, and W. Vainchenker, 2016, Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis: Blood, v. 127, p. 1317-1324.

McMullin, M. F., D. Bareford, P. Campbell, A. R. Green, C. Harrison, B. Hunt, D. Oscier, M. I. Polkey, J. T. Reilly, E. Rosenthal, K. Ryan, T. C. Pearson, B. Wilkins, and G. H. T. F. o. t. B. C. f. S. i. Haematology, 2005, Guidelines for the diagnosis, investigation and management of polycythaemia/erythrocytosis: Br J Haematol, v. 130, p. 174-95.

McMullin, M. F., D. Bareford, J. Craig, A. R. Green, C. Harrison, D. G. Oscier, J. T. Reilly, and M. D. U. S. Group, 2003, The optimal management of polycythaemia vera: Br J Haematol, v. 120, p. 543-4; author reply 544-5.

McMullin, M. F., J. T. Reilly, P. Campbell, D. Bareford, A. R. Green, C. N. Harrison, E. Conneally, K. Ryan, M. e. D. S. National Cancer Research Institute, and B. C. f. S. i. Haematology, 2007, Amendment to the guideline for diagnosis and investigation of polycythaemia/erythrocytosis: Br J Haematol, v. 138, p. 821-2.

Mead, A. J., and A. Mullally, 2017, Myeloproliferative neoplasm stem cells: Blood, v. 129, p. 1607.

Mesa, R. A., J. Mehta, H. Wang, Y. Wang, U. Iqbal, F. Neumann, Y. Zhang, and T. Colton, 2012, Epidemiology of Myeloproliferative Disorders in US - a Real World Analysis: Blood, v. 120, p. 2834.

Michiels, J. J., and E. Juvonen, 1997, Proposal for revised diagnostic criteria of essential thrombocythemia and polycythemia vera by the Thrombocythemia Vera Study Group: Semin Thromb Hemost, v. 23, p. 339-47.

Moulard, O., J. Mehta, J. Fryzek, R. Olivares, U. Iqbal, and R. A. Mesa, 2014, Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union: Eur J Haematol, v. 92, p. 289-97.

Mudireddy, M., N. Gangat, C. A. Hanson, R. P. Ketterling, A. Pardanani, and A. Tefferi, 2018, Validation of the WHO-defined 20% circulating blasts threshold for diagnosis of leukemic transformation in primary myelofibrosis, v. 8, p. 57.

Murphy, S., H. Iland, D. Rosenthal, and J. Laszlo, 1986, Essential thrombocythemia: an interim report from the Polycythemia Vera Study Group: Semin Hematol, v. 23, p. 177-82.

Murphy, S., P. Peterson, H. Iland, and J. Laszlo, 1997, Experience of the Polycythemia Vera Study Group with essential thrombocythemia: a final report on diagnostic criteria, survival, and leukemic transition by treatment: Semin Hematol, v. 34, p. 29-39.

Nangalia, J., and T. R. Green, 2014, The evolving genomic landscape of myeloproliferative neoplasms: ASH Education Program Book, v. 2014, p. 287-296.

Nangalia, J., C. E. Massie, E. J. Baxter, F. L. Nice, G. Gundem, and D. C. Wedge, 2013, Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2: New Engl J Med, v. 369, p. 2391-2405.

Nielsen, C., H. S. Birgens, B. G. Nordestgaard, and S. E. Bojesen, 2013, Diagnostic value of JAK2 V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population: Br J Haematol, v. 160, p. 70-9.

Nielsen, C., H. S. Birgens, B. G. Nordestgaard, L. Kjær, and S. E. Bojesen, 2011, The JAK2 V617F somatic mutation, mortality and cancer risk in the general population: Haematologica, v. 96, p. 450.

Nielsen, C., S. E. Bojesen, B. G. Nordestgaard, K. F. Kofoed, and H. S. Birgens, 2014, JAK2 V617F somatic mutation in the general population: myeloproliferative neoplasm development and progression rate: Haematologica, v. 99, p. 1448.

Oh, S. T., E. F. Simonds, C. Jones, M. B. Hale, Y. Goltsev, K. D. Gibbs, J. D. Merker, J. L. Zehnder, G. P. Nolan, and J. Gotlib, 2010, Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms: Blood, v. 116, p. 988.

Orazi, A., D. P. O'Malley, and D. A. Arber, 2006, Illustrated Pathology of the Bone Marrow: Cambridge, Cambridge University Press.

Ortmann, C. A., D. G. Kent, J. Nangalia, Y. Silber, D. C. Wedge, J. Grinfeld, E. J. Baxter, C. E. Massie, E. Papaemmanuil, S. Menon, A. L. Godfrey, D. Dimitropoulou, P. Guglielmelli, B. Bellosillo, C. Besses, K. Döhner, C. N. Harrison, G. S. Vassiliou, A. Vannucchi, P. J. Campbell, and A. R. Green, 2015, Effect of Mutation Order on Myeloproliferative Neoplasms: New England Journal of Medicine, v. 372, p. 601-612.

Pardanani, A. D., R. L. Levine, T. Lasho, Y. Pikman, R. A. Mesa, and M. Wadleigh, 2006, MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients: Blood, v. 108, p. 3472-3476.

Passamonti, F., E. Rumi, L. Arcaini, E. Boveri, C. Elena, D. Pietra, S. Boggi, C. Astori, P. Bernasconi, M. Varettoni, E. Brusamolino, C. Pascutto, and M. Lazzarino, 2008, Prognostic factors for thrombosis, myelofibrosis, and leukemia in essential thrombocythemia: a study of 605 patients: Haematologica, v. 93, p. 1645-1651.

Passamonti, F., E. Rumi, E. Pungolino, L. Malabarba, P. Bertazzoni, M. Valentini, E. Orlandi, L. Arcaini, E. Brusamolino, C. Pascutto, M. Cazzola, E. Morra, and M. Lazzarino, 2004, Life expectancy and prognostic factors for survival in patients with polycythemia vera and essential thrombocythemia: Am J Med, v. 117, p. 755-61.

Patel, J. P., M. E. Gonen M Fau - Figueroa, H. Figueroa Me Fau - Fernandez, Z. Fernandez H Fau - Sun, J. Sun Z Fau - Racevskis, P. Racevskis J Fau - Van Vlierberghe, I. Van Vlierberghe P Fau - Dolgalev, S. Dolgalev I Fau - Thomas, O. Thomas S Fau - Aminova, K. Aminova O Fau - Huberman, J. Huberman K Fau - Cheng, A. Cheng J Fau - Viale, N. D. Viale A Fau - Socci, A. Socci Nd Fau - Heguy, A. Heguy A Fau - Cherry, G. Cherry A Fau - Vance, R. R. Vance G Fau - Higgins, R. P. Higgins Rr Fau - Ketterling, R. E. Ketterling Rp Fau - Gallagher, M. Gallagher Re Fau - Litzow, M. R. M. Litzow M Fau - van den Brink, H. M. van den Brink Mr Fau - Lazarus, J. M. Lazarus Hm Fau - Rowe, S. Rowe Jm Fau - Luger, A. Luger S Fau - Ferrando, E. Ferrando A Fau - Paietta, M. S. Paietta E Fau - Tallman, A. Tallman Ms Fau - Melnick, O. Melnick A Fau - Abdel-Wahab, R. L. Abdel-Wahab O Fau - Levine, and R. L. Levine, Prognostic relevance of integrated genetic profiling in acute myeloid leukemia.

Pecquet, C., J. Staerk, R. Chaligné, V. Goss, K. A. Lee, X. Zhang, J. Rush, J. Van Hees, H. A. Poirel, J.-M. Scheiff, W. Vainchenker, S. Giraudier, R. D. Polakiewicz, and S. N. Constantinescu, 2010, Induction of myeloproliferative

disorder and myelofibrosis by thrombopoietin receptor W515 mutants is mediated by cytosolic tyrosine 112 of the receptor: Blood, v. 115, p. 1037.

Phekoo, K. J., M. A. Richards, H. Moller, and S. A. Schey, 2006, The incidence and outcome of myeloid malignancies in 2,112 adult patients in southeast England: Haematologica, v. 91, p. 1400-4.

Pietra, D., S. Li, A. Brisci, F. Passamonti, E. Rumi, and A. Theocharides, 2008, Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders, v. 111, p. 1686-1689.

Pietra, D., E. Rumi, V. V. Ferretti, C. A. D. Buduo, C. Milanesi, C. Cavalloni, E. Sant'Antonio, V. Abbonante, F. Moccia, I. C. Casetti, M. Bellini, M. C. Renna, E. Roncoroni, E. Fugazza, C. Astori, E. Boveri, V. Rosti, G. Barosi, A. Balduini, and M. Cazzola, 2015, Differential clinical effects of different mutation subtypes in CALR-mutant myeloproliferative neoplasms, v. 30, p. 431.

Pikman, Y., B. H. Lee, T. Mercher, E. McDowell, B. L. Ebert, and M. Gozo, 2006, MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia: PLoS Med, v. 3, p. e270.

Price, G., K. L Davis, S. Karve, G. Pohl, and R. A Walgren, 2014, Survival Patterns in United States (US) Medicare Enrollees with Non-CML Myeloproliferative Neoplasms (MPN): PloS one, e90299 p.

Prochazka, A. V., K. Lundahl, W. Pearson, S. K. Oboler, and R. J. Anderson, 2005, Support of evidence-based guidelines for the annual physical examination: a survey of primary care providers: Arch Intern Med, v. 165, p. 1347-52.

Radia, D., and H. L. Geyer, 2015, Management of symptoms in polycythemia vera and essential thrombocythemia patients: Hematology Am Soc Hematol Educ Program, v. 2015, p. 340-8.

Reilly, J. T., M. F. McMullin, P. A. Beer, N. Butt, E. Conneally, A. Duncombe, A. R. Green, N. G. Michaeel, M. H. Gilleece, G. W. Hall, S. Knapper, A. Mead, R. A. Mesa, M. Sekhar, B. Wilkins, and C. N. Harrison, 2012a, Guideline for the diagnosis and management of myelofibrosis: British Journal of Haematology, v. 158, p. 453-471.

Reilly, J. T., M. F. McMullin, P. A. Beer, N. Butt, E. Conneally, A. S. Duncombe, A. R. Green, G. Mikhaeel, M. H. Gilleece, S. Knapper, A. J. Mead, R. A. Mesa, M. Sekhar, and C. N. Harrison, 2014, Use of JAK inhibitors in the management of myelofibrosis: a revision of the British Committee for Standards in Haematology Guidelines for Investigation and Management of Myelofibrosis 2012, v. 167, p. 418-420.

Renzi, C., K. L. Whitaker, and J. Wardle, 2015, Over-reassurance and undersupport after a 'false alarm': a systematic review of the impact on subsequent cancer symptom attribution and help seeking: BMJ Open, v. 5, p. e007002.

Renzi, C., K. L. Whitaker, K. Winstanley, S. Cromme, and J. Wardle, 2016, Unintended consequences of an 'all-clear' diagnosis for potential cancer symptoms: a nested qualitative interview study with primary care patients: Br J Gen Pract, v. 66, p. e158-70.

Ridgeway, G., 2015, gbm: Generalized Boosted Regression Models.

Rollison, D. E., N. Howlader, M. T. Smith, S. S. Strom, W. D. Merritt, L. A. Ries, B. K. Edwards, and A. F. List, 2008, Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs: Blood, v. 112, p. 45-52.

Roman, E., A. Smith, S. Appleton, S. Crouch, R. Kelly, S. Kinsey, C. Cargo, and R. Patmore, 2016, Myeloid malignancies in the real-world: Occurrence, progression and survival in the UK's population-based Haematological Malignancy Research Network 2004-15: Cancer Epidemiology, v. 42, p. 186-98.

Rotunno, G., C. Mannarelli, P. Guglielmelli, A. Pacilli, A. Pancrazzi, and L. Pieri, 2013a, Impact of calreticulin mutations on clinical and hematological phenotype and outcome in essential thrombocythemia.

Rumbaut, R. E., and P. Thiagarajan, 2010, Integrated Systems Physiology: from Molecule to Function to Disease, Platelet-Vessel Wall Interactions in Hemostasis and Thrombosis: San Rafael (CA), Morgan & Claypool Life Sciences

Rumi, E., A. S. Harutyunyan, D. Pietra, J. D. Milosevic, I. C. Casetti, M. Bellini, N. C. Them, C. Cavalloni, V. V. Ferretti, C. Milanesi, T. Berg, E. Sant'antonio, E. Boveri, C. Pascutto, C. Astori, R. Kralovics, and M. Cazzola, 2014, CALR exon 9 mutations are somatically acquired events in familial cases of essential thrombocythemia or primary myelofibrosis: Blood.

Rumi, E., D. Pietra, V. Ferretti, T. Klampfl, A. S. Harutyunyan, and J. D. Milosevic, 2013, JAK2 or CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes.

Rumi, E., E. Sant'Antonio, E. Boveri, D. Pietra, C. Cavalloni, E. Roncoroni, C. Astori, and L. Arcaini, 2018, Diagnosis and management of prefibrotic myelofibrosis: Expert Rev Hematol, v. 11, p. 537-545.

Sagripanti, A., A. Ferretti, A. Nicolini, and A. Carpi, 1996, Thrombotic and hemorrhagic complications in chronic myeloproliferative disorders: Biomed Pharmacother, v. 50, p. 376-82.

Sanada, M., T. Suzuki, L. Y. Shih, M. Otsu, M. Kato, and S. Yamazaki, 2009, Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms, v. 460, p. 904-908.

Schalling, M., A. Gleiss, B. Gisslinger, A. Wölfler, V. Buxhofer-Ausch, G. Jeryczynski, M.-T. Krauth, I. Simonitsch-Klupp, C. Beham-Schmid, J. Thiele, and H. Gisslinger, 2017, Essential thrombocythemia vs. pre-fibrotic/early primary myelofibrosis: discrimination by laboratory and clinical data, v. 7, p. 643.

Scott, L. M., W. Tong, R. L. Levine, M. A. Scott, P. A. Beer, and M. R. Stratton, 2007, JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis: New Engl J Med, v. 356, p. 459-468.

Shih, A. H., O. Abdel-Wahab, J. P. Patel, and R. L. Levine, 2012, The role of mutations in epigenetic regulators in myeloid malignancies, v. 12, p. 599.

Shih, L. Y., and C. T. Lee, 1994, Identification of masked polycythemia vera from patients with idiopathic marked thrombocytosis by endogenous erythroid colony assay: Blood, v. 83, p. 744-748.

Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer, 2005, ROCR: visualizing classifier performance in R.: Bioinformatics, v. 21, p. 2.

Smith, A., D. Howell, S. Crouch, D. Painter, J. Blase, H.-I. Wang, A. Hewison, T. Bagguley, S. Appleton, S. Kinsey, C. Burton, R. Patmore, and E. Roman, 2018, Cohort Profile: The Haematological Malignancy Research Network (HMRN); a UK population-based patient cohort, p. dyy044-dyy044.

Smith, A., E. Roman, D. Howell, R. Jones, R. Patmore, and A. Jack, 2010, The Haematological Malignancy Research Network (HMRN): a new information strategy for population based epidemiology and health service research: Br J Haematol, v. 148, p. 739-53.

Somers, R. H., 1962, A New Asymmetric Measure of Association for Ordinal Variables, v. 27, p. 799-811.

Steensma, D. P., R. Bejar, S. Jaiswal, R. C. Lindsley, M. A. Sekeres, R. P. Hasserjian, and B. L. Ebert, 2015, Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes: Blood, v. 126, p. 9.

Swerdlow, S. H., C. International Agency for Research on, and O. World Health, 2008, WHO classification of tumours of haematopoietic and lymphoid tissues: Lyon, France, International Agency for Research on Cancer.

Team, R. C., 2018, R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria.

Tefferi, A., 2018, Primary myelofibrosis: 2019 update on diagnosis, risk-stratification and management: Am J Hematol.

Tefferi, A., P. Guglielmelli, D. R. Larson, C. Finke, E. A. Wassie, L. Pieri, N. Gangat, R. Fjerza, A. A. Belachew, T. L. Lasho, R. P. Ketterling, C. A. Hanson, A. Rambaldi, G. Finazzi, J. Thiele, T. Barbui, A. Pardanani, and A. M. Vannucchi, 2014a, Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis: Blood, v. 124, p. 2507-13; quiz 2615.

Tefferi, A., T. L. Lasho, C. M. Finke, R. A. Knudson, R. Ketterling, and C. H. Hanson, 2014b, CALR vs JAK2 vs MPL- mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons.

Tefferi, A., A. M. Vannucchi, and T. Barbui, 2018, Polycythemia vera treatment algorithm 2018: Blood Cancer J, v. 8, p. 3.

Tefferi, A., E. A. Wassie, T. L. Lasho, C. Finke, A. A. Belachew, R. P. Ketterling, C. A. Hanson, A. Pardanani, N. Gangat, and A. P. Wolanskyj, 2014c, Calreticulin mutations and long-term survival in essential thrombocythemia: Leukemia.

Tenen, D. G., R. Hromas, J. D. Licht, and D. E. Zhang, 1997, Transcription factors, normal myeloid development, and leukemia, v. 90, p. 489-519.

Therneau, T., 2015, A Package for Survival Analysis in S.

Therneau, T., and P. M. Grambsch, 2000, Modeling Survival Data: Extending the Cox Model: New York, Springer.

Thiele, J., H. M. Kvasnicka, L. Müllauer, V. Buxhofer-Ausch, B. Gisslinger, and H. Gisslinger, 2011, Essential thrombocythemia versus early primary myelofibrosis: a multicenter study to validate the WHO classification.

Titmarsh, G. J., A. S. Duncombe, M. F. McMullin, M. O'Rorke, R. Mesa, F. De Vocht, S. Horan, L. Fritschi, M. Clarke, and L. A. Anderson, 2014, How common are myeloproliferative neoplasms? A systematic review and meta-analysis, v. 89, p. 581-587.

Vainchenker, W., and S. N. Constantinescu, 2012, JAK/STAT signaling in hematological malignancies, v. 32, p. 2601.

Vainchenker, W., F. Delhommeau, S. N. Constantinescu, and O. A. Bernard, 2011, New mutations and pathogenesis of myeloproliferative neoplasms: Blood, v. 118, p. 1723.

van Bokhoven, M. A., M. C. H. Pleunis-van Empel, H. Koch, R. P. T. M. Grol, G.-J. Dinant, and T. van der Weijden, 2006, Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice: BMC Family Practice, v. 7, p. 75-75.

van der Weijden, T., M. A. van Bokhoven, G. J. Dinant, C. M. van Hasselt, and R. P. Grol, 2002, Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice: Br J Gen Pract, v. 52, p. 974-80.

Vardiman, J. W., J. Thiele, D. A. Arber, R. D. Brunning, M. J. Borowitz, and A. Porwit, 2009a, The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes: Blood, v. 114, p. 937-951.

Vardiman, J. W., J. Thiele, D. A. Arber, R. D. Brunning, M. J. Borowitz, A. Porwit, N. L. Harris, M. M. Le Beau, E. Hellström-Lindberg, A. Tefferi, and C. D. Bloomfield, 2009b, The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes: Blood, v. 114, p. 937-51.

Venables, W. N., and B. D. Ripley, 2002, Modern Applied Statistics with S., New York, Springer.

Visser, O., A. Trama, M. Maynadié, C. Stiller, R. Marcos-Gragera, R. De Angelis, S. Mallone, C. Tereanu, C. Allemani, U. Ricardi, and H. C. Schouten, 2012, Incidence, survival and prevalence of myeloid malignancies in Europe, v. 48, p. 3257-3266.

Warnes, G. R., B. Bolker, T. Lumley, and R. C. Johnson, 2013, gmodels: Various R programming tools for model fitting.

Wasserman, L. R., 1971, The management of polycythaemia vera: Br J Haematol, v. 21, p. 371-6.

Wilkins, B. S., W. N. Erber, D. Bareford, G. Buck, K. Wheatley, C. L. East, B. Paul, C. N. Harrison, A. R. Green, and P. J. Campbell, 2008, Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes.

Wolanskyj, A. P., S. M. Schwager, R. F. McClure, D. R. Larson, and A. Tefferi, 2006, Essential thrombocythemia beyond the first decade: life expectancy, long-term complication rates, and prognostic factors: Mayo Clin Proc, v. 81, p. 159-66.

Wolber, E. M., J. Fandrey, U. Frackowski, and W. Jelkmann, 2001, Hepatic thrombopoietin mRNA is increased in acute inflammation: Thromb Haemost, v. 86, p. 1421-4.

Zhao, S., Y. Lin, W. Xu, W. Jiang, Z. Zha, and P. Wang, 2009, Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha, v. 324, p. 261-265.