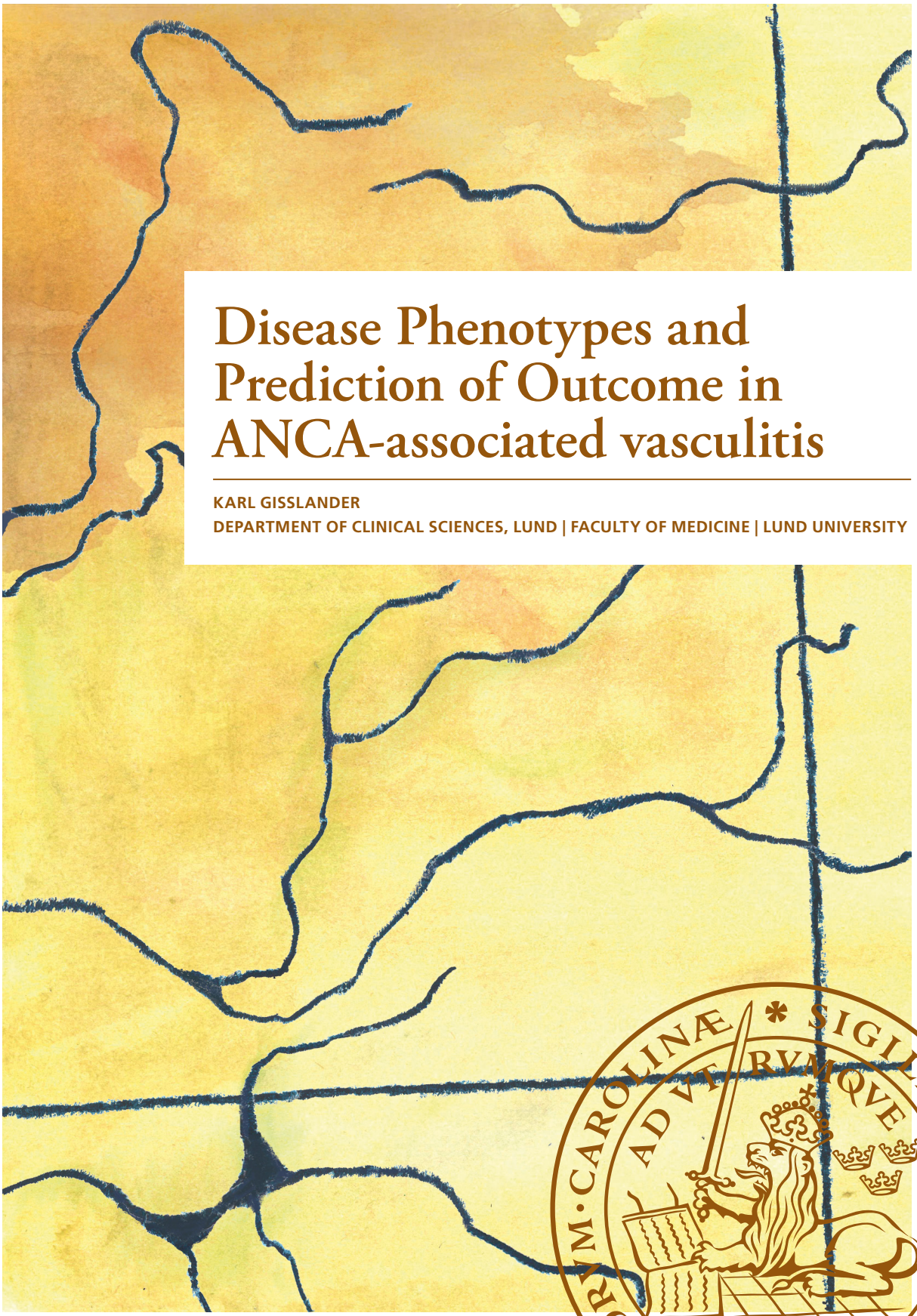
Disease Phenotypes and Prediction of Outcome in ANCA-associated vasculitis

Gisslander, Karl

2024

Document Version:
Publisher's PDF, also known as Version of record



Disease Phenotypes and Prediction of Outcome in ANCA-associated vasculitis

KARL GISSLANDER
DEPARTMENT OF CLINICAL SCIENCES, LUND | FACULTY OF MEDICINE | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2024:134
ISBN 978-91-8021-632-6
ISSN 1652-8220

Disease Phenotypes and Prediction of Outcome in ANCA-associated vasculitis

Karl Gisslander

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on 8th of November 2024 at 09.00 in Lottasalen, Department of Rheumatology, Skåne University Hospital

Faculty opponent
Professor Raashid Luqmani, University of Oxford, United Kingdom

Organization: LUND UNIVERSITY

Document name: Doctoral dissertation **Date of issue:** 2024-11-08

Author(s): Karl Gisslander **Sponsoring organization:**

Title and subtitle: Disease Phenotypes and Prediction of Outcome in ANCA-associated vasculitis

Abstract:

Objectives
Anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV) is a rare systemic autoimmune disease characterised by inflammation and destruction of small blood vessels. With research into the disease suffering from small sample sizes, this thesis aims to (1) address research data fragmentation in AAV through the integration of real-world observational registries, (2) stratify patterns of symptoms at disease onset, and (3) build models for the prediction of disease outcome.
Methods
Data from six European vasculitis registries were integrated using two complementary approaches. First the registries were combined using Semantic Web technologies, allowing federated access to aggregated data through a dedicated web interface. The quality of the underlying data was explored, and the characteristics, treatments and disease outcome of European patients described. Secondly, data was pooled in a central data storage. Using the central data, model-based clustering was used to study and stratify the diverse phenotypic presentations at disease onset. Lastly, prognostic models for key disease outcomes were built using survival modelling.
Results
The federated integration was successful, although some data quality concerns were identified, allowing access to an unprecedented cohort size of 5282 patients. Symptomatology, type of treatments used, mortality rates and rates of end-stage kidney disease were highly variable between the participating registries. Using model-based clustering, five clusters were identified, with distinct phenotypes, biochemical presentations, and disease outcomes – primarily stratified by kidney impairment and systemic inflammation. Building predictive models for disease outcome, known predictors of disease outcome were reidentified and compiled into comprehensive models, outperforming existing models in terms of predictive accuracy.
Conclusion
This thesis presents the first successful federated integration of distributed vasculitis datasets, allowing access to a cohort of unprecedented size. It further reinforces that AAV is beyond a binary construct and that the disease heterogeneity may be better described by five subcategories. While accurate prediction of disease outcome at the time of diagnosis is possible, the benefit of implementation of prediction models for the guidance of clinical decision-making needs further evaluation.

Key words: ANCA, vasculitis, epidemiology, cluster, classification, prediction, outcome

Classification system and/or index terms (if any) Supplementary bibliographical information

Language: English **Number of pages**: 120

ISSN and key title: 1652-8220 Lund University, Faculty of Medicine Doctoral Dissertation Series 2024:134

ISBN: 978-91-8021-632-6

Recipient's notes Price Security classification

Signature Date 2024-09-29

Disease Phenotypes and Prediction of Outcome in ANCA-associated vasculitis

Karl Gisslander

LUND
UNIVERSITY

Coverphoto by Studio Dartlander

Faculty of Medicine
Department of Clinical Sciences, Lund

ISBN 978-91-8021-632-6
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

We now use the country itself, as its own map,
and I assure you it does nearly as well
LEWIS CARROLL

Table of Contents

Abstract

Objectives

Anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV) is a rare systemic autoimmune disease characterised by inflammation and destruction of small blood vessels. With research into the disease suffering from small sample sizes, this thesis aims to (1) address research data fragmentation in AAV through the integration of real-world observational vasculitis registries, (2) stratify patterns of symptoms at disease onset, and (3) build models for the prediction of disease outcome.

Methods

Data from six European vasculitis registries were integrated using two complementary approaches. First the registries were combined using Semantic Web technologies, allowing federated access to aggregated data through a dedicated web interface. The quality of the underlying data was explored, and the characteristics, treatments and disease outcome of European patients described. Secondly, data was pooled in a central data storage. Using the central data, model-based clustering was used to study and stratify the diverse phenotypic presentations at disease onset. Lastly, prognostic models for key disease outcomes were built using survival modelling.

Results

The federated integration was successful, although some data quality concerns were identified, allowing access to an unprecedented cohort size of 5282 patients. Symptomatology, type of treatments used, mortality rates and rates of end-stage kidney disease were highly variable between the participating registries. Using model-based clustering, five clusters were identified, with distinct phenotypes, biochemical presentations, and disease outcomes – primarily stratified by kidney impairment and systemic inflammation. Building predictive models for disease outcome, known predictors of disease outcome were reidentified and compiled into comprehensive models, outperforming existing models in terms of predictive accuracy.

Conclusion

This thesis presents the first successful federated integration of distributed vasculitis datasets, allowing access to a cohort of unprecedented size. It further reinforces that AAV is beyond a binary construct and that the disease heterogeneity may be better described by five subcategories. While accurate prediction of disease outcome at the time of diagnosis is possible, the benefit of implementation of prediction models for the guidance of clinical decision-making needs further evaluation.

List of papers

I. **Gisslander K**, Rutherford M, Aslett L, Basu N, Dradin F, Hederman L, Hrušková Z, Kardaoui H, Lamprecht P, Lichołai S, Musiał J, O'Sullivan D, Puéchal X, Scott J, Segelmark M, Straka R, Terrier B, Tesař V, Tesi M, Vaglio A, Wandrei D, White A, Wójcik K, Yaman B, Little MA, Mohammad AJ, on behalf of the FAIRVASC consortium. Data quality and patient characteristics in European ANCA-associated vasculitis registries: data retrieval by federated querying. Ann Rheum Dis. 2024 Jan 2;83(1):112-120.

II. **Gisslander K**, White A, Aslett L, Hrušková Z, Lamprecht P, Musiał J, Nazeer J, Ng J, O'Sullivan D, Puéchal X, Rutherford M, Segelmark M, Terrier B, Tesař V, Tesi M, Vaglio A, Wójcik K, Little MA, Mohammad AJ, on behalf of the FAIRVASC consortium. Data-driven subclassification of ANCA associated vasculitis – model-based clustering of a federated international cohort. Lancet Rheumatol. 2024 Aug 22:S2665-9913(24)00187-5.

III. **Gisslander K**, White A, Segelmark M, Hrušková Z, Lamprech P, Najibi M, Puechal X, Rathmann J, Rutherford M, Terrier B, Vaglio A, Wójcik K, Little MA, Mohammad AJ, on behalf of the FAIRVASC consortium. Prognostic models for ANCA-associated vasculitis: a real-life observational cohort study of relapse, infection, end-stage kidney disease and mortality. (manuscript)

Abbreviations

AAV	Anti-neutrophil cytoplasmic antibody-associated vasculitis
ACR	American College of Rheumatology
AIC	Akaike information criterion
ANCA	Anti-neutrophil cytoplasmic antibody
ASN	American Society of Nephrology
AUC	Area under the curve of a receiver operating characteristics curve
BIC	Bayesian information criterion
BVAS	Birmingham Vasculitis Activity Score
cANCA	Cytoplasmic anti-neutrophil cytoplasmic antibody pattern
CI	Confidence interval
CHCC	Chapel Hill Consensus Conference
CRP	C-reactive protein
DNA	Deoxyribonucleic acid
eGFR	Estimated glomerular filtration rate
EGPA	Eosinophilic granulomatosis with polyangiitis
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
ESKD	End-stage kidney disease
EULAR	European Alliance of Associations for Rheumatology
EVI	Equal volume, varying shape, and orientation
FAIR	Findable, accessible, interoperable, reusable
FAIRVASC	Findable, accessible, interoperable, reusable, vasculitis
FVSG	French Vasculitis Study Group (registry)
GeVas	Joint Vasculitis Registry in German-speaking countries
GPA	Granulomatosis with polyangiitis
ICD	International Classification of Diseases
IIF	Indirect immunofluorescence
i~HD	European Institute for Innovation through Health Data

IMS	Inflammatory multisystem cluster
HLA	Human leukocyte antigen
MHC	Major histocompatibility complex
MPA	Microscopic polyangiitis
MPO	Myeloperoxidase
MPO-K	Anti-MPO kidney cluster
NET	Neutrophil extracellular traps
OMERACT	Outcome-measures in Rheumatology
pANCA	Perinuclear anti-neutrophil cytoplasmic antibody pattern
POLVAS	Polish Vasculitis Registry
PR3	Proteinase 3
PR3-K	Anti-PR3 kidney cluster
PROM	Patient-reported outcome measure
RCT	Randomised clinical trial
RKD	Ireland's Rare Kidney Disease registry
SK	Severe kidney cluster
VDI	Vasculitis damage index
VVI	Varying volume, varying shape, and orientation
WHO	World Health Organization
YR	Young respiratory cluster

Introduction

Anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV) is a rare systemic autoimmune disorder comprised of three diseases or disease subtypes, granulomatosis with polyangiitis (GPA), microscopic polyangiitis (MPA), and eosinophilic granulomatosis with polyangiitis (EGPA). All disease subtypes are characterised by inflammation and destruction of predominantly small blood vessels and share an association with autoantibodies, ANCA. However, the exact aetiopathology is poorly understood.

Although AAV may involve small blood vessels in any organ or tissue, the upper and lower respiratory tract and kidneys are most often affected. The rarity of the disease along with heterogenous symptomatology and a wide range of differential diagnoses makes AAV challenging to diagnose and study. Despite this, there have been considerable improvements in the diagnosis, treatment, and prognosis for patients with AAV. Still there is no cure, but what was once a disease with dire short-term prognosis, today is a chronic condition, unfortunately, involving frequent disease relapses.

While much is still to be learned about AAV, the considerable progress made in recent years is largely attributable to increased international collaboration. Being a rare disease, this is a necessity. This thesis describes, and benefits from such international collaboration, to explore the phenotypic spectrum of disease and to improve prognostication of disease progression. This is done using modern-day technologies such as the Semantic Web, unsupervised, and supervised learning. Of course, this has not always been possible, so let us start from the beginning.

A brief history of ANCA-associated vasculitis

In 1982 Davies and colleagues submitted a short report to the British Medical Journal, where they described the presence of "a factor that stained the cytoplasm of neutrophil leucocytes by indirect immunofluorescence" in eight patients with biopsy-proven necrotising crescentic glomerulonephritis.[1] This is the first report of ANCAs, but the history of AAV started much earlier and under other names.

Although vascular and rheumatic diseases have been recognised since antiquity, the first macroscopic description of what is believed to be systemic vasculitis was made

in 1852 by the Austrian pathologist Karl Rokitansky.[2, 3] The constellation of symptoms and histopathology were further described and named in 1866 by Adolf Kußmaul and Rudolf Maier as 'periarteritis nodosa,' a term later changed to polyarteritis nodosa.[4, 5]

Over the coming decades it became evident that the term polyarteritis nodosa encompassed a spectrum of different clinical presentations. In 1923 Friedrich Wohlwill described microscopic polyarteritis nodosa, while Heinz Klinger in 1931 described an atypical polyarteritis nodosa which Friedrich Wegener later recognized as a distinct disease with ear-nose-throat symptoms and histopathological findings characterised by a mixture of vasculitis and granuloma formation.[6, 7, 8, 9] Jacob Churg and Lotte Strauss would in 1949 describe cases of asthma and eosinophilia associated with granulomatous lesions of the vessel wall, and set this apart from classical polyarteritis nodosa, introducing the disease later bearing their names.[10]

These three, seemingly all related diseases, would in two cases be known by honorific eponyms, Wegener's granulomatosis (GPA) and Churg-Strauss syndrome (EGPA), while Wohlwill's name was omitted in the naming of microscopic polyarteritis nodosa (MPA). The use of eponyms persisted until the American Society of Nephrology (ASN), the American College of Rheumatology (ACR) and the European Alliance of Associations for Rheumatology (EULAR) in the early 2010's recommended a gradual shift to the disease-descriptive nomenclature we use today.[11, 12] By this time the pathogenic role of ANCAs had been widely described, but the umbrella term AAV not yet commonly used.

Nomenclature

The nomenclature of AAV is unarguably confusing. First, it is important to distinguish the difference in naming, defining, classifying, and diagnosing a disease. A nomenclature provides a standardised vocabulary aiding clinic and research in a uniform naming of diseases. The first attempt at a standardised vocabulary for the systemic vasculitides was published as the "Nomenclature of systemic vasculitides" in 1994 but is perhaps better known as the Chapel Hill Consensus Conference criteria (CHCC 1994).[13] Here the systemic vasculitides were subcategorised by the size of the predominantly involved blood vessel and subsequently named. The AAVs (although the umbrella-term AAV was not yet recognised) were categorised as small-vessel vasculitis and named Wegener's granulomatosis, microscopic polyangiitis (microscopic polyarteritis) and Churg-Strauss syndrome. The CHCC 1994 was updated and expanded in 2012, which is the current standardised nomenclature of systemic vasculitis (Figure 1).[11] Today, AAV is defined as a concept, categorised as a small vessel vasculitis with three disease subtypes (stripped of eponyms), granulomatosis with polyangiitis (GPA), microscopic

polyangiitis (MPA) and eosinophilic granulomatosis with polyangiitis (EGPA). However, to further add to the confusion, EGPA may sometimes be omitted when discussing AAV.[14]

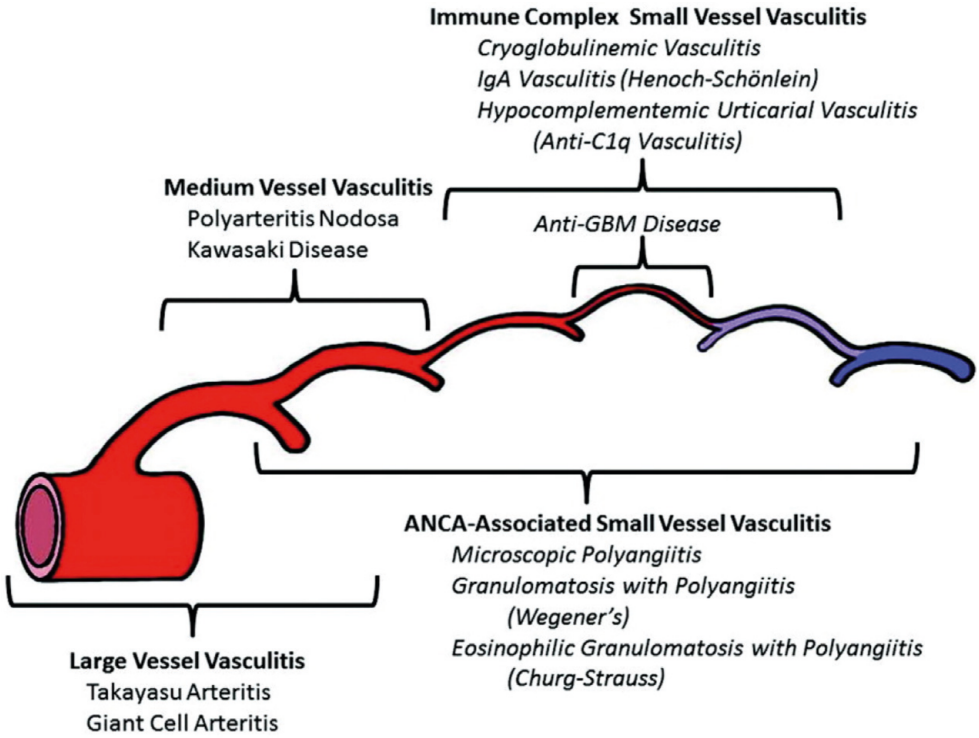


Figure 1. The 2012 Chapel Hill Consensus Conference criteria nomenclature of vasculitis
Distribution of vessel involvement by large, medium and small vessel vasculitis. Note the substantial overlap. The figure depicts (from left to right) aorta, large artery, medium artery, small artery/arteriole, capillary, venule and vein. Reproduced with permission from Wiley.[11]

Classification

In contrast to a nomenclature, diagnostic criteria and classification criteria provide guidance in the diagnosis and classification of diseases and disease subtypes.[15] Diagnostic criteria are a set of signs, symptoms, and tests with the purpose of accurately identifying patients in routine clinical practice (i.e., to distinguish vasculitis from non-vasculitis). Classification criteria on the other hand are standardised definitions for homogenous case enrolment in research (i.e., require a diagnosis of vasculitis and are used to differentiate types of vasculitis). Diagnostic criteria require high sensitivity and specificity and are notoriously hard to develop.

To date, no diagnostic criteria for AAV have been established. Classification criteria for the systemic vasculitides, however, have been widely available since first developed in 1952 by Pearl M. Zeek.[16] Over the latter half of the 20th century several alternative classifications emerged, gaining limited traction.[17, 18, 19, 20, 21]

In 1990, the ACR proposed criteria for seven forms of vasculitis, including Wegener's granulomatosis, and Churg-Strauss syndrome, but not microscopic polyangiitis.[22, 23, 24] Being developed before the wide-spread use of ANCA-testing in clinical practice, ANCAs are not included. Despite this, the ACR 1990 criteria have been widely used for case definition in AAV research.

In 2007, Watts and colleagues introduced the European medicines agency (EMA) algorithm as a consensus application of the ACR 1990 criteria and the CHCC 1994 nomenclature for uniform classification of the AAVs, including MPA.[25] To avoid double classification, the algorithm provides a stepwise classification system based on the specificity of the original classification criteria. The algorithm has been validated, extensively used in epidemiological research, and updated to incorporate the CHCC 2012 nomenclature.[25, 26 27, 28 29]

The latest classification criteria, and the first to incorporate ANCA and modern-day medical imaging, are the 2022 ACR/EULAR criteria (Table 1).[30, 31, 32] However, concerns have been raised that the new classification criteria contradict the definitions proposed in the CHCC 2012, allowing for granulomas in MPA. As the criteria does not suggest a step-wise classification, the classification of the same individual to multiple disease sub-classes is permitted.[33]

Common for all classification criteria in vasculitis are that they have been developed based on the notion of pre-defined distinct diseases or disease subtypes, based on clinical features, serological and histological findings. However, in recent years a re-classification of GPA and MPA, based on the type of ANCA pattern has been proposed, cited to be supported by stronger genetic association, more accurate prediction of clinical outcome and response to treatment.[34, 35, 36, 37, 38] EGPA on the other hand may be better reflected as either ANCA positive or ANCA negative EGPA.[39] These serology-based classifications have gained traction, with ANCA-positivity often required for inclusion in clinical trials and subsequently used for stratification.[40] However, an ANCA-based classification fails to address the percentage of patients previously classified with ANCA negative GPA or MPA.[27] Supported by a data-driven subclassification based on clinical features and serology using unsupervised machine learning some authors propose an extended subclassification of GPA/MPA based on the severity of disease and the ANCA pattern.[14 41]

Table 1. The 2022 American College of Rheumatology and European Alliance of Associations for Rheumatology Classification Criteria of the ANCA-associated vasculitides

These classification criteria should be applied to classify a patient when a diagnosis of small- or medium-vessel vasculitis has been made. Alternative diagnoses mimicking vasculitis should be excluded prior to applying the criteria.

Granulomatosis with polyangiitis	
Clinical criteria	
Nasal involvement: bloody discharge, ulcers, crusting, congestion, blockage, or septal defect/performation	+3
Cartilaginous involvement (inflammation of ear or nose cartilage, hoarse voice or stridor, endobronchial involvement, or sadle nose deformity	+2
Conductive or sensorineural hearing loss	+1
Laboratory, imaging, and biopsy criteria	
Postive test for cytoplasmic antineutrophil cytoplasmic antibodies (cANCA) or antiproteinase 3 (anti-PR3) antibodies	+5
Pulmonary nodules, mass, or cavitation on chest imaging	+2
Granuloma, extravascular granulomatous inflammation, or giant cells on biopsy	+2
Inflammation, consolidation, or effusion of the nasal/paranasal sinuses, or mastoiditis on imaging	+1
Pauci-immune glomerulonephritis on biopsy	+1
Positive test for perinuclear antineutrophil cytoplasmic antibodies (pANCA) or antimyeloperoxidase (anti-MPO) antibodies	-1
Blood eosinophil count ≥ 1 $x10^9$/liter	-4
Sum the score: ≥ 5 needed for classification of granulomatosis with polyangiitis	
Microscopic polyangiitis	
Clinical criteria	
Nasal involvement: bloody discharge, ulcers, crusting, congestion, blockage, or septal defect/performation	-3
Laboratory, imaging, and biopsy criteria	
Positive test for perinuclear antineutrophil cytoplasmic antibodies (pANCA) or antimyeloperoxidase (anti-MPO) antibodies	+6
Fibrosis or interstitial lung disease on chest imaging	+3
Pauci-immune glomerulonephritis on biopsy	+3
Postive test for cytoplasmic antineutrophil cytoplasmic antibodies (cANCA) or antiproteinase 3 (anti-PR3) antibodies	-1
Blood eosinophil count ≥ 1 $x10^9$/liter	-4
Sum the score: ≥ 5 needed for classification of microscopic polyangiitis	
Eosinophilic granulomatosis with polyangiitis	
Clinical criteria	
Obstructive airway disease	+3
Nasal polyps	+3
Mononeuritis multiplex	+1
Laboratory, imaging, and biopsy criteria	
Blood eosinophil count ≥ 1 $x10^9$/liter	+5
Extravascular eosinophilic-predominant inflammation on biopsy	+2
Postive test for cytoplasmic antineutrophil cytoplasmic antibodies (cANCA) or antiproteinase 3 (anti-PR3) antibodies	-3
Haematuria	-1
Sum the score: ≥ 6 needed for classification of eosinophilic granulomatosis with polyangiitis	

Modified from [30, 31, 32]

The anti-neutrophil cytoplasmic antibodies

After first being described by Davies and colleagues, van der Woude and colleagues in 1985 confirmed the presence of ANCA in systemic vasculitides and noted its potential as a novel diagnostic biomarker.[1, 42] In 1988, Falk and Jennette described two types of autoantibodies with distinct immunofluorescence patterns, one with reactivity with myeloperoxidase (MPO) producing a perinuclear immunostaining of neutrophils (pANCA), and one with no myeloperoxidase reactivity producing a diffuse cytoplasmic pattern (cANCA).[43] The diffuse cytoplasmic pattern was shortly after attributable to autoantibodies against a novel serine proteinase, proteinase 3 (PR3).[44]

The antigens

The neutrophil is the most abundant leukocyte in the blood, critical in antibacterial defence, but also modulates the inflammatory process. MPO is a peroxidase enzyme, abundantly expressed in the primary granules of neutrophils and is involved in the microbicidal reaction in the phagosome.[45] The primary function of PR3, however, is more oblique. While, expressed in the primary granules of neutrophils, PR3 is also expressed on the neutrophil surface, an expression increased by apoptosis. Here PR3 prevent the clearance of apoptotic cells, promoting non-resolving inflammation. PR3 may further disseminate throughout the body as a soluble protein, or via extracellular vesicles released from the cell membrane, contributing to systemic inflammation.[46]

The mechanisms of loss of immunological tolerance, autoantibody formation against MPO and PR3, seen as the key event in the pathogenesis of AAV, are largely unknown. However, the ANCAs present well before the onset of clinical disease.[47]

Detection

Today, two main methods are used for the detection ANCAs, indirect immunofluorescence (IIF) and immunoassay. Following a 2017 consensus statement, immunoassay is the preferred screening method of the two for the diagnosis of AAV, due to an improved specificity of immunoassays and the large variability between IIF methods.[48, 49, 50] This was a revision of a previous consensus, that IIF was to be used as a primary screening, with samples containing ANCA then being tested by immunoassay.[51] A number of immunoassay methods are available, primarily enzyme-linked immunosorbent assays (ELISA), but also fluoroenzyme immunoassay, chemiluminescent assay and multiplexed flow immunoassay, as well as second (capture) and third (anchor) generation antigen presentations.[49] Due to a lack of standardisation, the individual titres of ANCA do not correspond between

antigen-specific assays, impacting assessment of ANCA titre progression within one patient and group-level comparisons when using different assays.[52]

Diagnostic capabilities

The diagnostic capabilities of ANCA have been well established, and ANCA serotype is consistently associated with different clinical presentations and outcomes. Although considerable overlap exists, PR3-ANCA is primarily associated with a clinical presentation consistent with GPA (e.g., systemic inflammatory response and upper and lower respiratory tract disease), while MPO-ANCA is associated with symptoms of MPA (e.g., kidney and lung involvement).

However, despite the name, ANCAs are not a mandatory feature in AAV. Approximately, 5-10% of patients appear to be ANCA negative.[27] ANCA negativity is most common in EGPA (approximately 55-65%) and less common in GPA (approximately 5%) and MPA (approximately 5-10%).[53] The clinicopathologic findings and prognosis of patients with ANCA negative AAV are similar to those of ANCA-positive patients.[54] Given the evidence of the key role of ANCA in the pathogenesis of AAV, it has been proposed that conventional assays fail to detect some epitopes of ANCA or the presence of other autoantigens than MPO or PR3. Roth and colleagues have shown that seronegative patients may have an MPO-ANCA epitope masked in serum by a ceruloplasmin fragment, while several other studies have shown other autoantigens, notably human lysosome-associated membrane protein 2 and pentraxin-3 in cases of seronegative AAV.[55, 56, 57, 58, 59, 60]

In addition to not being mandatory, ANCAs are not specific to AAV, especially when analysed with immunofluorescence.[61] A cytoplasmic or perinuclear ANCA pattern can be seen in inflammatory bowel disease, malignancies, and infections (especially endocarditis), all important differential diagnoses to AAV.[48] ANCAs can even be present in low titres in healthy individuals, but may then have different epitope specificities.[55, 62] Additionally, a cytoplasmic or perinuclear ANCA pattern can be caused by reactivity not only to MPO or PR3 but to other antigens, such as lactoferrin (potentially identified already in 1959), elastase and bactericidal permeability-increasing protein.[63, 64, 65] Dual positivity for PR3- and MPO-ANCA may also occur and is frequently associated with drug-induced vasculitis.[66]

Despite ANCA negative disease and the existence of naturally occurring ANCA in healthy individuals, the evidence for the loss of immunological tolerance to MPO and PR3 and the development of autoantibodies as key events in the pathogenesis of AAV is compelling. The most compelling evidence being that of the in vivo pathogenicity of anti-MPO and anti-PR3 in murine models of vasculitis.[67, 68]

Pathogenesis

AAV is characterised by microvascular endothelial inflammation leading to tissue damage and loss of function. The pathogenic pathways to this are complex and involve both the innate and adaptive immune system (Figure 2). The hypothesised mechanism is in short that genetic, epigenetic, and environmental factors trigger loss of immunological tolerance to MPO or PR3 resulting the development of autoantibodies, ANCAs. ANCAs activate neutrophils, which locate to susceptible microvasculature and induce damage, resulting in further release of the autoantigen, mediating further immunological activation and injury.

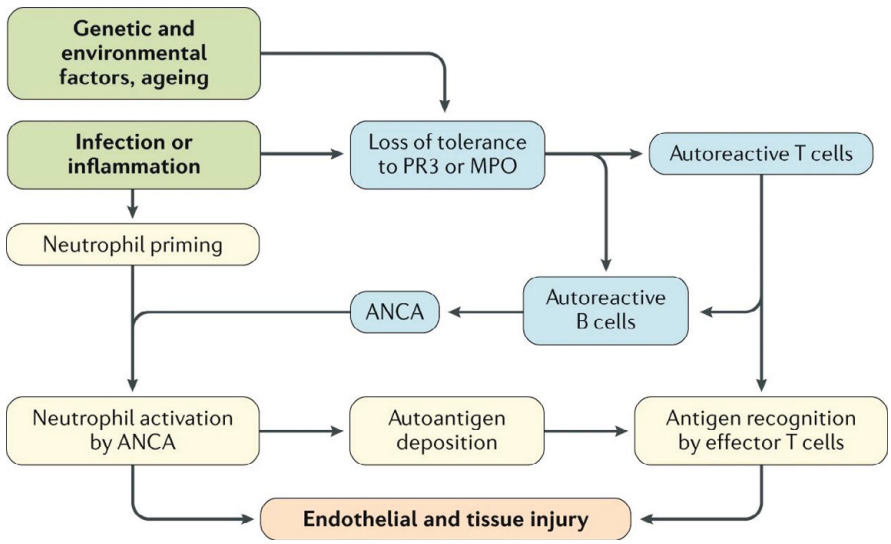


Figure 2. Simplified schematic of the hypothesised pathogenesis of AAV.
Risk factors (green) leading to autoreactivity (blue) and an effector phase (yellow) ultimately resulting in endothelial damage and tissue injury (orange). Reproduced with permission from Springer Nature.[53]

Loss of tolerance

Immunotolerance, the unresponsiveness of the immune system to substances or tissues that have the capacity to induce an immune response is essential. The breakdown of self-tolerance may result in autoimmune disorders while dysregulated responses to foreign antigens can lead to hypersensitivity and allergic disease. Immunotolerance is achieved through both central and peripheral mechanisms. Defects in both are present in AAV.

Establishing central tolerance, self-reactive T and B cells are eliminated during their development in the thymus and bone marrow, respectively. This process is evidently imperfect as ANCA autoantibodies are present in healthy individuals.[62] Tan and colleagues, showed how mice depleted of the autoimmune regulator transcription factor (naturally promoting thymic MPO expression and enhance central deletion of autoreactive T cells), produce more autoreactive anti-MPO T cells, higher ANCA titres and glomerular injury when immunised with MPO.[69]

However, peripheral mechanisms are also of importance. Peripheral tolerance ensures that self-reactive immune cells that escape central tolerance mechanisms remain inactive. In the same study as described above, Tan and colleagues, further demonstrated how depletion of peripheral regulatory T cells also led to the production of more autoreactive anti-MPO T cells, higher ANCA titres and glomerular injury.[69] The potential role of peripheral mechanisms in AAV, through regulatory T and B cells, have also been demonstrated in human studies.[70, 71, 72]

Neutrophil activation

The suggested key consequence of loss of immunotolerance is the production of ANCAs and the following activation of neutrophils. ANCAs bind to autoantigens and activate neutrophils by altering adhesion molecule expression, changing cell structure rigidity, and stimulating the formation of extracellular fibre networks (NETs) composed of neutrophil DNA and proteins like PR3 and MPO.[73, 74 75 76, 77] Importantly they induce damage through the production of reactive oxygen species.[78] The activated neutrophils are recruited to vulnerable microvascular beds through the action of adhesion molecules and chemokines, where they induce injury.

This process is exacerbated through priming of the neutrophils through pro-inflammatory cytokines, resulting in the display of the autoantigens on the cell surface. Central in the priming of neutrophils and the pathogenesis of AAV is also the complement system, and in particular the cell surface interaction of C5a and C5a-receptor. Of the three pathways of complement activation (i.e., the classical, the lectin and alternative pathway) evidence suggest the alternative pathway as the main contributor in AAV. The impact of the complement system is not limited to neutrophil priming but may stimulate chemotaxis, neutrophil activation and further bridge inflammation and coagulation.[79]

Tissue injury

Why disease manifestations have a predilection for certain microvasculatures and organs is unknown. The disease manifesting primarily in small arteries have been suggested to be partly explained by increased membrane rigidity of ANCA-activated neutrophils, leading to their entrapment in smaller capillaries.[75] However,

vascular beds, both within and between different organs, vary in morphology and function, and may influence immunological cell proliferation, migration, transendothelial passage and injury-response patterns.[80] The end result, necrotising injury with disrupted blood vessel walls, haemorrhage, release of plasma proteins into the vascular and perivascular tissue, coagulation, resulting in fibrinoid necrosis, seems to be similar across all affected organ systems.

In addition to features of necrotising vasculitis, in GPA, extravascular granulomatous inflammation of the upper and lower respiratory tract is a key manifestation. The mechanisms involved in granuloma formation in AAV are incompletely understood but may be predisposed by inflammation, infection, and dysbiosis of the upper airways.[81]

Cellular immunity

As been alluded to, not only innate and humoral immunity contributes to the pathogenesis. ANCA-activated neutrophils release their autoantigen, making it available for recognition by effector T cells. In murine models of anti-MPO glomerulonephritis a role for both $CD4^+$ and $CD8^+$ T cells in tissue injury has been shown.[82, 83] Furthermore, $CD4^+$ T cells promote the production of ANCA, and $CD8^+$ T cell gene expression is associated with clinical disease and disease outcome.[53, 84] The cellular immunity may have further implications for the prognosis of vasculitis, with signatures associated with T-cell exhaustion (e.g., loss of effector functions) being associated with favourable outcome in autoimmunity, but unfavourable outcome in infection.[85]

Pathology of chronicity and disease relapse

Mechanisms involved in disease induction are likely to be relevant also in disease chronicity and relapse. It has been suggested that chronic nasal colonisation of Staphylococcus aureus may facilitate relapse.[86] Furthermore, a disease manifestation beyond necrotising vasculitis and granulomatosis that has gained increased interest in recent years is pulmonary fibrosis. As opposed to granuloma, common in anti-PR3 positive disease, pulmonary fibrosis is present mainly in anti-MPO disease.[87] The aetiology, is however, unclear, and interstitial lung disease can be present both as a prodrome and a complication of manifest AAV. Recurrent, sub-clinical intra-alveolar haemorrhage has been suggested in the pathogenesis, but also a direct role of MPO-ANCA through stimulation of fibroblast proliferation and release of fibrosis inducing proteolytic enzymes.[87]

Differences in eosinophilic granulomatosis with polyangiitis

The implication of eosinophil dysfunction in the pathogenesis is unique for EGPA. However, the differences between the ANCA-positive and ANCA-negative forms of disease presentation, and how EGPA differs from other types of AAV remain poorly understood. Supported by differences in genetic association and disease presentation, current evidence suggests distinct differences in the pathogenesis of ANCA-negative and ANCA-positive EGPA, but the phenotypic expression is not dichotomous.[39, 88, 89] In ANCA-negative disease, dysfunction of the mucosal barrier is implied by genetic associations, while ANCA-positive disease displays associations with mutations of the human leukocyte antigen (HLA) region, indicative of an autoimmune disorder.[39] The role of the eosinophil has been further implicated through the presence of eosinophils in the vasculitis lesions of EGPA, and convincingly demonstrated through the clinical effect of interleukin-5 (IL-5) and IL-5 receptor blockade (IL-5 promote eosinophil proliferation and function) in symptom control.[90, 91]

Aetiology

As apparent by the previous section, the exact aetiology of AAV remains unknown. However, current evidence suggests an interplay of genetic, epigenetic, and environmental factors.

Ageing

AAV can present at any period of life, but disease incidence increases progressively with age.[27] Although at first glance paradoxical, this is not uncommon for autoimmune diseases. With complex pathways to loss of immunotolerance, the failure of immunological tolerance checkpoints could accumulate over a lifetime, but also be induced by declining immune competence. Despite being an immune system wide process, the accumulation of effector memory T cells, seen in AAV, and seen in the ageing immune system, have been implicated as a facilitator of chronic inflammation in elders.[92] Interestingly, incidence not only increases with age, but the phenotypic expression also differs between age groups.[93]

Genetics

Although several genes associated with susceptibility or resistance to AAV have been identified, the most frequent and strongest ones are found in the HLA region, encoding the MHC II complex.[94] GPA (and anti-PR3 positive disease) is most strongly associated with the HLA-DP region, while MPA (and anti-MPO positive

disease) with the HLA-DQ.[35] Similarly, the HLA DQ region is associated with EGPA, but only in anti-MPO positive disease.[39] Associations between the MHC and autoimmune diseases have been known since the 1970s and remain the strongest genetic risk factors in many autoimmune diseases.[95, 96]

However, estimating the heritability (how much variation in phenotype or disease that can be explained by genetic variants) is challenging. The reported variance attributable to MHC alleles for autoimmune diseases varies from 2 – 30 %, with 6% reported for EGPA.[39, 96] Differences in frequency of alleles strongly associated with MPA have been seen among East Asian, and European/North American populations, reflecting the epidemiological differences seen in disease pattern, again highlighting the importance of the MHC region in disease presentation.[97]

Yet, genetical associations are not limited to the HLA region. Variation in the PRTN3 region, encoding PR3 is associated with GPA (and anti-PR3 positive disease).[35] Similarly, the SERPINA1 gene, encoding α1-antitrypsin (an inhibitor of PR3) is associated with GPA (and anti-PR3 positive disease).[35] Clinically, α1-antitrypsin deficiency is known to be associated with systemic vasculitis.[98] When comparing anti-PR3 positive patients with and without heterozygosity for α1-antitrypsin deficiency, patients with heterozygosity exhibited more extensive organ involvement and poorer prognosis.[99] Both the genetic variants in PRTN3 and SERPINA1 genes favours an increased expression and higher circulating levels of PR3, facilitating the synthesis of anti-PR3 autoantibodies.[100]

Other genes associated with AAV are related to T cell regulation, PTPN22 and CTLA4, and to endothelial cell functions and T and B cell homeostasis, BACH2.[101] In ANCA-negative EGPA, gene associations are seen related to eosinophil inflammation and respiratory barrier function, IRF1/IL5 and GPA33, respectively.[101]

Familial aggregation is a key indicator of complex genetic diseases. Both aggregation of the same autoimmune condition (familial autoimmune disease) and aggregation of diverse autoimmune diseases (familial autoimmunity) are common.[102] In AAV, however, reports of familial aggregation, are scarce, and limited to case series.[103] Knight and colleagues, have in two population-based studies investigated both familial autoimmune disease and familial autoimmunity in GPA. These indicate an increased risk of GPA in relatives with the disease, but a low risk in absolute terms.[104] Similarly, there is a moderate increase in risk of autoimmune and autoinflammatory disorders in relatives of patients with GPA.[105]

Epigenetics and post-translational modifications

In addition to genetic, epigenetic processes regulate gene expression and may influence disease development and phenotype. Methylation of DNA and histones are two such mechanisms identified in AAV, influencing the expression of PRTN3

and MPO, the genes encoding the ANCA-autoantigens.[106] Interestingly, the DNA methylation levels at these sites were lower in patients with active disease compared to patients in remission, and increased methylation of PRTN3 was protective of disease relapse.[106]

Post-translational modifications also play a role outside of gene expression, especially through the regulation of the ANCA effector functions. Glycosylation, the formation of glycoconjugates (the addition of sugars to proteins and lipids), have wide extent physiological and pathological implications, including modulation of the inflammatory response.[107, 108, 109] In AAV, both hypo- and hyperglycosylation of the antibodies are seen, but its correlation with disease activity conflicting.[110, 111, 112]

Likewise, post-translational modification of the PR3 and MPO antigens also occur and may alter their antigenicity, and changes in antibody avidity has been observed in relapsing disease.[113, 114] Interestingly, glycosylation has implications beyond the direct post-translational modifications of immune system components. Interaction between endothelial cells and leukocytes, crucial in recruitment to the tissue, is regulated by adhesion molecules affected by endothelial cell-surface glycosylation, suggesting a contribution of post-translational modification in inflammatory vascular disease.[115, 116]

Environment

Several microbial, occupational, and geoepidemiological exposures have been implicated in the aetiology and activity of disease in AAV. Notable is also, as mentioned, that the incidence of disease increases with age, implying the potential for accumulation of damage related to environmental risk factors.[27]

Microbial agents

There are several suggested mechanisms for how microbes may induce autoimmunity and vasculitis: molecular mimicry (similarities between foreign and self-peptides), epitope spreading (an immune response against a specific antigen may expand to include other epitopes over time), bystander activation (immunological activation without an antigen), and cryptic antigen (e.g., tissue damage may reveal previously immunologically "hidden" antigens).[117]

Although several bacteria and fungi have been implicated in the disease aetiology, the most convincing results are regarding chronic nasal colonisation of Staphylococcus aureus.[118] Patients with GPA have a higher rate of chronic nasal colonisation, which is also associated with an increased risk of relapsing disease.[86] This is not seen in MPA.[119] Related to the nasal colonisation of Staphylococcus aureus in GPA, several studies have investigated the wider nasal microbiome.[120, 121, 122] Dysbiosis is associated with active disease, precede clinical disease onset, and normalise in remission.[122] Additionally, the impact of dysbiosis of the gut

microbiome in the development and presence of autoimmune disease have raised much interest in recent years.[123, 124] In AAV, dysbiosis of the intestinal microbiota may be related to disease activity.[125] The evidence for the influence of infections and microbial agents (i.e., Staphylococcus aureus) on disease aetiology is further strengthened by evidence of a protective effect of trimethoprim-sulfamethoxazole on disease relapse.[126, 127, 128]

Viral pathogens are associated with autoimmune disease, importantly Epstein-Barr virus in multiple sclerosis, rheumatoid arthritis, and systemic lupus erythematosus.[129] However, evidence to support a role for a viral pathogen in AAV is scarce.[118] During the Covid-19 pandemic case-reports were presented with onset of disease following Sars-Cov-2 infection, and Covid-19 is known to exhibit features of and mimic vasculitis.[130, 131, 132] Similarly, rare case-reports have been made regarding onset of AAV following Covid-19 vaccination.[133] Reports on influenza vaccination induced autoimmune disease onset and relapse are relatively common but evidence for this association have not been found in larger scale studies.[134, 135]

Occupational exposures

Although, a Swedish case-control study found no evidence for an association between occupation and the development of GPA, several occupational or industrially generated exposures have been implicated in the aetiology of AAV.[136] The most prevalently described being silica and environmental dusts, exposure of which, with available evidence, are considered risk factors.[137] Relatedly, there are reports of an increased incidence following natural disasters with increased airborne particulate matter, although the reports are conflicting.[138, 139, 140] Farming, via the exposure to inhaled antigens have also been described as associated with AAV.[141]

Other occupational exposures implicated (but with low-grade evidence) are mercury, industrial solvents, and carbon monoxide. Interestingly, smoking seems to have a protective effect on the development of vasculitis, but results are conflicting.[118]

Geoepidemiology

Several environmental risk factors related to geoepidemiology have been suggested in AAV. Seasonality or temporal clustering of disease have been repeatedly reported. Results are conflicting but reports of increased incidence during wintertime are frequent.[118] In addition to this there are conflicting studies on cyclic disease occurrence, with a periodicity of none to one to eight year cycles.[142, 143] However, studies on the seasonality of disease are complicated by the highly variable time from symptom onset to diagnosis and the largely unknown disease prodrome.

Relatedly, a latitudinal gradient has been noted in AAV. Some evidence suggests GPA being more common in the North of the Northern hemisphere, and MPA in the South, with a largely reciprocal finding in the Southern hemisphere.[144, 145] It has been hypothesised that this is caused by sun exposure, and following levels of 1,25-dihydrohydroxyvitamin D3 ($1,25(OH)_2$ D_3), with known immunomodulary effects.[146] However, levels of $1,25(OH)_2$ D_3 are hard to measure, interindividual variance within the same latitude large and a clear latitudinal gradient lacking.[147] Despite this, studies have shown lower serum 25-OH vitamin D levels (the measurable inactive metabolite of $1,25(OH)_2$ D_3) in AAV patients compared to healthy controls, with levels falling during relapse, and similar results, related to ambient ultraviolet radiation in an epidemiological study.[148, 149, 150]

Drugs

In addition to environmental exposures various drugs and medications have been associated with development of AAV or disease-like symptoms. The anti-thyroid drug propylthiouracil, and to some extent other antithyroid medications, are known to cause MPO-ANCA production and is associated with the development of vasculitis.[151, 152] Other drugs implicated are minocycline (an antibiotic) and hydralazine (an antihypertensive).[153] The use of cocaine is also associated with AAV, both through mimicking vasculitis (sinonasal destruction following snorting) and through the anti-helminthic agent levamisole (commonly used in cocaine adulteration).[154] Clinically, drug-induced vasculitis often presents with less severe symptomatology, more prominent skin involvement, and concurrent anti-MPO and anti-PR3 positivity. Symptoms often subside with termination of the drug, but immunosuppressive treatment may be required.[153]

Epidemiology

The study of the epidemiology of AAV was long was hampered by the lack of universally accepted classification criteria and the rarity of the disease. Since the 1990s considerable progress has been made in understanding the incidence ("How many people developed the disease during the specified time?") and prevalence ("How many people have the disease during the specified time?") of AAV. Studies have, however, been largely limited to the Global North.[155]

Incidence

Although disputed, an increase in the incidence over time has been reported. While it is possible that the disease truly is becoming more and more common, other possible explanations include the change and development of classification criteria, widely available ANCA serology testing, and increased healthcare provider

awareness.[27, 155] Interestingly, the increase in incidence is largely driven by MPA. This might reflect a demographic change in the Global North, with MPA being a disease subtype predominant in the elderly. The classification of MPA has also changed over time, with widely used classification criteria not available until 2007, and the disease being recognised by the World Health Organisation (WHO) in the International Classification of Diseases (ICD) as late as 2004.[25, 156]

The mean annual incidence of AAV in southern Sweden during the period 1997 to 2019 was estimated to 30.1 per million adults (95% CI 27.0-33.1), (GPA: 15.4 [95% CI 13.3-17.6], MPA: 12.8 [95% CI 10.8-14.8], EGPA: 1.8 [95% CI 1.1-2.6]), with no change over time.[27] The distribution of patient cases between the three disease subtypes are largely consistent among European and North American cohorts, with some studies, as previously discussed, indicating a latitudinal gradient.[155] However, in East Asia, MPA is the most common disease presentation and GPA, EGPA and PR3-positive disease exceedingly rare.[157]

In addition to regional differences, gender differences have been observed. Most studies in AAV demonstrate a slight predominance in men, as opposed to most autoimmune disorders.[155, 158] Although conflicting results, current evidence suggests GPA being more common in men, and MPA in women. Disparities in study results have been suggested to reflect demographic differences in age distribution between populations.[155] AAV show a clear increase in incidence with age. Rathmann and colleagues, show a peak incidence in individuals 74 years and older, results largely replicated in comparable studies.[27, 155] A difference between the disease subtypes is also present, with MPA (and MPO-positivity) occurring more frequently in older individuals, than GPA (and PR3-positvity). AAV is however, not a disease limited to elderly, and can develop at any time in life. Although studies are scarce, disease may, but rarely, develop in children. In 2018 the annual incidence of AAV in southern Sweden was estimated to 3.2 per million children (95% CI 1.1-5.4), (GPA: 1.4 [95% CI 0-2.8], (MPA: 1.4 [95% CI 0-2.8], EGPA 0.4 [95% CI 0-1.1]).[159] In a case-report, passive placental transfer from mother to neonate with subsequent disease development in the newborn, has been reported. This is also the first report demonstrating the immunopathogenic potential of MPO-ANCA in humans.[160]

Prevalence

While studies of the prevalence of AAV are scarce compared to incidence studies, most reports indicate an increasing prevalence.[155] As AAV is a life-long condition, without any known cure, the increasing prevalence might reflect increasing incidence but also improved survival. The point-prevalence in 2020, in southern Sweden, was estimated to 428.4 (95% CI 350.7-506.0) per million adults (GPA: 241.6 [95% CI 183.3-299.9], MPA: 150.1 [95% CI 104.2-196.1], EGPA: 36.6 [95% CI 13.9-59.3]). This is a significant increase compared to data from the same

population in 2003, 353.6 (95% CI 275.6-431.6) per million adults.[27] With an estimated prevalence of approximately 400 per million adults, AAV is to be regarded a rare disease.[161]

Health-economics

Despite being a rare disease, AAV has substantial health-care economic effects. In addition to the considerable morbidity, placing a heavy burden on the individual, the elevated all-cause healthcare resource use compared to the background population, and following costs, is a health-economic burden.[162, 163, 164, 165] The age, disease and, treatment related comorbidities are complex, but the increased expenditures are highly attributable to medications and hospitalisation.[163] Healthcare resource use is correlated with the use of glucocorticoids.[163] This may be attributable to both morbidities associated with disease activity, but also the treatment side-effects. This highlights the fine line, of over-treating with the risk of treatment related complications (e.g., infections) and under-treating with the risk of disease relapse.

Signs and symptoms

The disease presentations of AAV are complex and heterogenous. While the different types of AAV may share clinical features of non-specific systemic inflammation, such weight loss, arthralgia, myalgia, and malaise, the organ specific features are seen in different frequencies in the different types of disease (although considerable overlap exists) (Figure 3).

The key feature is that of necrotising vasculitis (of small and medium sized vessels) with few or no immune deposits (i.e., pauci-immune) affecting any organ system of the body. In GPA and EGPA, but not MPA granulomatous inflammation may be present.

Granulomatosis with polyangiitis

The hallmark features of GPA are necrotising granulomatous inflammation and vasculitis involving the upper and lower respiratory tract. Consequently, it often presents with ear-nose-throat symptoms (sinusitis, nosebleed, nasal crusting, chronic otitis-media, and destruction of the nasal cartilage). Lung involvement is also commonly seen in with symptoms such as pulmonary nodules and cavitation, and pulmonary capillaritis with lung haemorrhage, manifesting as shortness of breath, cough and potentially haemoptysis. Necrotising glomerulonephritis and kidney symptomatology is frequent, but less so than in MPA. Any organ system

may be affected, but eye involvement (e.g., peripheral ulcerative keratitis, scleritis, uveitis, and retinal vascular manifestations) is more often seen in comparison to the other types of AAV.

GPA is most often a systemic disease but may present with localised ear-nose-throat symptoms, posing considerable diagnostic difficulties. Most patients have positive ANCA serology (95%), with 65-75% being PR3-positive, and MPO-positivity seen in 20-30%.[53]

Microscopic polyangiitis

Necrotising glomerulonephritis is very common in MPA and seen in almost all patients with symptoms such as haematuria, proteinuria, and hypertension. Lung involvement with pulmonary capillaritis is also frequent. Symptoms from the lung (diffuse alveolar haemorrhage) and kidney (glomerulonephritis), may co-present as the pulmo-renal syndrome.

While MPA is most often a systemic disease, kidney-limited disease presentation is relatively frequent. Like in GPA, this may pose diagnostic challenges. Non-vascular inflammation and granulomatous inflammation are per definition absent in MPA.

One disease presentation more frequently seen in MPA compared to the other AAVs is interstitial lung disease. Most common is a radiographic or histopathological pattern of usual interstitial pneumonia (UIP) that most often presents before or concomitantly with the vasculitis diagnosis.[87] However, MPO-positive interstitial lung disease without other signs of vasculitis may occur. In a retrospective cohort study including 34 patients with idiopathic pulmonary fibrosis and MPO-ANCA positivity, 28% developed clinical symptoms of systemic vasculitis.[166] Seroconversion to ANCA-positivity in patients with seronegative idiopathic pulmonary fibrosis may also occur.[167, 168] MPA is associated with MPO-ANCA (55-65%), less so PR3-ANCA (20-30%) and is infrequently (5-10%) ANCA negative.[53]

Eosinophilic granulomatosis with polyangiitis

Asthma and eosinophilia in blood and tissue in association with necrotising granulomatous inflammation or vasculitis are hallmark features of EGPA. The symptomatology can be localised to the upper (e.g., nasal polyposis and other non-destructive sinonasal symptoms) and lower respiratory tract (e.g., nodular lung disease), but is often systemic. EGPA frequently involves the skin (e.g., haemorrhagic lesions and nodules) and the peripheral nervous system (e.g., peripheral neuropathy and mononeuritis multiplex). Involvement of the heart (e.g., cardiomyopathy), and the gastrointestinal system (e.g., eosinophilic gastroenteritis) are also common, and poor prognostic factors.[88]

In the absence of histologic proof of necrotising vasculitis or eosinophil-rich granulomatous inflammation there are considerable challenges in distinguishing EGPA from primary and secondary hypereosinophilia, or organ-restricted hypereosinophilic presentation (e.g., eosinophilic pneumonia and eosinophilic myocarditis). These difficulties are further exacerbated by EGPA frequently being ANCA negative (55-65%). ANCA-positive EGPA is almost exclusively of MPO-type and PR3-positive EGPA is very rare. Interestingly, the phenotype differs between ANCA-negative and ANCA-positive disease. ANCA-negative EGPA is associated with cardiomyopathy, lung, and gastrointestinal involvement. ANCA-positive disease on the other hand is associated skin disease, peripheral neuropathy and mononeuritis, and kidney involvement.[88]

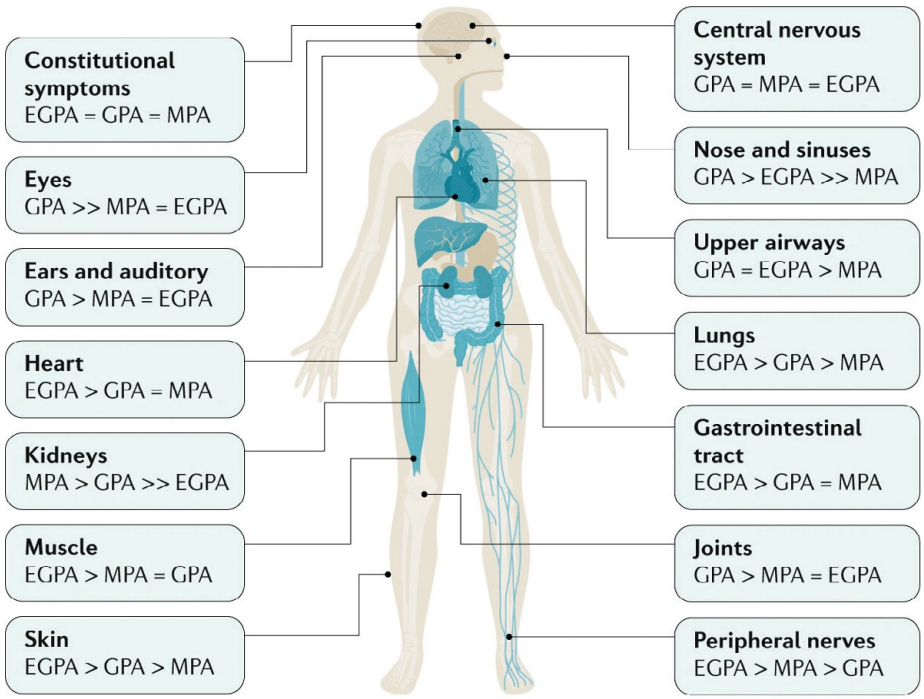


Figure 3. The approximate relative frequencies of organ system involvement in AAV
Reproduced with permission from Springer Nature.[53]

Diagnosis

There are no diagnostic criteria, and no diagnostic tests for AAV. Consequently, making a diagnosis is challenging and initial misdiagnoses as another systemic rheumatic disease, infection, and malignancy are common.[169] Some differential

diagnoses, such as infective endocarditis not only share clinical features but may also present with a positive ANCA serology.

The plethora of disease mimics, the rarity of the condition, and the lack of pathognomonic features often leads to a significant diagnostic delay. Frequent healthcare encounters prior to a diagnosis are common.[170] In a 2021 survey by the Vasculitis Patient-Powered Research Network 73% of patients were initially misdiagnosed, with a median of five different misdiagnoses. The median time from symptom onset to a correct diagnosis was seven months, and 82% of patients reported that the diagnostic delay had negative health consequences.[171] The diagnostic delay in AAV is associated with considerable risk of morbidity and mortality, but also increased healthcare costs.[172, 173] However, the diagnostic delay has been considerably reduced following the widespread use of ANCA testing.[174]

The diagnosis of AAV is a clinical one, supported by serology and histology. Investigations to support a diagnosis and refute differential diagnoses may differ based on symptomatology.

ANCA-testing, in cases of high pre-test probability, dramatically improves the diagnostic certainty, but both false negatives (5-10% of AAV is ANCA negative) and false positives (ANCA are seen in healthy individuals and be associated with differential diagnoses) still occur. However, a higher titre of ANCA may help to discriminate AAV from disease mimics.[175]

Assessing disease activity

Disease activity refers to the potentially reversible aspects of a disease (as opposed to damage). In AAV the main validated tool for assessing disease activity is the Birmingham Vasculitis Activity Score (BVAS), first published by Luqmani and colleagues in 1994, now in its third version. [176, 177] The BVAS form include signs and symptoms from nine organ-systems attributable to active vasculitis. The form provides a numeric score from 0 to 63, shown to correlate with prognosis, and other activity measures. BVAS is recommended for use in clinical trials and clinical practice, to assess disease activity at symptom onset, and for the assessment of remission and relapse.

Disease subtype specific scoring systems, such as the BVAS for Wegener's granulomatosis (i.e., GPA) (BVAS/WG) have been developed and used in clinical trials.[178] Other activity assessment tools include the physician's global assessment (PGA), the disease extent index (DEI), and the prognostic Five-Factor Scores.

Furthermore, a dichotomous division of organ/life-threatening and non-organ/life-threatening disease manifestations often occur in the literature. While these concepts (previously described as severe and non-severe AAV) are ill-defined, the BVAS form provides some guidance through the "major" and "minor" items. This may

help in deciding management strategies or define the severity of the disease manifestation at onset or relapse.[179] Similarly, the Five Factor Score is used for management decision guidance in EGPA.[180]

Treatment and management

The treatment of AAV has undergone considerable changes since the introduction of immunosuppressants in the mid 20th century. Since its discovery by chemists Edward Calvin Kendall and Tadeusz Reichstein, and first clinical use by physician Philip Showalter Hench in 1949, cortisone has been a mainstay in the treatment of systemic rheumatic disease. However, severe, debilitating, and even mortal treatment side-effects caused by cortisone toxicity were common.[181] In 1954 the first report of the use of a cytotoxic agent in AAV was published, describing symptom relief following intravenous nitrogen mustard in a patient with GPA (Interestingly, snake venom was also tested. Today cobra venom factor is used for complement depletion in research models).[182]

In 1979 Fauci and colleagues, presented paradigm-shifting data of 17 patients with severe systemic necrotising vasculitis treated with the cytotoxic agent cyclophosphamide (a much less toxic nitrogen mustard-derived alkylating agent).[183] Fourteen achieved complete remission. Over time, side-effects (e.g., haemorrhagic cystitis, infertility, and malignancies) of prolonged cyclophosphamide exposure became evident.[184] Despite issues of treatment side-effects and the emergence of new therapeutic alternatives, glucocorticoids and cyclophosphamide (albeit with changed routines) are still the cornerstones of the pharmacological treatment.

Today's pharmacological treatment of AAV is biphasic, induction of disease remission and following achieved remission, remission maintenance. Foundational for the advances in the treatment of AAV are several randomised clinical trials (RCTs) (Table 2).

Table 2. List of key randomised clinical trials and main results

Trial	Main results
CYCAZAREM (2003)[185]	The non-inferiority of azathioprine to cyclophosphamide in maintenance of disease remission in AAV (GPA or MPA)
NORAM (2005)[186]	The non-inferiority of methotrexate to cyclophosphamide in remission induction in non-severe AAV (GPA or MPA)
MEPEX (2007)[187]	The superiority of plasma exchange to pulsed intravenous glucocorticoids for kidney survival in severe kidney AAV (GPA or MPA)
CYCLOPS (2009)[188]	The non-inferiority of dose-sparing pulsed intravenous cyclophosphamide to daily oral cyclophosphamide for remission induction in AAV (GPA or MPA)
IMPROVE (2010)[189]	The inferiority of mycophenolate mofetil to azathioprine for maintenance of disease remission in AAV (GPA or MPA)
RAVE (2010)[190]	The non-inferiority of the B-cell depleting, anti-CD20 monoclonal antibody Rituximab to cyclophosphamide in the induction of remission in AAV (GPA or MPA)
RITUXVAS (2010)[191]	The non-inferiority of Rituximab to cyclophosphamide in the induction of remission in AAV (GPA or MPA)
MAINRITSAN (2014)[192]	The superiority of Rituximab to azathioprine in remission maintenance in AAV (GPA or MPA)
REMAIN (2017)[193]	The superiority of continued treatment (azathioprine/prednisolone) to withdrawal (at 24 months after diagnosis) for remission maintenance in AAV (GPA or MPA)
MAINRITSAN 2 (2017)[194]	No difference in rate of relapse between individually tailored (B-cell or ANCA reapperance) to fixed Rituximab (every 6 months) regimens in AAV (GPA or MPA)
MIRRA (2017)[90]	The superiority of IL-5 inhibitor Mepolizumab over placebo in relapsing and refractory EGPA
MYCYC (2019)[195]	The non-inferiority of mycophenolate mofetil to cyclophosphamide for remission induction (but higher rate of relapse) in AAV (GPA or MPA)
MAINRITSAN 3 (2020)[196]	The superiority of prolonged Rituximab to placebo in remission maintenance beyond 18 months in AAV (GPA or MPA)
PEXIVAS (2020)[197]	No effect of plasma exchange with respect to survival or kidney survival in severe AAV (GPA or MPA), and the non-inferiority of a reduced-dose glucocorticoid regimen compared to standard of care with respect to survival and kidney survival.
ADVOCATE (2021)[198]	The non-inferiority (week 26) and superiority (week 52) of remission maintenance by the complement [C5a receptor] inhibitor Avacopan to glucocorticoids in AAV (GPA or MPA)
MANDARA (2024)[91]	The non-inferiority of IL-5 receptor inhibitor Benralizumab to Mepolizumab for the induction of remission in relapsing or refractory EGPA

Current pharmacological treatment

The latest guidelines for the rheumatological management of AAV, were presented by EULAR in 2022 (Figure 4).[179] Stratifying active GPA or MPA in organ-/life-threatening and non-organ-/life-threatening disease, induction of remission is recommended with a combination of glucocorticoids and rituximab or cyclophosphamide, or a combination of glucocorticoids and rituximab, respectively.

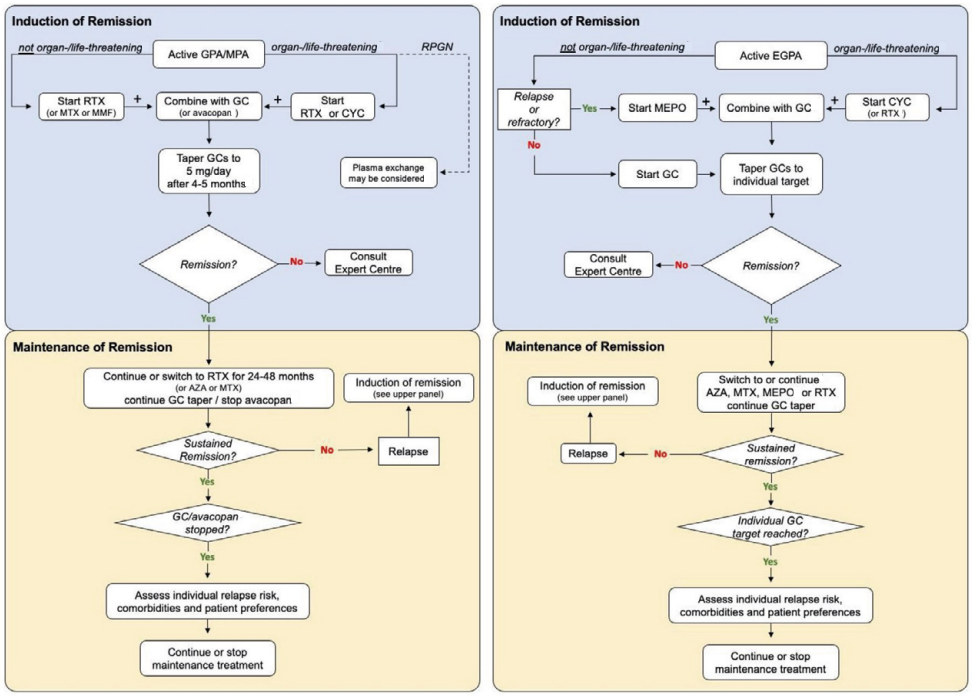


Figure 4. 2022 EULAR algorithm for the treatment of GPA/MPA (left) and EGPA (right)
RPGN = rapidly progressive glomerulonephritis, RTX = rituximab, MTX = methotrexate, MMF = mycophenolate mofetil, GC = glucocorticosteroids, CYC = cyclophosphamide, MEPO = mepolizumab, AZA = azathioprine. Reproduced from EULAR recommendations for the management of ANCA-associated vasculitis: 2022 update, Hellmich B, Sanchez-Alamo B, Schirmer JH, et al, 83:30-47, 2024, with permission from BMJ Publishing Group Ltd.[179]

Rituximab is preferred over cyclophosphamide in relapsing disease.[199] The recommended starting dose of oral glucocorticoids is 50-75 mg of prednisolone per day with a following stepwise reduction to 5 mg per day by 4-5 months.[197] Avacopan may be used in combination with rituximab or cyclophosphamide to further reduce glucocorticoid exposure.[198] Plasma exchange may be considered as part of the induction remission therapy in case of serum creatinine >300 µmol/L.[200] After remission induction maintenance of remission is recommended to be continued for two to four years or longer with rituximab.[193, 196, 201]

In EGPA, organ-/life-threatening disease is recommended to be treated with a combination of glucocorticoids and cyclophosphamide.[179] For remission induction of EGPA without organ-/life-threatening symptoms treatment with glucocorticoids are recommended, with mepolizumab recommended for remission induction in relapsing or refractory disease.[90] Remission maintenance therapy in life-threatening disease is recommended using glucocorticoids and rituximab, mepolizumab, azathioprine, or methotrexate. In the absence of organ-/life-threatening symptoms

glucocorticoids alone, or in relapsing disease, glucocorticoids in combination with mepolizumab are recommended.[179]

However, the above medications are associated with several side-effects and risk of complications and adverse events. Additional treatment or prophylaxis to address these are required and recommended.

Some considerations at induction of immunosuppressants

Immunosuppressive treatment is associated with side-effects and risks of complications. Some of the considerations at induction of immunosuppressants are:

● Fertility: Cyclophosphamide confers a risk of premature gonadal failure and sterility, and alternative treatments or fertility-preserving measures should be considered.

● Osteoporosis: Glucocorticoids increase the risk of osteoporosis (other factors that might be of relevance in AAV contributes such as for example reduced kidney function). Prophylaxis should be considered.

● Infections: Immunosuppressive treatment is associated with an increased risk of infections. The use trimethoprim-sulfamethoxazole decreases this risk and is recommended. Vaccinations are recommended.

Non-pharmacological management

The 2022 EULAR management guidelines rests on four overarching principles, highlighting shared decision-making between patient and physician, multidisciplinary management, and access to specific vasculitis expertise. This indicates a considerable change, evident in all fields of medicine, towards person-centred care and increased patient involvement. In AAV, the importance of multidisciplinary management cannot be understated. Being a heterogenous systemic disease with complex phenotypic presentation and a plethora of differential diagnoses, collaboration between medical specialties is key. However, despite recent advances in the development of standardised management guidelines, the care received is still highly varied.[202] Furthermore, concordance may be lacking, with patients not following the treatment prescription by the health care provider. While data specific for AAV is lacking, wider studies of rheumatic diseases show medication adherence (acting in accordance with the treatment regimen) and persistence (continuing the treatment for the prescribed duration) varying from 9%

to 94% and 23% to 80% respectively.[203] Structured education programs may increase disease-specific knowledge and behavioural changes.[204]

Future of treatment

While a cure for AAV remains elusive, research is ongoing in re-establishing self-tolerance in autoimmune disease.[205, 206] In addition to this, chimeric antigen receptor T-cell (CAR-T) therapies, targeting and killing B-cells have shown promising results of sustained disease remission in other systemic rheumatic diseases and murine models of AAV.[207, 208, 209] However, there are reports of secondary T-cell neoplasms, highlighting a potential mutagenesis of the genetically modified administered T-cells.[210]

While what the future holds is impossible to know, several clinical trials using both new and established therapies are currently ongoing. With complement inhibition being the latest addition to the treatment alternatives, evidence is emerging for the potential of an extended therapeutic use.[211]

Disease progression

As evident from the management guidelines, close follow-up, and vigilance, by both the patient and the healthcare system, are needed in AAV. Today, the disease-course has changed following improved pharmacological treatment. Prior to the introduction immunosuppressants the prognosis for individuals with necrotising vasculitis was abysmal, with an average survival of five months.[212] With the focus previously largely being on the immediate survival, focus today has extended to the control of a chronic relapsing disease. Akin to extended focus on person-centred care, patient-reported outcome has been increasingly important in the disease follow-up. In AAV, and other autoimmune and musculoskeletal diseases, the OMERACT (Outcome-measures in Rheumatology) initiative is driving how both patient- and disease relevant outcomes are measured in clinical trials.[213] A composite outcome measure for AAV, encompassing several clinically relevant outcomes, is under development.[214]

Patient-reported outcomes

Discrepancies in perspective of the disease, its impact, and areas of importance in management between physicians and patients are not uncommon in rheumatic diseases, AAV included.[215, 216] While only outcomes, such as survival, kidney survival and disease remission and relapse have been the primary outcome measures

of clinical trials, patient-reported outcome measures (PROMs) have been increasingly incorporated in trial protocols.[217]

Two areas consistently highlighted by patients, and affecting quality of life, are fatigue and pain. Interestingly, fatigue does not correlate with measures of disease activity or disease progression. While a patient may achieve physician-assessed clinical disease remission, debilitating fatigue often persists.[218, 219] Instead of disease activity, fatigue is often associated with anxiety, depression, and pain.[220, 221] As with fatigue, pain is also often present in patients assessed to be in stable clinical remission.[222] Diffuse musculoskeletal complaints, primarily affecting the lower limbs are common, the causes of which are largely unknown but both chronic damage and sub-clinical inflammation have been suggested.[218]

The interplay of pain, psychological distress, and fatigue and its mechanisms in AAV are largely unexplored. However, persistent fatigue and pain are prominent features of all systemic autoimmune rheumatic diseases, indicating shared mechanisms. Strong evidence supporting how fatigue and pain should be managed in AAV are lacking, but behavioural changes and exercise have been implicated in a smaller feasibility trial.[223] Furthermore, avacopan has been reported to significantly improve health-related quality of life.[224]

The consequences of an AAV diagnosis go beyond that of personal symptomatology. Decreased participation in work and social activities are frequently reported, and economic consequences not uncommon.[225, 226]

Complications and comorbidities

With improved survival, disease-related comorbidities and complications of immunosuppressive medications are increasingly identified as major adverse outcomes of AAV. These include, but are not limited to cardiovascular and thromboembolic events, infections, malignancies, and hormonal dysfunction.

Cardiovascular events

Myocardial infarction and cerebrovascular insult are leading causes of morbidity and mortality in AAV.[227, 228, 229] These cardiovascular events may occur both in the acute symptomatology of active disease and secondary to disease and treatment related changes to the cardiovascular system.

The risk of both myocardial infarction and stroke are highest at the time of diagnosis, potentially related to the heightened inflammatory state or direct vascular mechanisms.[230, 231] Long-term, accelerated atherosclerosis and diffuse endothelial dysfunction have been seen in patients with AAV and glucocorticoid treatment is associated with other metabolic cardiovascular risk factors such as diabetes and hypertension.[232, 233] Although, the relative risk is highest early in the disease course,

cardiovascular events are one the main causes of death beyond the first year after diagnosis.[229] Anti-MPO-positive patients are at higher risk of cardiovascular death.[227]

Thromboembolic events

In a population-based study from southern Sweden almost one in five patients with AAV suffered thromboembolic events (deep vein thrombosis or pulmonary embolism), with high rates seen in other observational studies and clinical trials. [234, 235, 236, 237] Like with cardiovascular events, the incidence is highest early in the disease course. The mechanisms behind the evident hypercoagulable state in AAV are not well-understood. An interplay between the immune system dysregulation and coagulation is evident. Activated neutrophils may induce hypercoagulability through both tissue-factor-expressing microparticles and NETs, in a mechanism facilitated by C5a and ANCA.[238] In addition to this, several measures of an increased state of hypercoagulability including increased platelet count, fibrinogen, prothrombin, and fibrin degradation product D-dimer are seen in active disease.[239] Interestingly, the levels of, at least some, coagulation factors remain increased also in disease remission.[240]

Infections

Infections are common in AAV and are the leading cause of mortality.[227, 229, 241, 242] The exact rate of infections differs greatly between studies, largely attributable to varying definitions and methodology. In a population-based study from southern Sweden, 40% of patients suffered at least one infection requiring hospitalisation.[241] Infections occur most frequently in the first year following diagnosis.[241, 243] This may be attributable to the substantial immunosuppression during this time. Higher disease activity, higher s-Creatinine, dialysis dependency, and higher cumulative doses of cyclophosphamide and glucocorticoids have all been associated with an increased risk of infections.[241, 243, 244, 245] However, the use of rituximab has not been shown to be associated with fewer infectious complications.[190, 246] Instead, the use of rituximab may be complicated by hypogammaglobulinemia, predisposing infections.[246, 247]

Other risk factors identified are older age, lung involvement of active vasculitis, and pre-existing chronic obstructive pulmonary disease.[248] The respiratory tract is also the most frequent site for infections in patients with AAV, followed by the urinary tract and septicemia.[241, 243]

In addition to mortality, infections add considerably to the morbidity and the importance of thwarting infections is increasingly recognised. Initially primarily used as a prophylactic against opportunist Pneumocystis jirovecii infection, trimethoprim-sulfamethoxazole has been shown to reduce the total risk of severe infections and are recommended to be used in patients receiving high dose glucocorticoids, rituximab, or cyclophosphamide.[179, 243, 248]

Malignancies

Historically the rates of malignancies in AAV patients have been increased compared to the background population, but the rates appear to be declining.[249, 250] The high cumulative doses of cyclophosphamide in earlier treatment protocols likely contributed to the increased rates, of especially non-melanoma skin cancer, urothelial cancer, and leukemia.[251, 252] No increase in risk of malignancies is seen in patients receiving rituximab, compared to general population.[253] In patients receiving low cumulative doses of cyclophosphamide there is no increased incidence of malignancies other than squamous cell carcinoma.[254, 255] Malignancy is one of the leading long-term causes of mortality in AAV, especially in younger patients.[229]

Irreversible organ damage

Damage in AAV is irreversible impairment, not responding to immunosuppressive therapy. Damage may be induced by both disease and treatment and is seen during long-term follow up in more than 90% of patients.[256, 257] Introduced in 1998, the Vasculitis Damage Index (VDI) is a cumulative scoring system for vasculitis damage, consisting of 64 items in eleven subcategories.[258] The items may be attributable to either disease, treatment, or other comorbidities, and all items are of the same weight. Damage needs to be set apart from disease activity, which may be resolved following immunosuppressive treatment. As such, an item can only be scored after being present for three months or more, and only if emerging after vasculitis onset. VDI is one of the key outcome measures in AAV, as identified by OMERACT.[213]

Relapse

Clinical relapses, that is the return of disease activity after remission, are common in AAV. For relapse to occur, clinical remission must be achieved. Exactly what remission entails may differ between studies, but the absence of disease activity (a BVAS of 0) paired with a low dose of corticosteroids (prednisolone less than or equal to 7.5 mg/day) over a defined period of time are the EULAR recommendations for clinical trial design.[259] While overt clinical symptoms, or clear biomarkers of active inflammation (e.g., elevated C-reactive protein), might be easy to detect, there is evidence that subclinical inflammation persists even in patients considered to be in clinical remission. This is, however, not true for all patients. Sustained remission, without need for immunotherapy, while uncommon, is achieved.[260]

Today, there are identified biomarkers for disease activity, but none have been validated for clinical use in relapse prediction. Urinary soluble CD163 (a protein associated with activated glomerular macrophages) is a biomarker with high precision in separating kidney flare from kidney flare mimics, and may be closest

to wide use in clinical practice.[261, 262] Other markers include circulating leukocyte subsets, urinary leukocytes, and other urinary leukocyte proteins.

Biomarkers indicative of prognosis are rarer. $CD8^+$ T cell gene expression has been shown to be indicative of subsequent relapse, as have changes in serum calprotectin levels.[84, 263] Additionally, the prognostic capabilities of serial ANCA testing are widely debated. Already proposed in 1985 by van der Woude and colleagues, it was further investigated in a prospective study in 1990 with results indicating that changes in ANCA titres predict changes in disease activity, and a benefit of pre-emptive treatment in preventing relapse.[42, 264] Since then studies have shown conflicting results, with meta-analyses showing limited utility of serial ANCA measurements to guide treatment decisions.[265, 266] Despite this, it has been argued that serial measurements during disease remission may serve a purpose in certain clinical settings. Longitudinal ANCA measurements may be more useful in patients with severe disease manifestations, such as alveolar haemorrhage and glomerulonephritis, and after remission induction treatment with rituximab. [267, 268, 269, 270] The potential prognostic capabilities of antibody titres align with observations of an association of relapse with plasma cell frequencies and B-cell reconstitution in patients treated with rituximab.[271, 272]

PR3-ANCA positivity is known to be associated with disease relapse. The same is true for clinical diagnosis and phenotypic expression, with relapse being more common in GPA than MPA, and in patients with respiratory and cardiovascular disease involvement. Invertedly, higher s-Creatinine levels at diagnosis are associated with a lower risk of relapse.[273] Evident from clinical trials is also that the rate of relapse depends on the choice of induction and maintenance treatment. In keeping with this the exact rate of relapse is highly variable between different clinical trials and observational studies. The five-year relapse rate in patients induced with cyclophosphamide is estimated at 47%.[274]

End-stage kidney disease

Patients with AAV presenting with impaired kidney function are at high risk of developing end-stage kidney disease (ESKD). While the exact definition may vary between studies, an estimated glomerular filtration rate (eGFR) of less than 15 per minute per 1.73 m^2 of body-surface area or dialysis dependency is indicative of ESKD. Considering the significant impact of renal replacement therapy on the quality of life and the increased mortality in patients with impaired kidney function, several attempts of predicting the progression to ESKD have been made. In 2018, Brix and colleagues, developed the renal risk score for early risk prediction of ESKD, combining clinical and histopathological parameters.[275] The renal risk score is one of several systems attempting to stratify severity of kidney involvement in AAV. Others include the Berden classification and the Mayo Clinic/Renal

Pathology Society Chronicity Score, both focusing on histopathological changes.[276, 277, 278]

In a population-based study from southern Sweden 18.4% of patients developed ESKD, with MPO-ANCA positivity being associated with worse kidney survival.[37] In the long term follow up of EUVAS clinical trials the rates of ESKD were 20% and 9% for MPA and GPA, respectively.[256]

Chronic dialysis seems to be protective of disease relapse but is associated with an increased risk of infections. As such, treatment discontinuation in patients in sustained dialysis in absence of extrarenal disease manifestations should be considered.[279] However, assessing disease activity in patients on dialysis is notoriously difficult, and the decision to discontinue maintenance needs to be made on an individual basis.

In a large European cohort study, the rate of patients recovering independent kidney function within 90 days after commencing acute dialysis was 5.1%, which is relatively high compared to control groups with kidney disease caused by glomerulonephritis, diabetes, and non-diabetes.[280] The same study showed that 22% of patients with AAV in renal replacement therapy received a kidney transplantation. However, there is little consensus how to manage kidney transplantation in AAV.[281] Overall ten-year graft survival is 50-70%.[280, 281, 282] In a European survey by Little and colleagues, ANCA positivity at the time of the transplant was associated with severe vasculopathy of the graft (OR: 4.4 [95% CI 1.1-16.8]), which in turn was associated with reduced graft survival.[281] A short interval between disease remission and transplant was a predictor of patient mortality.

Mortality

While survival in AAV have been dramatically improved, the mortality rate compared to the background population is still increased. In a population-based study from southern Sweden the standardised mortality ratio was 2.8 (95% CI: 2.0-3.7).[283] This is comparable to other observational cohorts.[227, 284] The mortality in long-term follow-up studies of clinical trials, also show worse survival compared to the background population (when matched for age, sex, calendar year, and country), with 14.2%. 19.9%, 28.8% and 36.3% more deaths at 5, 10, 15 and 20 years following diagnosis respectively.[229] As been described under the respective sub-headings the main causes of death are infections, cardiovascular events, and malignancies. The mortality due to active vasculitis today is comparatively low (3.6% of all deaths in the long-term follow up of clinical trials).[229] However, the disease phenotype at the time of diagnosis still has prognostic implications.[285]

Few prognostic models for outcome of AAV exists today. In 1996 the French Vasculitis Study Group developed a score for assessing the risk of mortality in

EGPA, polyarteritis nodosa, and MPA, the Five Factor Score.[286] The presence of five identified items (proteinuria >1g/day, s-Creatinine ≥140 μmol/L, gastrointestinal involvement, central nervous system involvement and cardiomyopathy) at the time of diagnosis generate one point each. The five-year mortality for zero, one and two or more points were 11.9%, 25.9%, 46.0% respectively. In 2009 the score was updated to also include GPA.[287] In this version, five items are still indicative of prognosis, but ear-nose-throat involvement is associated with better survival (the absence of which is scored one point [only to be used in GPA and EGPA]), while age > 65 years, cardiac involvement, kidney involvement and gastrointestinal involvement are negative prognostic factors. Today, the original 1996 version is the one primarily used, especially in EGPA, where it may guide the choice of treatment.[180]

Real-world data

This short review should have made it evident that while much is known about AAV, much is still to be understood. One of the main obstacles of gaining further knowledge is the rarity of the condition. To tackle this, patient registries, that is systematically organised databases for the collection and storing of observational data, have become increasingly popular in rare disease research. In AAV, several regional and national registries have been developed, providing so-called real-world data (e.g., electronic-health records, health insurance claims or, patient-registries – effectively any data not sourced from clinical trials).

Through real-world data essentially all aspects introduced in the preceding review may be studied, highlighting the asset that is patient registries and observational cohorts. However, one major concern remains. These datasets are generally fragmented, siloed, and lack data sharing mechanisms. In this thesis, we address this research data fragmentation through the integration of real-world observational cohorts and registries in AAV. Using the integrated data, patterns of symptoms are explored and grouped, and models built for prediction of adverse disease outcome.

Rationale and aims

The aim of this thesis is to address research data fragmentation in AAV through the integration of real-world observational cohorts and registries. Using the integrated data I further aim to find, explore, and stratify patterns of symptoms, and build models for the prognostication of disease relapse, infection, ESKD, and mortality.

Rooted in the wider Findable, Accessible, Interoperable, Reusable, Vasculitis (FAIRVASC) project, the patient perspective has been considered throughout this thesis work, with patient representatives involved since inception. Under the auspices of FAIRVASC, discussion workgroups among individuals with lived experience of AAV have been carried out to explore what registries are, what information is stored, data protection and patient rights, and research priorities. From a lived experience perspective, the use of pre-existing information to find patterns of symptoms and predicting adverse outcome were two of the identified research priorities.

Specific aims

In Study I we aim to (1) describe the data structure, federation and harmonisation process, (2) explore data quality and (3) give an overview on the baseline characteristics, treatment and outcomes of AAV across six European registries.

In Study II we aim to identify phenotypically distinct subgroups and develop a data-driven subclassification of AAV.

In Study III we aim build and assess predictive models for disease outcome (i.e., major relapse, infection, ESKD and mortality) in AAV, using commonly collected and recorded data at the time of diagnosis.

Methods

Overarching study design

This thesis work was made within the FAIRVASC project, a European collaboration to enable access to a network of six AAV registries and observational cohorts.

The project addressed a concern present in all studies of rare disease, sample-size. Quantitative clinical research relies upon analysis of a sufficiently large number of observations to allow statistical inference. With a low prevalence, the patient numbers of AAV are a small in any single cohort, and existing cohorts have been fragmented, siloed, and not standardised to allow for interoperability.

Resting on the FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles for scientific data management and stewardship, introduced by Wilkinson and colleagues in 2016, the FAIRVASC project set out to connect European vasculitis registries using a Semantic Web approach.[288] The Semantic Web is a vision to extend and transform the World Wide Web from a collection of isolated documents to an interlinked information space, enabling automated data processing. This is achieved through standard web technologies such as hyper-text transfer protocol (HTTP) and Uniform Resource Identifiers (URIs) but expands the original concept by enabling encoding of semantics with the data. *Semantics*, in this context, centers around assigning contextual meaning to data, ensuring that it can be effectively used, integrated, and understood by humans, but importantly also machines. Semantic integration is attained through a set of technologies such as Resource Description Framework (RDF), SPARQL Protocol and RDF Query Language (SPARQL), and ontologies (Table 3).

It is here important to distinguish between two general approaches to address data fragmentation, *centralisation* and *federation*. In centralisation, the data-controllers release data from their local repositories to a centralised pool, where the data then can be accessed. Due to the sensitive nature of health-care data, this poses regulatory, legal and ethical challenges related to data protection and privacy. To address these challenges, federation has emerged as an alternative. Here data stays under the data controller, but remote access is enabled. To avoid data-leakage this access is constrained to pre-defined queries or statistical models, effectively bypassing governance obstacles. However, while in some contexts federation can

achieve comparable performance as centralisation, there are still statistical limitations to what can be achieved.

Table 3. Semantic Web dictionary

Word	Description
Hyper-text transfer protocol (HTTP)	The foundational communication protocol used on the web to transfer data between a client and a server, enabling the retrieval and display of web pages.
Uniform Resource Identifier (URI)	A string of characters that uniquely identifies a resource on the internet, such as a web page.
Resource Description Framework (RDF)	A standard model for representing information on the web, using a triple-based structure of subject, predicate, and object to describe relationships between resources in a way that can be understood and processed by machines.
SPARQL Protocol and RDF Query Language (SPARQL)	The query language used to retrieve and manipulate data stored in RDF format.
Ontology	A structured representation, formal naming and defintions of the categories, properties and relationships within a domain, providing a shared vocabulary.
Knowledge graph	The amalgamation of the above concepts, providing an interconnected representation of knowledge that machines can understand and reason over.

While the FAIRVASC project at its core was a project of federation, the requirement for complex statistical modelling, necessitated an ad hoc centralised approach. As such, here Study I rely on federation, while Study II and Study III relies on centralisation. However, the need for interoperability and semantic integration remained, regardless of method of data access.

The data structure and variables collected across the registries in the FAIRVASC project were different. To support the analysis of data across these registries, a process of harmonisation was required. This was achieved through the development of the vasculitis-specific FAIRVASC ontology and the transformation of unstructured tabular registry data to a knowledge graph data format.[289]

Tabular data and the RDF graph

Tabular data is the usual format for clinical research data. In the example of assigning a patient a diagnosis, the row would usually represent the patient, identified through a unique id, and column represent a specific attribute (in this case the diagnosis). In the the intersection of the row and column (the cell), the value of the specific attribute for the patient would be represented (the type of diagnosis).

Often, for tabular data, the information in the cell would not be self-explanatory but require a separate data-dictionary (also in the format of a table).

Example of tabular data

id	diagnosis
001	1

Example of data dictionary for tabular data

Variable name	Description	Data type	Choices
id	Unique patient identifier	String	
diagnosis	Type of vasculitis diagnosis	Categorical	1. GPA | 2. MPA | 3. EGPA

In RDF, the same information would be described as a triple. Patient 001 (subject) has diagnosis (predicate) granulomatosis with polyangiitis (object), and represented using (a for the sake of argument somewhat simplified) FAIRVASC ontology as :

<http://w3id.org/FAIRVASC/patient/001>

<http://w3id.org/FAIRVASC/hasDiagnosis>

<http://orpha.net/ORDO/Orphanet_900>

Here the object links to the Orphanet rare disease-ontology term for granulomatosis with polyangiitis.

The ontology development and data-transformation were done using an iterative approach, guided by competency questions identified by the clinical experts of the project. The end-result was the creation of a cohort of AAV of unprecedented size,

accessible through either a dedicated web-interface (Study I) or through a central data-store (Study II and III).

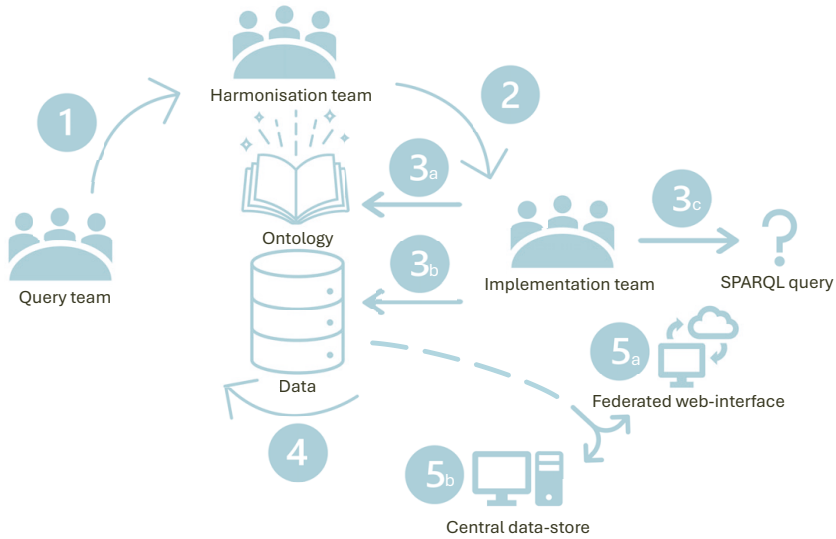


Figure 5. The iterative FAIRVASC interoperability process
1. The *query team* of vasculitis experts pose a competency question. 2. The *harmonisation team* of registry experts identify the terms needed to answer the question and find suitable common names for the terms (preferably using existing ontologies). 3. The *implementation team* of data managers and computer scientists: a) implement the identified terms in the FAIRVASC ontology, b) map the local registry data to the ontology terms using RDF, and c) write the SPARQL query to retrieve the updated information. 4. Feedback from the implementation informs the next iteration. 5. The harmonised data is: a) uploaded to a local server exposed to the web-interface, and b) sent to the central data store for pooling.

Study I

In Study I, we set out to describe the data structure, federation and harmonisation process within FAIRVASC to a clinical audience, explore and describe the quality of the data in the registries, and give an overview on the characteristics, treatment and outcomes of AAV across the participating registries (Table 4).

Table 4. Participating registries meta-data summary

	Czech	FVSG	GeVas	POLVAS	RKD	Skåne
Included diagnoses	EGPA, MPA, GPA, unclassified AAV	EGPA, MPA, GPA	EGPA, MPA, GPA, unclassified AAV	EGPA, MPA, GPA, unclassified AAV	EGPA, MPA, GPA, unclassified AAV	EGPA, MPA, GPA
Classification method	EMA, clinical	CHCC 2012, ACR 1990	CHCC 2012, ACR 1990, clinical	CHCC 2012, ACR 1990	EMA	EMA
Type of recruitment area	National	National	National	National	National	Regional
Name of recruitment area	Czechia	France	Germany, Austria, Switzerland	Poland	Ireland	Skåne, Sweden
Care setting	Secondary/tertiary	Secondary/tertiary	Vasculitis centres	Secondary/tertiary	Secondary/tertiary	Population coverage
Medical speciality	Nephrology	All	All	All	Nephrology	All
Data source	Encounters	Encounters	Encounters	Chart review	Encounters	Chart review
Period of recruitment	2013 – continuous	1983 – continuous	2019 – continuous	2016 – continuous	2012 – continuous	1997 – 2019

Czech: the Czech Registry of AAV, FVSG: the French Vasculitis Study Group Registry, GeVas: the Joint Vasculitis Registry in German-speaking Countries, POLVAS: the Polish Vasculitis Registry, RKD: Irelands's Rare Kidney Disease Registry, Skåne: Sweden's Skåne Vasculitis Cohort, Unclassified AAV: Unclassified ANCA-associated vasculitis, EMA: European Medicines Agency algorithm, CHCC 2012: 2012 Revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides, ACR 1990: The American College of Rheumatology 1990 criteria.

Study design and participants

The participating registries were the Czech Registry of AAV, the French Vasculitis Study Group registry (FVSG), the Joint Vasculitis Registry in German-speaking Countries (GeVas), the Polish Vasculitis Registry (POLVAS), Ireland's Rare Kidney Disease (RKD) registry, and Sweden's Skåne Vasculitis Cohort (Figure 6). Included were all patients with a diagnosis of AAV, regardless of classification method.

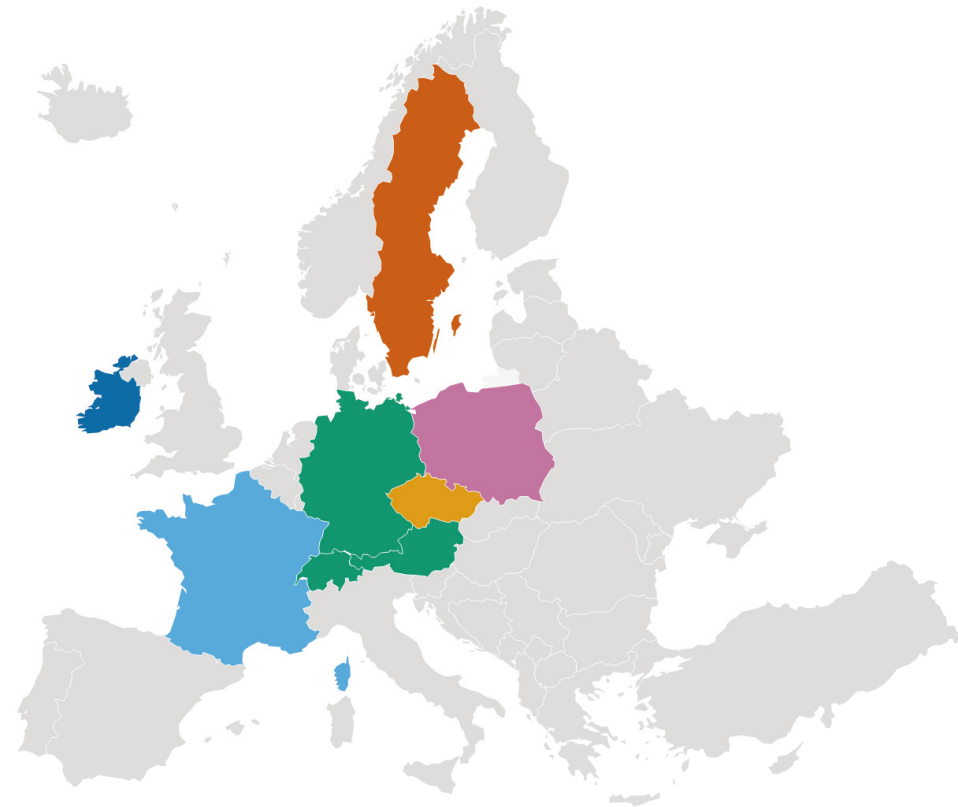


Figure 6. The six participating vasculitis registries of the FAIRVASC project
Czech Registry of AAV (light orange), the French Vasculitis Study Group registry (light blue), the Joint Vasculitis Registry in German-speaking Countries (green), the Polish Vasculitis registry (pink), Ireland's Rare Kidney Disease registry (dark blue), and Sweden's Skåne Vasculitis Cohort (dark orange). Note that the areas of recruitment may not cover the full geographical areas as represented by this map.

Procedures

This study rested upon federation of data. The general workflow can be summarised as: Registry data were quality controlled, harmonised to the common schema developed with the FAIRVASC ontology, and uploaded to a local server with an exposed endpoint. Predefined queries were dispatched from a web interface to the local servers where the query was automatically processed and aggregated non-subject level results returned to the web interface (Figure 7). In the following sections these steps will be further clarified.

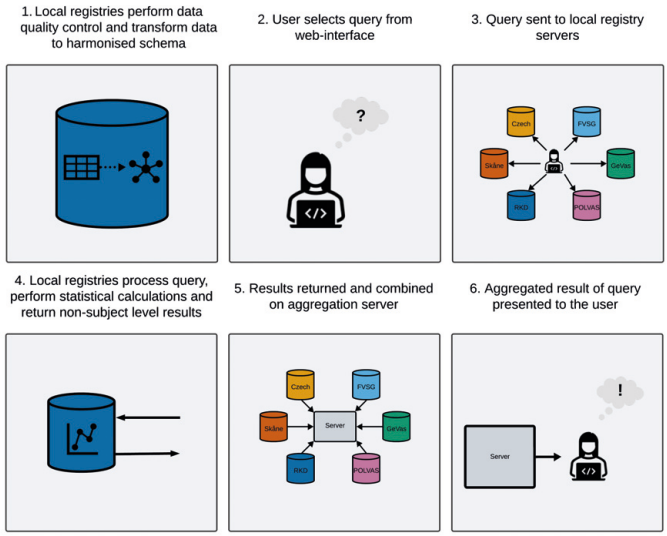


Figure 7. The FAIRVASC federated query workflow
Registry data are quality controlled and harmonised to a common schema. Harmonised data are uploaded to a local server with an exposed endpoint. A researcher dispatches a predefined query from a web interface to the local servers. Non-subject level results sent back. Data are combined and presented to the web-interface user.

Data quality framework

The data quality in the registries were assessed across four core domains: uniqueness, consistency, completeness and correctness. The domains were selected in collaboration with the European Institute for Innovation through Health Data (i~HD), based on a pool of nine candidate dimensions in the i~HD data quality assessment framework.[290]

Based on these four domains, we developed a data quality worksheet. Using this worksheet, data quality was assessed at each registry site by local investigators, unblinded. The results of the data quality assessment were presented as variable specific percentages, stratified by registry.

In assessing *uniqueness* (absence of duplication of data), we investigated duplicate patient identifiers and potential cases of readmission of the same patient under another unique identifier within each registry. We identified possible cases of duplication by finding individuals sharing date of birth and gender, and further compared these by hand.

We assessed *consistency* (adherence of datatype to what is specified in the registries' data dictionaries) by examining nine key variables in AAV research (gender, date of birth, immunoassay ANCA-type, BVAS at diagnosis, serum creatinine at diagnosis, C-reactive protein (CRP) at diagnosis, induction treatment, date of death and date of ESKD). Under the consistency domain, we also applied logic tests to variables containing dates ("Was date of death greater than date of diagnosis" and "Was date of death greater than date of birth"). We further tested plausibility of numeric values for three variables ("Was serum creatinine at diagnosis within a biologically plausible range of 0–5000 μmol/L", "Was BVAS at diagnosis within the plausible range of 0–63?", and "Was CRP at diagnosis within a plausible range of 0–1000 mg/L").

We assessed the absence of missing data (*completeness*) across the same variables. Lastly, we assessed *correctness* of the same variables against an electronic health record for at least ten patients per registry.

Data harmonisation

While the harmonisation of data has been described under *Overarching study design*, some additional clarifications are needed here. The process of harmonisation did not alter the source registry data. This data stayed in the format it was, be it a single Excel spreadsheet, as a Research Electronic Data Capture (REDCap) project, or some other relational database management system.

In a process of semi-automated *data mapping*, the new RDF data model was created to sit "on top" of the source registry data, on a local server exposed to the web. Via a dedicated web interface, the data could then be queried using SPARQL (Figure 8).

The federation infrastructure

While there are alternative architectures for federation infrastructures, the architecture of FAIRVASC rested on an aggregation server. The local registry servers hosting the harmonised data were connected to the aggregation server, which was exposed to a password protected web interface (Figure 9). Personal passwords were administered by a data access committee, and the type of query posed, username and time of the query logged on the local servers.

From the web interface a user could select from a range of predefined queries, retrieving non-subject level data. To further ensure data protection and privacy, low cell counts were filtered.

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX bvas: <http://w3id.org/BVAS#>

SELECT (COUNT(DISTINCT ?patient) AS ?gpa_n)
WHERE {
  {
    SERVICE <http://czech_example.org/sparql> {
      ?patient a fvc:Patient.
      ?patient fvc:hasPatientOverview ?PatientOverview.
      ?PatientOverview fvc:hasDiagnosis ?Diagnosis.
      ?Diagnosis fvc:mainDiagnosis ?maindiagnosis.
      FILTER (?maindiagnosis = <http://www.orpha.net/ORDO/Orphanet_900>)
    }
  }
  UNION
  {
    SERVICE <http://fvsg_example.org/sparql> {
      ?patient a fvc:Patient.
      ?patient fvc:hasPatientOverview ?PatientOverview.
      ?PatientOverview fvc:hasDiagnosis ?Diagnosis.
      ?Diagnosis fvc:mainDiagnosis ?maindiagnosis.
      FILTER (?maindiagnosis = <http://www.orpha.net/ORDO/Orphanet_900>)
    }
  }
#... All endpoints are listed in the same fashion
}
```

Figure 8. Federated SPARQL query example
An example of a federated query retrieving the number of patients with a diagnosis of granulomatosis with polyangiitis from two registries. Note that the endpoints are made up for this example.

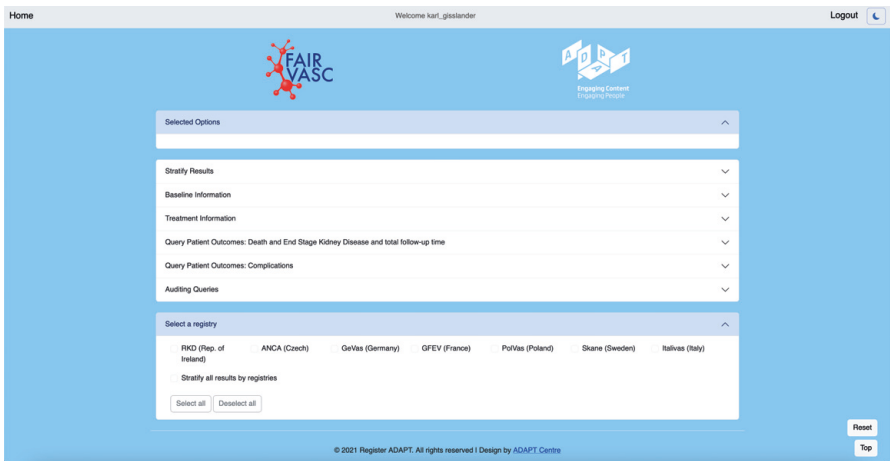


Figure 9. FAIRVASC web interface
Snapshot of the FAIRVASC query interface from 2024-04-07. Note that the Italian vasculitis registry (Italvas) have been added to consortium since Study I.

Patient characteristics retrieval

For this study we analysed patients recruited to the participating registries with a diagnosis of AAV up to 31st of January 2023, with descriptive analysis of demographics, type of diagnosis, organ involvement, serology, serum creatinine level at diagnosis, treatments, follow-up and outcome. Using the pre-defined federated queries available via the interface we retrieved this information from the local registries.

Statistical analysis

As a descriptive study, relying on the native capabilities of SPARQL, the statistical analysis was limited. Continuous variables were summarised with a mean and standard deviation. Categorical variables were summarised as frequencies with valid percentages. The incidence rate of ESKD and all-cause mortality rate per 1000 person-years with a 95% confidence interval, combined across the registries, were estimated using random effects meta-analyses, as available in the R package *metafor*.[291] To ensure data protection and privacy, the federated queries did not allow for the flow of subject level data from one site to another or to the aggregation server, which prohibited the investigation of summary statistics requiring the global rank of data to be known (e.g., quantiles).

Study II

In Study II, we set out to identify phenotypically distinct subgroups and develop a data-driven subclassification of ANCA-associated vasculitis, using model-based clustering of the large real-world dataset generated through the efforts of Study I. Due to the statistical limitations of the federated approach, described under *Overarching study design*, Study II required the assembling of a central data repository. The central repository allowed access to a subject level dataset, with data minimised for the purpose of Study II and Study III. Following the development of a data sharing agreement between the participating registries we assembled a data pool in Lund, Sweden.

Study design and participants

Again, patients from the Czech registry of AAV, the FVSG registry, the GeVas registry, the POLVAS registry, the RKD registry, and Sweden's Skåne Vasculitis Cohort were included in the study. Inclusion further required a diagnosis of either GPA or MPA (irrespective of classification criteria), a valid date of diagnosis, and the absence of unacceptable data quality concerns.

While classified as AAV, EGPA has limited overlap with GPA and MPA with respect to genetics, pathophysiology, clinical presentation, and therapy. Furthermore, we considered there to be a paucity of data elements essential for the discrimination of EGPA (e.g., asthma, nasal polyposis, and eosinophilia) available in the participating registries. Consequently, in Study II, patients with a diagnosis of EGPA were excluded.

Model-based clustering being an exploratory analysis, we conducted no sample size calculation. However, we anticipated that the size of our observational study cohort would be large enough to reach a minimum number of observations per cluster for adequate statistical power.[292]

Procedures

To explore the phenotypic spectrum of AAV we used a total of 17 mixed-type variables as input for the model. Age at diagnosis, and serum creatinine and CRP concentrations at diagnosis, were regarded as continuous, while gender and the involvement of constitutional symptoms, the musculoskeletal system, skin, eyes, mucosa, ear–nose–throat, lung, the cardiovascular system, the gastrointestinal system, kidneys, and the central and peripheral nervous system at diagnosis were regarded as binary. ANCA was regarded as a three-level nominal consisting of ANCA negative, PR3-ANCA positive, and MPO-ANCA positive. These variables were selected through consensus agreement within the author group, guided by data availability and domain-expertise of features considered to be of clinical importance in the stratification of AAV.

In the absence of an antigen-specific immunoassay, patients were assigned to PR3-ANCA or MPO-ANCA if immunofluorescence showed cytoplasmic or perinuclear staining pattern, respectively. Cases positive for both PR3-ANCA and MPO-ANCA were reassigned based on immunofluorescence (i.e., cANCA to PR3-ANCA and pANCA to MPO-ANCA) or clinical diagnosis (i.e., GPA patients were assigned PR3-ANCA and MPA patients assigned MPO-ANCA)

Due to inconsistency in variable definitions compared with the other registries, data concerning serum creatinine and organ pattern from the POLVAS registry were defined as “highest ever” and “at any time”, respectively.

Patients with nine or more missing variables were excluded from analysis. Remaining missing values were imputed using multiple imputation, as available in the R-package *mice*, guided by a framework for multiple imputation in cluster analysis by Basagaña and colleagues.[293, 294] We used predictive mean matching (continuous data), logistic regression imputation (binary data) and polytomous regression imputation (nominal data) for the imputation. In addition to the 17 variables described above, outcome data, registry affiliation, and the type of vasculitis diagnosis were included in the imputation model to generate predictor

values but were not imputed. Patients known to be under acute dialysis with a missing serum creatinine value were randomly assigned a value ranging from 500 μmol/L to 1000 μmol/L.

Statistical analysis

We performed model-based clustering of the imputed datasets using a parsimonious mixture of two latent Gaussian variable models as available in the R package *clustMD*.[295] Two mixture models were fitted for 1–5 clusters using a Monte Carlo expectation maximisation algorithm with 500 iterations. The two mixture models used both assume conditional independence between variables and allow for the shape of each cluster to vary. One model (EVI) assumes that the overall volume of each cluster is the same, while this property can also vary for the other type of model (VVI). We selected the optimal combination of number of clusters and model through the highest average approximate Bayesian information criterion (BIC) over the imputed datasets.

As there were potential differences in the assigned cluster affiliation between the imputed datasets, patients were assigned to the cluster with the highest average probability over the imputed datasets fitted with the optimal model and number of clusters, following relabelling to consistent cluster labels using Stephens' method.

To explore the identified clusters, we summarised the demographic, phenotypic and serological characteristics, clinical diagnosis, and outcome. Summary statistics were presented as valid percentage, excluding missing data. The impact of the heterogenous recruitment settings among the source registries was assessed by testing differences in cluster assignment between registries using χ^2 tests. Internal validity, and the impact of the multiple imputation, was assessed through cluster-wise assessment of cluster stability over the ten imputed datasets using the Jaccard coefficient.

While the use of an external validation cohort is common in supervised classification, it is less common and less straightforward for the validation of clustering results. This being an exploratory analysis with no objective ground-truth, we here opted for the use of repeatedly held out data for the validation of the clusters, as adapted from O'Hagan and colleagues.[296]

To further explore our clustering model, we compared the outcome of the clusters with respect to survival and competing risks time-to-ESKD using Kaplan–Meier estimators and cumulative incidence functions, respectively. As the POLVAS registry did not record the date of ESKD, patients from this registry were excluded from this time-to-event analysis. Patients lost to follow-up were censored at their last known contact date with their respective registry. The potential predictive value of our clustering model was compared to subclassifications based on clinical diagnosis, ANCA specificity, and a previous cluster-based stratification by Mahr

and colleagues.[41] In these models we assessed the predictive accuracy towards survival and kidney survival using Cox proportional hazards models, and Fine–Gray models for competing risk, respectively. Models were fit with only the subclassification labels or with gender, age, and source registry as covariates, and assessed using the Akaike information criterion. The models were pooled over the imputed datasets, following the framework developed by Basagaña and colleagues.[293]

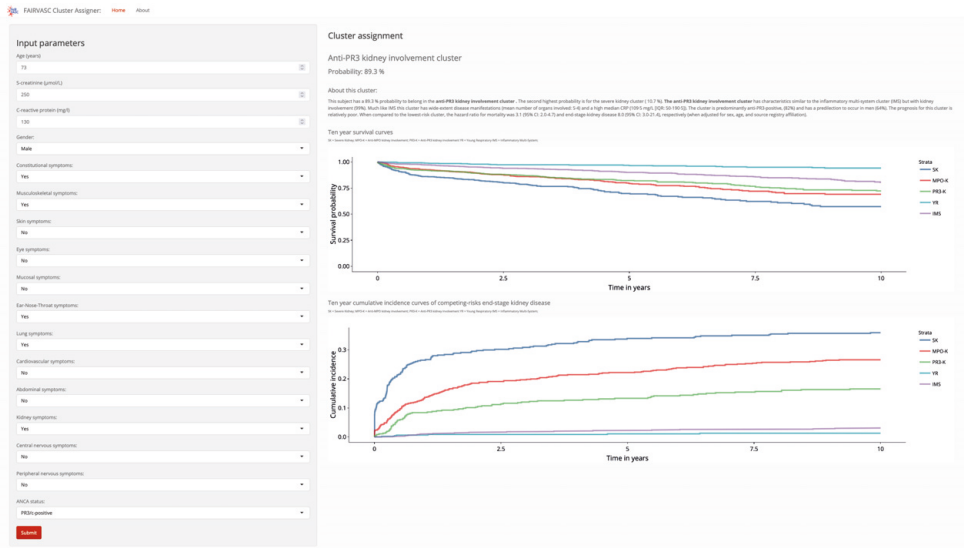


Figure 10. The FAIRVASC web application for cluster assignment
A snapshot of the web application allowing for input of parameters and assignment of a cluster affiliation.

Lastly, we built a web application to allow for cluster assignment of previously unseen patient data, using the R package *shiny* (Figure 10).[297] The applications allow for the submission of data and presents probabilities for an individual to belong to the respective clusters, following a simulated expectation-step of the averaged parameters from the cluster models of the main analysis. In addition to limiting the potential range of input, poorly fitting data is flagged. This is achieved through the comparison of the log-likelihood of the component that a data point is assigned to under the estimated clustering model to its likelihood under a uniform noise component that covers roughly the range of the observed data. If the data point fits the uniform component better than the cluster model, the input is flagged (i.e., the patient data does not have a strongly recognisable phenotypic pattern).

Study III

In Study III, we set out to build and assess predictive models for adverse disease progression in AAV, using commonly collected and recorded data at the time of diagnosis. Again, due to the statistical limitations of the federated architecture, this study made use of a central data repository, minimised for the research purpose.

Study design and participants

Inclusion in Study III required a new diagnosis GPA or MPA regardless of classification criteria (again excluding EGPA), with a known date of diagnosis and a complete and standardised disease activity assessment available at the time of diagnosis. This last requirement meant that data from the POLVAS registry was excluded from this study. As such, data was retrieved from the five other participating registries: the Czech registry, the FVSG registry, the GeVas registry, the RKD registry, and the Skåne vasculitis cohort.

However, due to differences in the source registry designs, the inclusion further differed depending on the outcome of interest. Data from all registries were included in the prediction of mortality and ESKD. The Skåne vasculitis cohort was excluded from the study of major relapse, and Czech registry, the FVSG registry, and the GeVas registry excluded from the prediction of serious infection.

Procedures

We retrieved the full BVAS assessment at the time of diagnosis, along with the levels of serum creatinine, gender, age and ANCA status of the patients at the time of diagnosis. Using the 2021 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) algorithm, we estimated the eGFR.[298]

We then retrieved outcome data regarding our four events of interest. Date of death, date of ESKD, and date of last follow up were retrieved directly. The date of the first infectious event with a Common Terminology Criteria for Adverse Event (CTCAE) score of three or higher, following the date of diagnosis for each patient were retrieved.

Lastly, we retrieved the estimated date of major relapse. We first employed a wash-out period of six months after which a patient was estimated to be in disease remission. The date of relapse was then defined as the first visit with major disease activity (i.e., greater to or equal to three minor items or the presence of one or more major activity item as reported by a BVAS assessment) following the estimated disease remission.

We further imputed missing data using random forest imputation, as available in the *R*-package *missForest*.[299] For serum creatinine the missing data was manually imputed through random sampling within the boundaries defined by their disease activity assessment.

Statistical analysis

First, we selected variables for the inclusion in prediction models using random survival forests, as available in the *R*-package *randomForestSRC*.[300, 301] Random survival forests are non-parametric supervised machine learning algorithms for time-to-event analysis that can capture complex relationships between predictors and outcome and identify the most important variables for the prediction of outcome.

The ten variables with highest variable importance in the prediction of each outcome, as identified by the algorithm, were selected for inclusion in the respective prediction model. The models were then built using the identified variables, using a Cox proportional hazards model for mortality prediction, and Fine-Gray models for ESKD, infection, and relapse prediction (all with mortality as a competing risk). We trained the models on the full dataset and assessed the performance through ten-fold cross-validation.

To assess the predictive accuracy of the models we visualised the area under the curve (AUC) of receiver operating characteristics curves and the prediction error defined by the Brier score, over the follow-up time. We further evaluated the calibration of the models at 1-year, 3-year, and 5-year post diagnosis. Using the AUC, we visualised the predictive performance of our models as compared to existing predictive models for the respective outcome. As sensitivity analyses, we evaluated the performance of the models for serious infections and ESKD separately for groups known to be receiving infection prophylaxis with trimethoprim-sulfamethoxazole, and plasma exchange, respectively.

Lastly, we estimated the potential the utility of the prediction models through decision curve analysis as available in the R-package *dcurves*, again using ten-fold cross-validation.[302] In decision curve analysis the net benefit of a model is calculated by putting false positives (harm) on the same scale as true positives (benefit). False positive rates are multiplied by an exchange rate (how many false positives are worth one true positive) defined by a probability threshold. Our proposed thresholds, while arbitrary, and interventions were as follow: For relapse; early discontinuation of immunosuppressive therapy if the five-year risk is below 5%. For infection, prolonged infection prophylaxis if the five-year risk exceeds 10%. For ESKD and mortality, general management guidance if the five-year risk is above 20%. To allow for outcome prediction of previously unseen data we built a web application, using the R package *shiny*.[297]

Ethical considerations

Within the FAIRVASC project the ethical approvals for the collection, sharing, and analysis of patient data were governed locally and obtained from the relevant ethics committee at each participating registry site. For the Skåne Vasculitis cohort, the studies included in this thesis were approved by the Regional Ethics Review Board in Lund (2010/517) and the Swedish Ethical Review Authority (2020-00697), requiring no informed consent. All other participating registry sites provided written informed consent directly from the patients.

Results

Study I

The process of integration of fragmented and siloed AAV datasets described in detail in the methods section (under the sub-headlines *Overarching study design* and *Study I*) of this thesis was successful. Here, we present the results of the source data quality assessment, and the results of the retrieval of harmonised data from the participating registries.

Data quality

When assessing uniqueness of data entries in the registries, we identified no duplicate patient identifiers. In the POLVAS registry, there were 2.2% potential duplicate entries (i.e., where the same patient was entered into the registry under more than one unique identifier), while no such duplications were found in the other registries.

Due to differences in registry design, not all variables were present in all registries. impacting how consistency, completeness, and correctness could be assessed. However, when consistency could be assessed the consistency of data type was 100% across all variables present in the registries. However, when testing the plausibility of data entries through logic tests, the results for dates were between 93.6% and 100% and, between 98.7% and 100% for numeric variables.

Completeness for demographic data (e.g., gender and date of birth) ranged from 95.1% to 100%, while completeness of laboratory data (e.g., ANCA, serum creatinine and CRP) were poorer, ranging from 49.5% to 99.2%. A BVAS assessment was available at diagnosis for 49.5%–100% of patients. While only four out of six registries provided explicit induction treatment data, at least one induction treatment type was recorded in 96.9% to 100% of the entries in these registries.

In patients who died or reached ESKD by the end of follow-up, a date of death or date of ESKD were available in 75%–100%. When comparing the registry entries towards the electronic health records for a subset of entries, the data correctness was between 60.0%–100% (Figure 11).

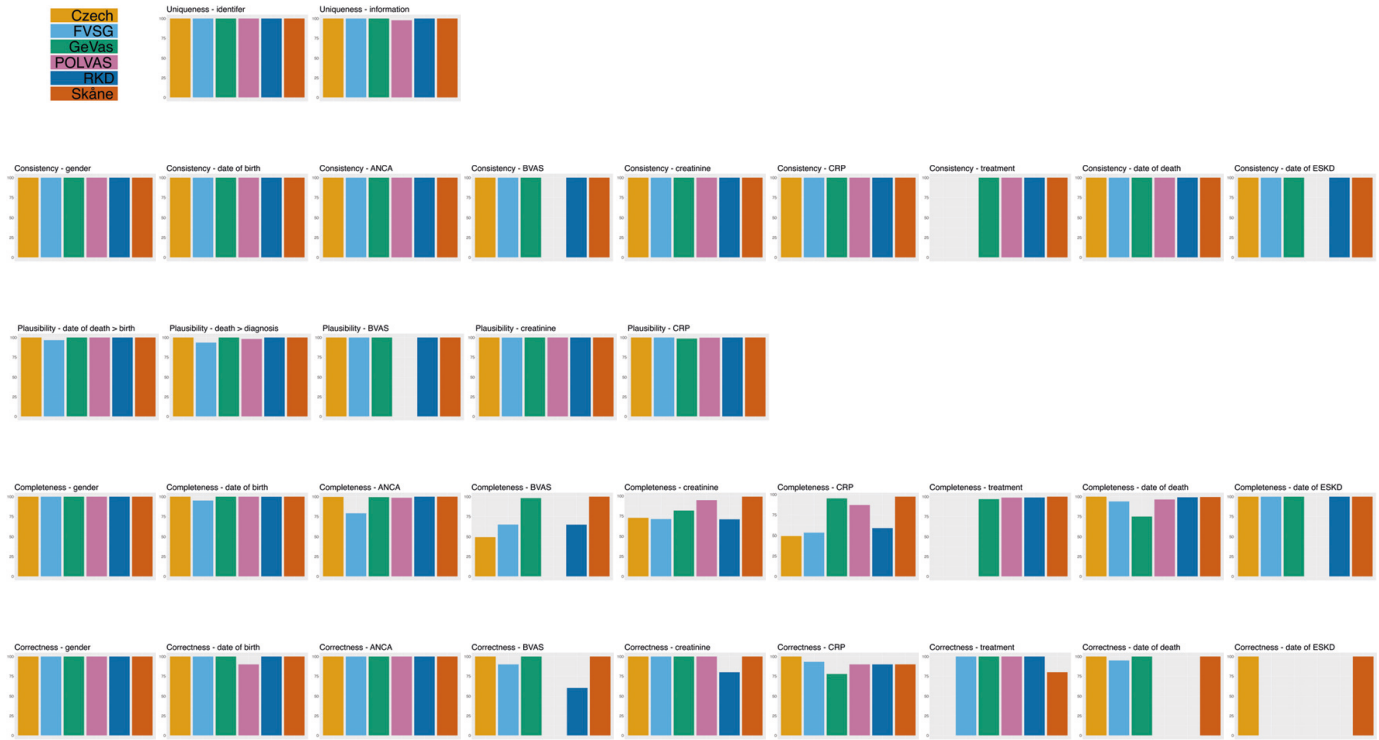


Figure 11. Data quality metrics per registry as percentages per key variable.
The absence of duplication of data (uniqueness), data-type consistency and plausibility of data values (consistency and plausibility), the absence of missing data (completeness) the correctness of data assessed against available electronic health records for a subset of patients (correctness). Variables that were missing from a registry because they were not targeted for collection are shown as the absence of a bar.

Retrieved patient characteristics

A total of 5282 patients (2568 [48.6%.] women and 2708 [51.3%] men) were retrieved, and included, across the six participating registries (Table 5). Of these, 2754 (52.1%) were classified as having GPA, 1580 (29.9%) MPA and 937 (17.7%) EGPA. The mean age at diagnosis was 56.0±16.7 years, of 4973 available.

There were 1840 (51.2%) PR3-ANCA positive, 1506 (41.8%) MPO-ANCA positive and 219 (9.0%) immunoassay ANCA negative patients of 3601 available. Most frequent was general symptoms (i.e., constitutional or musculoskeletal symptoms), while specific organ involvement, such as lung involvement was reported in 3281 (65.1%), ear-nose-throat involvement in 2860 (56.7%) and kidney involvement in 2534 (50.2%) of 5043 available. Cardiovascular and abdominal involvement were present in 822 (16.3%) and 658 (13.4%) of 5043 available, respectively. The mean serum creatinine level at diagnosis was 198±266 µmol/L of 3288 available across the registries.

Due to differences in source registry design, the retrieval of treatment information was limited, reflected by a paucity of available data. Cyclophosphamide was the most used primary induction treatment (2073 [72.5%] of 2858), followed by rituximab (505 [17.7%] of 2858). The use of pulsed intravenous glucocorticoids was highly variable and seen in a total of 1735 (55.2%) of 3141 patients. Similarly, there were differences in the use of plasma exchange, seen in a total of 421 (13.4%) of 3141 available. Azathioprine (911 [50.1%] of 1820 available) was the most used maintenance treatment, followed by mycophenolate mofetil (345 [18.9%]), rituximab (282 [15.2%]), and methotrexate (270 [14.8%]), although the practice differed between the registries.

When assessing the follow-up and outcome of the patients in the registries, there were a total of 30 548 person-years of follow-up, with a mean follow-up time of 6.2±5.8 years. During this follow-up, there were 767 deaths occurring with a known date of death, yielding a pooled all-cause mortality rate of 28.8 (95% CI 19.7 to 42.2) per 1000 patient-years. However, there was considerable heterogeneity among the registries (I^2=96%), with the highest mortality rate seen in the Skåne registry (62.8 [95% CI 54.4 to 72.4]), and the lowest in FVSG (19.0 [95% CI 17.0 to 21.3]) per 1000 patient-years (Figure 12). There was considerable heterogeneity also in the incidence-rate of ESKD, seen in a pooled estimate of 24.8 (95% CI 19.7 to 31.1) per 1000 patient-years (I^2=75%) (Figure 13).

Table 5. Patient characteristics summary stratified by registry

	Czech (n=335)	FVSG (n=2806)	GeVas (n=169)	POLVAS (n=932)	RKD (n=668)	Skåne (n=374)	Total (n=5282)
Demography and diagnosis							
Age, years	60.1 (15.2; 334)	55.1 (16.5; 2529)	59.5 (15.3; 169)	50.5 (15.9; 901)	59.1 (15.6; 666)	64.9 (16.2; 374)	56.0 (16.7; 4973)
Men	169 (50.4)	1441 (51.4)	83 (49.1)	431 (46.2)	384 (57.5)	200 (53.5)	2708 (51.3)
Women	166 (49.6)	1357 (48.4)	86 (50.9)	501 (53.8)	284 (42.5)	174 (46.5)	2568 (48.6)
GPA	143 (42.7)	1390 (49.6)	85 (50.3)	645 (69.3)	299 (44.7)	192 (51.3)	2754 (52.1)
MPA	178 (53.1)	683 (24.4)	54 (31.9)	169 (18.2)	337 (50.4)	159 (42.5)	1580 (29.9)
EGPA	5 (1.5)	731 (26.1)	28 (16.6)	118 (12.7)	32 (4.8)	23 (6.1)	937 (17.7)
Unspecified AAV	8 (2.4)	*	2 (1.2)	†	†	*	10 (0.2)
Laboratory							
PR3 positive	147/333 (44.1)	650/1283 (50.7)	79/169 (46.8)	457/774 (55.9)	320/668 (47.9)	187/374 (50.0)	1840/3601 (51.2)
MPO positive	170/333 (51.1)	615/1283 (47.9)	61/169 (36.1)	173/774 (21.1)	326/668 (48.8)	161/374 (43.0)	1506/3601 (41.8)
ELISA negative	16/333 (4.8)	*	26/169 (15.4)	134/774 (16.4)	17/668 (2.5)	26/374 (6.9)	219/3601 (9.0)
S-creatinine	222 (149; 253)	171 (282; 2017)	156 (187; 138)	*	288 (263; 509)	224 (223; 371)	198 (266; 3288)
Organ pattern involvement							
General	208/332 (62.7)	2093/2574 (81.3)	155/168 (92.3)	823/929 (88.6)	258/668 (38.6)	277/372 (74.5)	3814/5043 (75.6)
Mucous membrane, skin, eye	37/332 (11.1)	1359/2574 (52.8)	67/168 (39.9)	434/929 (46.7)	204/668 (30.5)	66/372 (17.7)	2167/5043 (43.0)
Ear-nose-throat	112/332 (33.7)	1595/2574 (61.9)	94/168 (55.9)	622/929 (67.0)	281/668 (42.1)	156/372 (41.2)	2860/5043 (56.7)
Lung	160/332 (48.2)	1792/2574 (69.6)	112/168 (66.7)	682/929 (73.4)	338/668 (50.1)	197/372 (52.9)	3281/5043 (65.1)
Cardiovascular	7/332 (2.1)	631/2574 (24.5)	17/168 (10.1)	127/929 (13.7)	20/668 (2.9)	20/372 (5.4)	822/5043 (16.3)
Abdominal	10/332 (3.0)	480/2574 (18.6)	11/168 (6.6)	114/929 (12.3)	34/668 (5.1)	9/372 (2.4)	658/5043 (13.4)
Kidney	310/332 (93.4)	720/2574 (27.9)	108/168 (64.3)	575/929 (61.9)	564/668 (84.4)	257/372 (69.1)	2534/5043 (50.2)
Nervous	51/332 (15.4)	1209/2574 (46.9)	64/168 (38.1)	266/929 (28.6)	96/668 (14.3)	51/372 (13.7)	1737/5043 (34.4)
Induction treatment							
Cyclophosphamide	*	395/761 (51.1)	87/149 (58.4)	734/922 (79.6)	562/661 (51.4)	294/365 (45.5)	2073/2858 (60.0)
Rituximab	*	112/761 (14.7)	75/149 (50.3)	112/922 (12.1)	177/661 (26.7)	29/365 (7.9)	505/2858 (17.7)
Plasma exchange	88/283 (31.1)	44/761 (5.8)	9/149 (6.0)	81/922 (8.9)	137/661 (20.7)	62/365 (16.9)	421/3141 (13.4)
Intravenous glucocorticoids	258/283 (91.2)	248/761 (32.6)	122/149 (81.9)	699/922 (75.8)	367/661 (55.5)	41/365 (11.2)	1735/3141 (55.2)
Maintenance treatment							

Azathioprine	*	*	26/69 (37.7)	357/857 (41.7)	322/563 (57.2)	206/331 (62.2)	911/1820 (50.1)
Rituximab	*	*	44/69 (63.8)	19/857 (2.2)	163/563 (28.9)	56/331 (16.9)	282/1820 (15.5)
Methotrexate	*	*	0/69 (0)	210/857 (24.5)	0/563 (0)	60/331 (18.1)	270/1820 (14.8)
Mycophenolate mofetil	*	*	0/69 (0)	172/857 (20.1)	122/563 (21.7)	51/331 (15.4)	345/1820 (18.9)
Outcome							
Death	56 (16.7)	350 (12.5)	3 (1.8)	113 (12.4)	127 (19.0)	187 (50.0)	836 (15.8)
ESKD ‡	49/203 (24.1)	298 (10.6)	5 (2.9)	145 (15.9)	127 (19.0)	55 (14.7)	679/5150 (13.2)
Follow-up, years	3.4 (3.3)	6.4 (5.7)	0.8 (0.9)	6.1 (5.8)	7.3 (6.8)	7.9 (6.2)	6.2 (5.8)

Data are n (%), n/N (%), mean (SD) or mean (SD; n), or median (IQR) or median (IQR; n) *Not included in registry collection. †Excluded in registry harmonisation. ‡End-stage kidney disease definitions: Czech (dialysis for >90 days; sustained CKD 5 (estimated glomerular filtration rate (eGFR)<15 mL/min/1.73 m^2) for >90 days; and/or kidney transplantation), FSVG (dialysis for more than 30 days or death within 30 days of start of dialysis), GeVas (renal replacement therapy; sustained dialysis or CKD 5 in two succeeding visits), Skåne/POLVAS (sustained dialysis), RKD (dialysis for >90 days; sustained CKD 5 (eGFR<15 mL/min/1.73 m^2) for >90 days; and/or kidney transplantation).

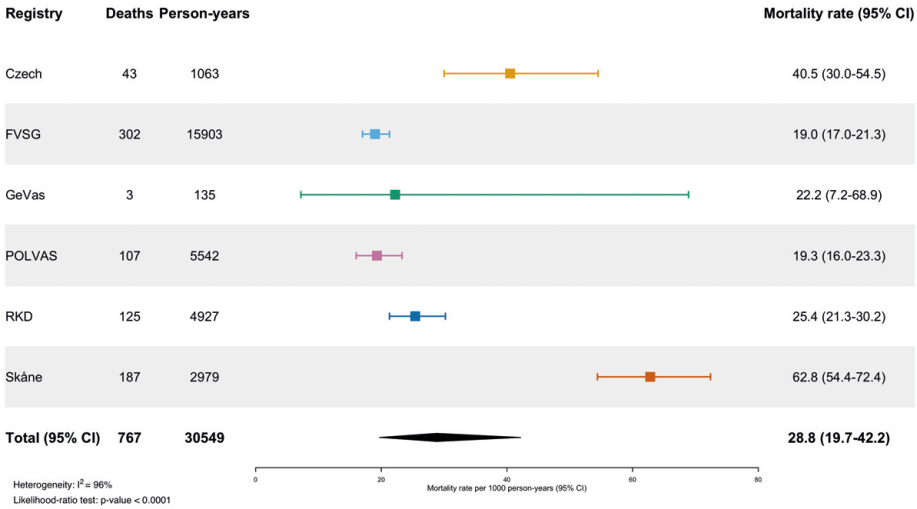


Figure 12. Mortality rate per registry and in total
All-cause mortality rate per 1000 patient years. Per registry and pooled estimates shown with 95% confidence intervals

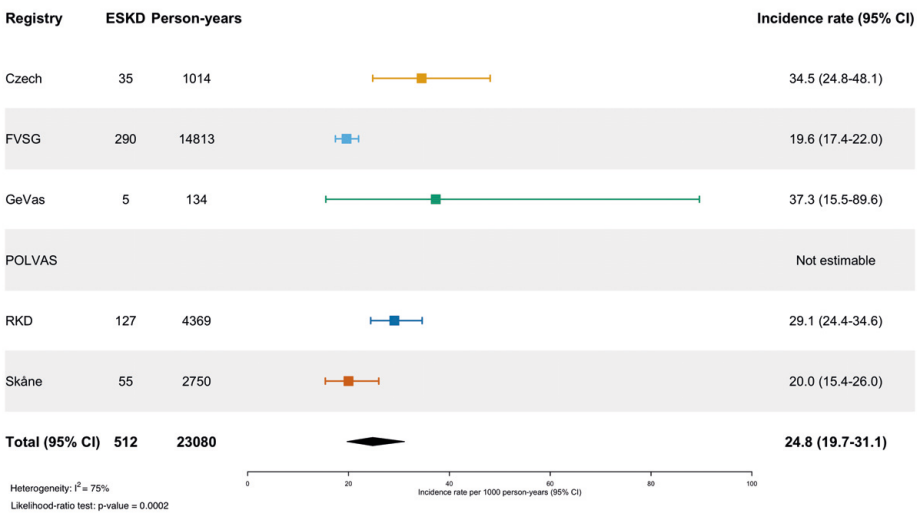


Figure 13. Incidence rate of end-stage kidney disease per registry and in total
Incidence rate of end-stage kidney disease per 1000 patient years. Per registry and pooled estimates shown with 95% confidence intervals.

Study II

In Study II we included a total of 3868 patients, diagnosed with either GPA (2434 [62.9%]) or MPA (1434 [37.1%]) (Table 6). Most cases were contributed from the FVSG registry (1780 [46.0%]), followed by the POLVAS registry (792 [20.5%]), the RKD registry (439 [11.3%]), the Czech registry (371 [9.6%]), the Skåne vasculitis cohort (351 [9.1%]), and the GeVas registry (135 [3.5%]). The mean age at the time of diagnosis was 57.2±16.4 years. There were 2006 (51.9%) men and 1861 (48.1%) women of 3867 included.

Similarly to in Study I, constitutional or musculoskeletal symptoms were most common, present in 2824 (75.5%) of 3738. The most common specific organ pattern presentation, was kidney involvement, seen in 2591 (67.0%), followed by lung symptoms in 1977 (51.1%) of 3865 and ear-nose-throat involvement 1869 (48.3%) of 3867.

Most of the included patient were PR3-positive (1949 (54.0%)), and 1344 (37.3%) MPO-positive, and 314 (8.7%) ANCA negative, of 3607 available. Concerning biochemical presentation at the time of diagnosis, the median serum creatinine and median CRP were 130.0 μmol/L (IQR:80.0-309.0) of 3521 available and 57.0 mg/L (IQR:15.0-133.0) of 2809 available, respectively.

The median follow-up was 4.2 (IQR:1.5-8.7) years; and during this time 642 (16.6%) reached ESKD and 702 (18.1%) patients died. The censoring proportion was 3166 (81.9%) for the survival analysis (all lost to follow-up). For the competing risks time-to-event analysis for ESKD the POLVAS registry was excluded, with a following sample size of 3076 for this analysis. Here, there were 502 (16.3%) events of ESKD, 463 (15.5%) competing deaths, with a censoring proportion of 2111 (68.6%) (all lost to follow-up).

In total, 3.2% of the data were missing, mostly in the variables CRP 1059 (27.4%), serum creatinine 347 (9.0%) and ANCA 261 (6.8%) variables. This data was imputed, as described in the methods section of this thesis, to generate multiple complete datasets.

Table 6. Baseline characteristics and follow-up data within the identified clusters

	SK (N = 555)	MPO-K (N = 782)	PR3-K (N = 683)	YR (N = 646)	IMS (N = 1202)
Demography					
Age, years	549; 66.7 (12.7)	776; 62.7 (13.8)	679; 55.9 (15.5)	640; 47.7 (17.0)	1,191; 55.0 (16.2)
Men	309/555 (55.7)	392/782 (50.1)	439/683 (64.3)	268/646 (41.5)	598/1,201 (49.8)
Women	246/555 (44.3)	390/782 (49.9)	244/683 (35.7)	378/646 (58.5)	603/1,201 (50.2)
Diagnosis					
GPA	211 (38.0)	185 (23.7)	570 (83.5)	530 (82.0)	938 (78.0)
MPA	344 (62.0)	597 (76.3)	113 (16.5)	116 (18.0)	264 (22.0)
Organ pattern					
Constitutional	323/526 (61.4)	248/723 (34.3)	600/676 (88.8)	249/632 (39.4)	854/1,183 (72.2)
Musculoskeletal	232/526 (44.1)	223/723 (30.8)	541/676 (80.0)	270/632 (42.7)	769/1,183 (65.0)
Cutaneous	40/555 (7.2)	97/782 (12.4)	288/683 (42.2)	135/642 (21.0)	362/1,201 (30.1)
Mucosa	4/551 (0.7)	5/774 (0.6)	53/680 (7.8)	11/634 (1.7)	34/1,192 (2.9)
Eyes	4/551 (0.7)	20/775 (2.6)	173/680 (25.4)	154/635 (24.3)	232/1,193 (19.4)
Ear-nose-throat	111/555 (20.0)	110/782 (14.1)	477/682 (69.9)	450/646 (69.7)	721/1,202 (60.0)
Lung	287/555 (51.7)	275/781 (35.2)	456/683 (66.8)	271/646 (42.0)	688/1,200 (57.3)
Cardiovascular	43/554 (7.8)	34/782 (4.3)	90/682 (13.2)	37/646 (5.7)	84/1,199 (7.0)
Gastrointestinal	33/553 (6.0)	25/780 (3.2)	118/682 (17.3)	17/644 (2.6)	80/1,201 (6.7)
Kidney	554/555 (99.8)	770/782 (98.5)	676/683 (99.0)	147/646 (22.8)	444/1,202 (36.9)
Central nervous system	13/552 (2.4)	15/770 (1.9)	66/682 (9.7)	56/636 (8.8)	141/1,194 (11.8)
Peripheral nervous system	52/552 (9.4)	32/770 (4.2)	170/682 (24.9)	95/634 (15.0)	335/1,192 (28.1)
Organs involved	555; 3.1 (1.2)	782; 2.4 (1.2)	683; 5.4 (1.4)	646; 2.9 (1.5)	1,202; 4.0 (1.5)
ANCA					
PR3/C-positive	198/532 (37.2)	131/759 (17.3)	547/665 (82.3)	340/567 (60.0)	733/1,084 (67.6)
MPO/P-positive	307/532 (57.7)	568/759 (74.8)	85/665 (12.8)	126/567 (22.2)	258/1,084 (23.8)
ANCA-negative	27/532 (5.1)	60/759 (7.9)	33/665 (5.0)	101/567 (17.8)	93/1,084 (8.6)
Laboratory					
Creatinine, µmol/L	546; 579.5 (311.3-817.0)	738; 239.5 (162.0-356.0)	657; 221.0 (154.0-364.0)	542; 80.0 (66.0-89.0)	1,038; 79.0 (65.0-90.0)

CRP, mg/L	425; 123.0 (77.0-178.0)	554; 16.0 (6.0-32.0)	544; 109.5 (50.0-190.5)	464; 6.0 (3.0-15.0)	822; 99.0 (60.0-161.0)
Outcome					
Death	169 (30.5)	161 (20.6)	158 (23.1)	36 (5.6)	178 (14.8)
ESKD*	231 (41.6)	221 (28.3)	136 (19.9)	11 (1.7)	43 (3.6)
Follow up, years	2.9 (0.6-5.8)	3.3 (1.4-6.6)	4.3 (1.5-9.0)	6.1 (2.0-11.1)	4.9 (1.8-9.6)

Data are n (%), n/N (%), mean (SD) or mean (SD; n), or median (IQR) or median (IQR; n) SK = severe kidney, MPO-K = anti-MPO kidney, PR3-K = anti-PR3 kidney, YR = young respiratory, IMS = inflammatory multisystem. *End-stage kidney disease definitions: Czech (dialysis for >90 days; sustained CKD 5 (estimated glomerular filtration rate (eGFR)<15 mL/min/1.73 m^2) for >90 days; and/or kidney transplantation), FSVG (dialysis for more than 30 days or death within 30 days of start of dialysis), GeVas (renal replacement therapy; sustained dialysis or CKD 5 in two succeeding visits), Skåne/POLVAS (sustained dialysis), RKD (dialysis for >90 days; sustained CKD 5 (eGFR<15 mL/min/1.73 m^2) for >90 days; and/or kidney transplantation).

Cluster identification

We fitted a total of ten models to each imputed dataset (1 to 5 clusters, with two types of mixture-models). We used the highest average BIC to identify the optimal model, which was seen in a five-cluster solution, retained for the reminder of Study II. The cluster assignments were unevenly distributed over the registries, with the most consistent distribution seen for the Skåne cohort.

While it should be noted that model-based clustering is not hierarchical in nature, the primary separator of cluster assignment was the presence of kidney involvement, separating the cohort into three clusters with kidney involvement and two clusters with limited kidney involvement. Of the three clusters with kidney involvement, one cluster was driven by severe kidney impairment, advanced age, and high degree of inflammation. We chose to name this, the *Severe kidney* cluster (SK). The other kidney involvement cluster, while also being driven by advanced age, saw a paucity of extra-renal involvement (especially constitutional, musculoskeletal, and ear-nose throat involvement), and a low degree of inflammation. As this cluster also saw a high prevalence of MPO-positivity we named this the *MPO kidney* cluster (MPO-K). The last kidney involvement cluster was driven by extra-renal manifestations (all organ systems, but especially constitutional and musculoskeletal symptoms), and a high degree of inflammation. As this cluster saw a high prevalence of PR3-positivity we called this the *PR3 kidney* cluster (PR3-K).

Of the two clusters with limited involvement from the kidney, both saw prevalent involvement of the upper respiratory tract, but were separated largely by the degree of inflammation, age, and the extent of organ involvement outside of the upper respiratory tract. One cluster saw younger patients, lower degree of inflammation and involvement largely isolated to the upper respiratory tract. We called this the *Young respiratory* cluster (YR). The other cluster saw greater involvement outside of the ear-nose throat region, especially constitutional and musculoskeletal symptoms, and a high degree of inflammation. Subsequently, we named this the *Inflammatory multi-system* (IMS) cluster (Figure 14).

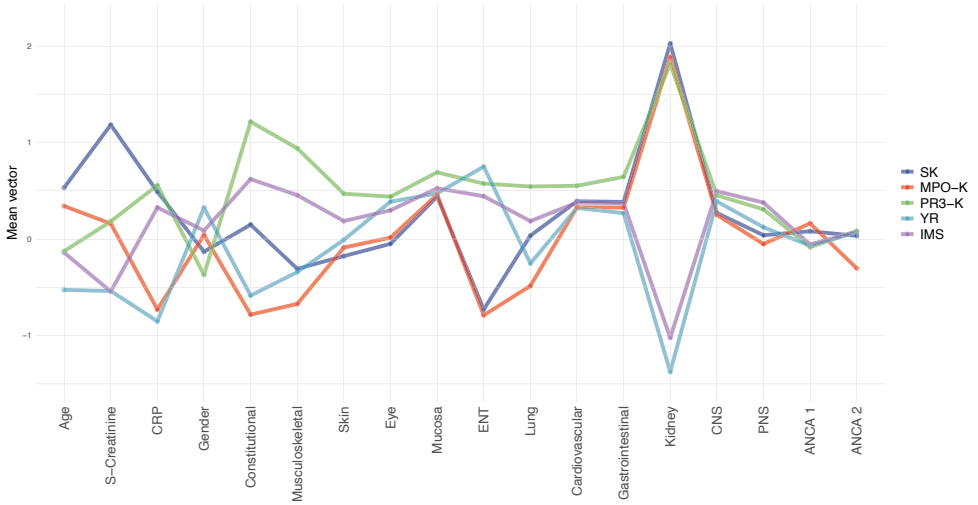


Figure 14. Mean vectors of each cluster across all variables
Illustrating the distinct patterns of variable importance and differences between clusters. Higher or lower mean vector values indicate the key characteristics differentiating the clusters. ANCA, a three nominal is here represented as two "one-hot encoded" levels, ANCA1 and ANCA2.

Cluster evaluation

To provide a clinical evaluation of our model we evaluated the five clusters with respect to survival and ESKD, with the best prognosis for both seen in the YR cluster and worst prognosis for both seen in the SK cluster (Figure 15 and Figure 16). We further compared model-fit for prediction of survival and time-to-ESKD to existing subclassification. Evaluating model-fit our cluster-membership model outperformed models fitted with diagnosis, ANCA status, and an existing cluster-stratification regarding prediction of patient survival (ΔAIC 88, 168 and 124) and kidney survival (ΔAIC 368, 399 and 186). To further allow for a fair comparison we adjusted the prediction model for age, gender, and source registry, retaining the differences for both prediction of survival (ΔAIC 34, 60 and 14) and kidney-survival (ΔAIC 349, 366 and 179). While no absolute threshold exists, a ΔAIC ≥ 10 strongly support the lower AIC model (in all cases being our identified model).

Here it should be noted, that while the description of the clusters was done with a fixed cluster assignment (the patients were assigned to the cluster with the highest average probability in the ten complete datasets following a maximum *a posteriori* estimation), the clinical evaluation allowed for floating cluster affiliations (as each patient may have been assigned to a different cluster in each imputed dataset). However, there was good agreement of subject allocation to clusters in the imputed datasets. As model-based clustering, as opposed to other prevalent clustering methods, is probabilistic, we could further assess the mean assignment probability of each cluster, here being 79.4%.

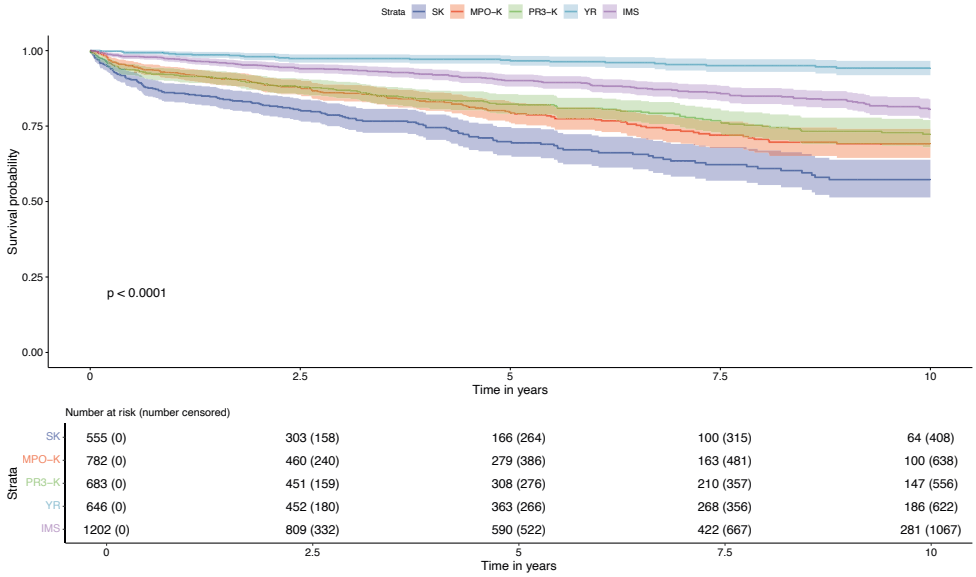


Figure 15. Ten-year survival by cluster affiliation
Kaplan-Meier curve of the ten-year survival with 95% confidence intervals. P-value through log-rank test.

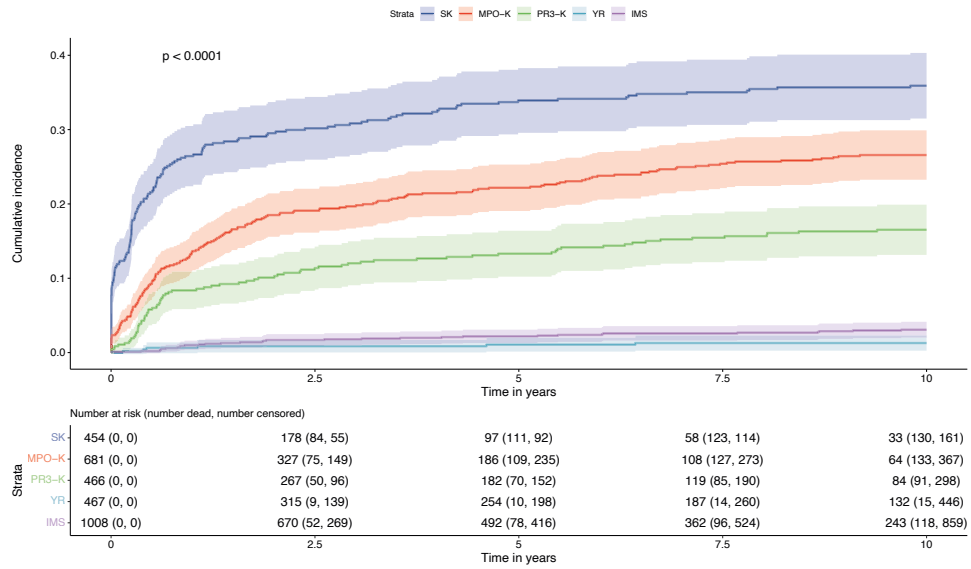


Figure 16. Ten-year cumulative incidence of end-stage kidney disease by cluster affiliation
Cumulative incidence function for the competing ten-year risks of end-stage kidney disease to death with 95% confidence interval. P-value with Gray's test.

To evaluate the robustness of the model, we assessed the cluster-wise stability using the Jaccard index. While no fixed thresholds exist, we interpreted our Jaccard indices ranging from 0.63 (SK) to 0.75 (IMS) as stable clusters. As an alternative to external validation, we performed six leave-one-registry out analyses. Here we assessed the empirical distribution function of the log-likelihoods for data simulated from the model, excluding one registry at the time. The results were stable, with model-fit improving with the exclusion of each registry (Figure 17).

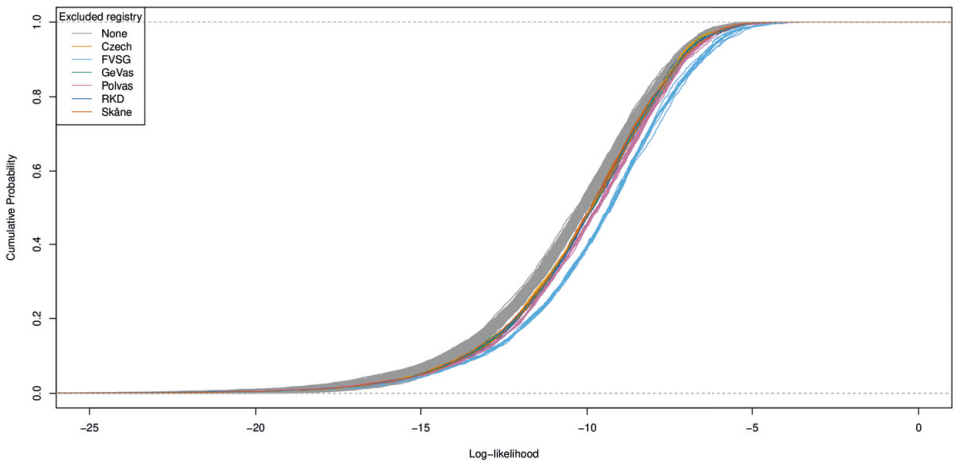


Figure 17. Empirical distribution function of the log-likelihoods for simulations of the main analysis, and six leave-one-registry out analyses
The further to the right the curve is, the better the model fit. Excluding registries improves model fit, indicating robustness of the model.

Lastly, we built the FAIRVASC clustering web application for the assignment of previously unseen data to one of the five identified clusters and made this publicly available at: https://fairvasc.shinyapps.io/clusteranalysis/.

Study III

In Study III we included a total of 2849 patients with ANCA-associated vasculitis (1726 [60.6%] with GPA and 1123 [39.4%] with MPA). A total of 353 (12.4%) of the patients faced ESKD, and 342 (12.0%) died over a median follow-up time of 4.4 (IQR: 1.5-5.0) years. A total of 2112/2498 (84.5%) reached the estimated disease remission, with 503/2112 (23.9%) of these experiencing a first major relapsing event over a median follow-up time of 5.0 (n: 2112, IQR: 2.4-5.0) years. A total of 190/736 (25.8%) patients had a first serious infectious event, over a median follow-up time of 4.7 (n: 736, IQR: 2.0-5.0) years (Table 7).

Table 7. Characteristics of the three datasets used in Study III

	Full study (n=2849)	Relapse study (n=2498)	Infection study (n=736)
Demography, registry and diagnosis			
Age, year	2,824; 58.5 (16.2)	2,473; 57.5 (16.0)	734; 62.6 (15.7)
Male	1,513/2,848 (53)	1,322/2,497 (53)	421/736 (57)
Female	1,335/2,848 (47)	1,175/2,497 (47)	315/736 (43)
Czech	157/2,849 (6)	157/2,498 (6)	0/736 (0)
FVSG	1,780/2,849 (62)	1,780/2,498 (71)	0/736 (0)
GeVas	127/2,849 (4)	127/2,498 (5)	0/736 (0)
RKD	434/2,849 (15)	434/2,498 (17)	434/736 (59)
Skåne	351/2,849 (12)	0/2,498 (0)	302/736 (41)
GPA	1,726/2,849 (61)	1,534/2,498 (61)	355/736 (48)
MPA	1,123/2,849 (39)	964/2,498 (39)	381/736 (52)
Disease activity – organ involvement			
Musculoskeletal*	1,560/2,849 (55)	1,357/2,498 (54)	364/736 (49)
Myalgia	844/2,849 (30)	698/2,498 (28)	206/736 (28)
Artrit	1,248/2,849 (44)	1,128/2,498 (45)	262/736 (36)
Constitutional*	1,603/2,849 (56)	1,407/2,498 (56)	331/736 (45)
Fever	1,002/2,849 (35)	868/2,498 (35)	179/736 (24)
Weightloss	1,182/2,849 (41)	1,084/2,498 (43)	219/736 (30)
Skin*	650/2,849 (23)	616/2,498 (25)	95/736 (13)
Infarct	59/2,849 (2)	55/2,498 (2)	9/736 (1)
Purpura	374/2,849 (13)	359/2,498 (14)	49/736 (7)
Ulcer	36/2,849 (1)	32/2,498 (1)	12/736 (2)
Gangrene	37/2,849 (1)	37/2,498 (1)	0/736 (0)
Other skin involvement	290/2,849 (10)	277/2,498 (11)	32/736 (4)
Mucosal involvement*	85/2,849 (3)	76/2,498 (3)	28/736 (4)
Mouth ulcer	76/2,849 (3)	69/2,498 (3)	26/736 (4)
Genital ulcer	5/2,849 (0)	5/2,498 (0)	0/736 (0)
Adnexal	5/2,849 (0)	3/2,498 (0)	2/736 (0)
Eye*	404/2,849 (14)	383/2,498 (15)	50/736 (7)
Proptosis	60/2,849 (2)	57/2,498 (2)	3/736 (0)
Scleritis	192/2,849 (7)	187/2,498 (7)	24/736 (3)
Conjunctivitis	137/2,849 (5)	125/2,498 (5)	16/736 (2)
Blurredvision	48/2,849 (2)	47/2,498 (2)	4/736 (1)
Loss of vision	19/2,849 (1)	18/2,498 (1)	1/736 (0)
Uveitis	26/2,849 (1)	26/2,498 (1)	5/736 (1)
Retinal changes	25/2,849 (1)	22/2,498 (1)	3/736 (0)
Ear-nose-throat*	1,271/2,849 (45)	1,137/2,498 (46)	254/736 (35)
Nasal	880/2,849 (31)	768/2,498 (31)	192/736 (26)
Paranasal	674/2,849 (24)	605/2,498 (24)	112/736 (15)
Subglottic	26/2,849 (1)	24/2,498 (1)	3/736 (0)
Conductive hearloss	236/2,849 (8)	212/2,498 (8)	55/736 (7)
Sensorineural hearloss	178/2,849 (6)	176/2,498 (7)	7/736 (1)
Non-severe lung involvement*	1,072/2,849 (38)	917/2,498 (37)	283/736 (38)
Wheeze	86/2,849 (3)	72/2,498 (3)	30/736 (4)
Nodule	618/2,849 (22)	555/2,498 (22)	116/736 (16)
Pleurisy	97/2,849 (3)	54/2,498 (2)	57/736 (8)
Infiltrate	594/2,849 (21)	512/2,498 (20)	174/736 (24)
Endobronchial involvement	32/2,849 (1)	30/2,498 (1)	3/736 (0)
Severe lung involvement*	477/2,849 (17)	447/2,498 (18)	98/736 (13)
Alveolar haemorrhage	453/2,849 (16)	425/2,498 (17)	93/736 (13)

Respiratory failure	103/2,849 (4)	97/2,498 (4)	30/736 (4)
Cardiovascular involvement*	205/2,849 (7)	188/2,498 (8)	23/736 (3)
Loss of pulses	12/2,849 (0)	10/2,498 (0)	2/736 (0)
Valvular	26/2,849 (1)	26/2,498 (1)	0/736 (0)
Pericarditis	90/2,849 (3)	84/2,498 (3)	11/736 (1)
Cardiac pain	36/2,849 (1)	28/2,498 (1)	8/736 (1)
Cardiomyopathy	22/2,849 (1)	22/2,498 (1)	2/736 (0)
Cardiac failure	46/2,849 (2)	45/2,498 (2)	2/736 (0)
Abdominal*	190/2,849 (7)	184/2,498 (7)	20/736 (3)
Peritonitis	24/2,849 (1)	24/2,498 (1)	6/736 (1)
Diarrhoea	47/2,849 (2)	42/2,498 (2)	5/736 (1)
Abdominal pain	155/2,849 (5)	153/2,498 (6)	11/736 (1)
Kidney inolvement*	1,854/2,849 (65)	1,608/2,498 (64)	570/736 (77)
Hypertension	1,854/2,849 (65)	1,608/2,498 (64)	570/736 (77)
Proteinuria	1,379/2,849 (48)	1,230/2,498 (49)	447/736 (61)
Haematuria	1,487/2,849 (52)	1,336/2,498 (53)	474/736 (64)
Creatinine change	833/2,849 (29)	673/2,498 (27)	193/736 (26)
Central nervous system involvement*	218/2,849 (8)	206/2,498 (8)	26/736 (4)
Headache	130/2,849 (5)	127/2,498 (5)	13/736 (2)
Meningitis	22/2,849 (1)	21/2,498 (1)	1/736 (0)
Confusion	16/2,849 (1)	14/2,498 (1)	2/736 (0)
Seizure	3/2,849 (0)	3/2,498 (0)	1/736 (0)
Stroke	41/2,849 (1)	37/2,498 (1)	8/736 (1)
Spinal lesion	4/2,849 (0)	3/2,498 (0)	2/736 (0)
Cranial nerve involvement	49/2,849 (2)	46/2,498 (2)	4/736 (1)
Peripheral nervous system involvement*	529/2,849 (19)	497/2,498 (20)	52/736 (7)
Peripheral neuropathy	473/2,849 (17)	451/2,498 (18)	28/736 (4)
Mononeuritis multiplex	303/2,849 (11)	291/2,498 (12)	27/736 (4)
Laboratory			
PR3/c-positive	1,355/2,640 (51)	1,167/2,289 (56)	372/736 (51)
MPO/p-positive	1,053/2,640 (40)	900/2,289 (43)	346/736 (47)
ANCA-negative	232/2,640 (9)	222/2,289 (10)	18/736 (2)
eGFR, mL/min/1.73 m^2	2,502; 50.6 (18.7, 91.0)	2,154; 51.9 (19.5, 91.5)	698; 50.8 (19.6, 86.3)
Outcome			
Death	342/2,849 (12)	233/2,498 (9)	145/736 (20)
ESKD†	353/2,849 (12)	311/2,498 (12)	106/736 (14)
Remission	2,112/2,498 (85)	2,112/2,498 (85)	383/434 (88)
Relapse	503/2,498 (20)	503/2,112 (24)	104/434 (24)
Infection	190/785 (24)	94/434 (22)	190/736 (26)
Follow up, years	2,849; 4.4 (1.5, 5.0)	2,498; 4.2 (1.4, 5.0)	736; 4.7 (2.0, 5.0)

Data are n (%), n/N (%), mean (SD) or mean (SD; n), or median (IQR) or median (IQR; n) *Composite categories (all below until next composite category). †End-stage kidney disease definitions: Czech (dialysis for >90 days; sustained CKD 5 (estimated glomerular filtration rate (eGFR)<15 mL/min/1.73 m^2) for >90 days; and/or kidney transplantation), FSVG (dialysis for more than 30 days or death within 30 days of start of dialysis), GeVas (renal replacement therapy; sustained dialysis or CKD 5 in two succeeding visits), Skåne (sustained dialysis), RKD (dialysis for >90 days; sustained CKD 5 (eGFR<15 mL/min/1.73 m^2) for >90 days; and/or kidney transplantation).

Variable importance

We identified several features of some importance for the prediction of relapse, but none stood out. For the prediction of infection, age was most important. Age and eGFR were identified as most important for the prediction of mortality and eGFR respectively (Figure 18).

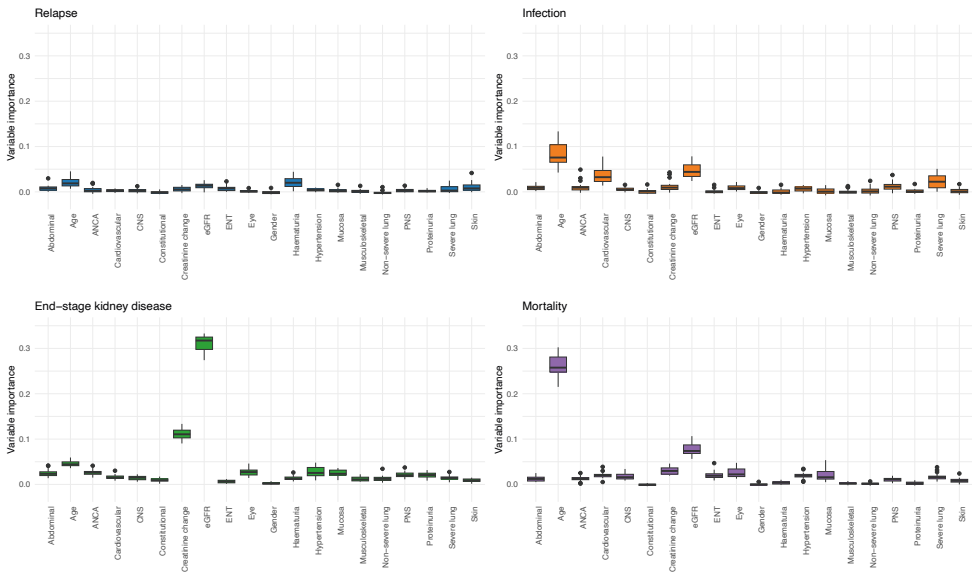


Figure 18. Variable importance per outcome as estimated by the random survival forests
Breiman-Cutler variable importance calculated using bootstrap subsampling.

When assessing the individual components of the prediction models, we saw an increased risk of relapse with the presence of skin involvement, ear-nose-throat involvement, haematuria, better kidney function, advanced age, and the absence of peripheral nerve involvement. A higher age was significantly associated with infection, while a low eGFR was the only independent predictor of ESKD. Age, cardiovascular involvement, a rise in serum creatinine or fall of serum creatinine clearance, eGFR, mucosal involvement, cardiovascular involvement, and the absence of ear-nose-throat involvement were all associated with mortality (Table 8).

Table 8. Subdistribution hazard ratios and hazard ratios of the variables in the prediction models for relapse, infection, end-stage kidney disease and mortality with 95% confidence intervals

	SHR (95% CI)	P-value
Relapse		
Haematuria	1.37 (1.11-1.69)	0.0029
Ear-nose-throat involvement	1.36 (1.13-1.63)	0.0011
Skin involvement	1.28 (1.04-1.57)	0.021
Creatinine change	1.23 (0.96-1.58)	0.095
Age*	1.13 (1.06-1.20)	0.00012
Musculoskeletal involvement	1.10 (0.91-1.33)	0.31
eGFR*	1.06 (1.02-1.09)	0.0040
Abdominal involvement	1.04 (0.74-1.47)	0.82
Peripheral nervous involvement	0.76 (0.60-0.97)	0.024
Hypertension	0.75 (0.53-1.07)	0.11
Infection		
Peripheral nervous involvement	1.53 (0.89-2.61)	0.12
Severe lung involvement	1.35 (0.89-2.05)	0.16
Age*	1.21 (1.08-1.36)	0.0016
Abdominal involvement	1.19 (0.46-3.08)	0.72
Creatinine change	1.19 (0.85-1.65)	0.31
Cardiovascular involvement	1.00 (0.41-2.45)	0.99
eGFR*	0.96 (0.90-1.01)	0.13
PR3-ANCA positive	0.79 (0.58-1.08)	0.14
Central nervous involvement	0.75 (0.27-2.03)	0.57
Eye involvement	0.63 (0.30-1.33)	0.23
ANCA negative	0.19 (0.03-1.19)	0.076
End-stage kidney disease		
Hypertension	1.18 (0.89-1.57)	0.25
Creatinine change	1.11 (0.86-1.43)	0.41
Age*	1.00 (0.93-1.08)	0.89
Eye involvement	0.92 (0.62-1.35)	0.66
Proteinuria	0.88 (0.69-1.13)	0.32
Abdominal involvement	0.78 (0.50-1.23)	0.28
Peripheral nervous involvement	0.76 (0.52-1.09)	0.14
eGFR*	0.66 (0.61-0.71)	<0.0001
Cardiovascular involvment	0.64 (0.39-1.05)	0.077
Mucosal involvement	0.39 (0.15-1.07)	0.068
Death	**HR (95% CI)**	
Age*	2.16 (1.94-2.40)	<0.0001
Cardiovascular involvement	2.06 (1.45-2.91)	<0.0001
Mucosal involvement	2.00 (1.09-3.67)	0.026
Severe lung involvement	1.58 (1.21-2.06)	0.00075
Creatinine change	1.54 (1.19-1.99)	0.0010
Central nervous involvement	1.21 (0.80-1.84)	0.37
Hypertension	1.17 (0.85-1.60)	0.33
Eye involvement	0.98 (0.64-1.49)	0.91
eGFR*	0.94 (0.90-0.98)	0.0094
Ear-nose-throat involvement	0.64 (0.50-0.83)	0.00065

*10-step increments

Outcome prediction

The predictive accuracy was good for mortality and ESKD, while the accuracy was acceptable for the prediction of serious infection but limited for relapse (Figure 19).

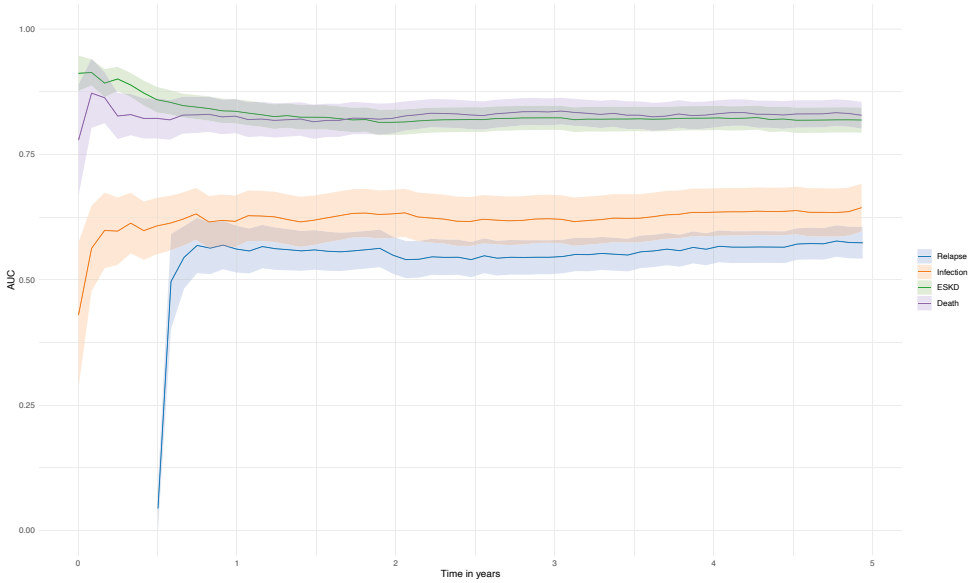


Figure 19. Model performance with 95% confidence interval over follow-up time for relapse, infection, end-stage kidney disease and mortality estimated through ten-fold crossvalidation
AUC: Area under the curve of reciever opereating characteristics evaluated monthly over follow up time. Note that the prediction of relapse starts at the estimated remission of six months. AUC of 0.5 is equivalent of chance and 1.0 a perfect predictor.

When comparing the AUC over time of our models with existing prognostic models for adverse disease progression in AAV, we saw improved predictive accuracy for both relapse and mortality. No comparators were available for infection and ESKD. For these models we evaluated the predictive accuracy for serious infection in patients known to receiving prophylaxis with trimethoprim-sulfamethoxazole, and for ESKD in patients known to receiving plasma exchange, with acceptable results (Figure 20).

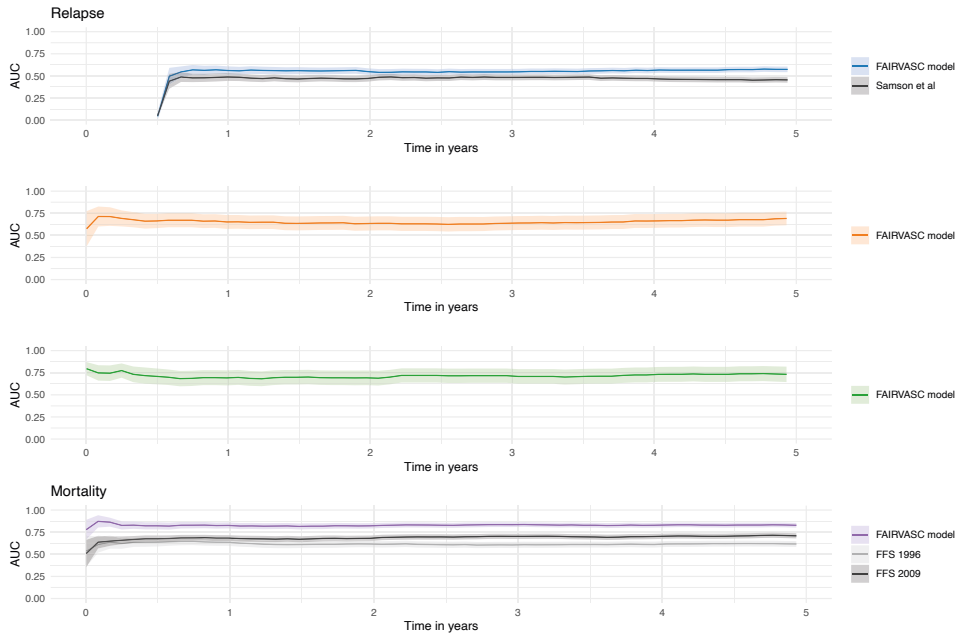


Figure 20. Comparison to existing models (relapse and mortality) and sensivity analyses (infection and end-stage kidney disease)
Relapse: The model presented in Study III and a model developed by Samson and colleagues.[303] Infection: The model presented in Study III tested in a subset of of 301 patients on trimethoprim-sulfamethoxazole. End-stage kidney disease: AUC: The model presented in Study III tested in a subset of of 181 patients receiving plasma exchange. Mortality: The model presented in Study III and Five Factor Scores developed by Guillevin and colleagues.[286, 287]

When estimating the applicability of implementation of the models for clinical practice we noted clear net benefit of model use for guiding intervention in relapse, infection and mortality, and some, although limited, use for ESKD guidance, with the pre-identified thresholds. The final prediction models were made available for implementation to new data at: https://fairvasc.shinyapps.io/prognosticmodels/ (Figure 21).

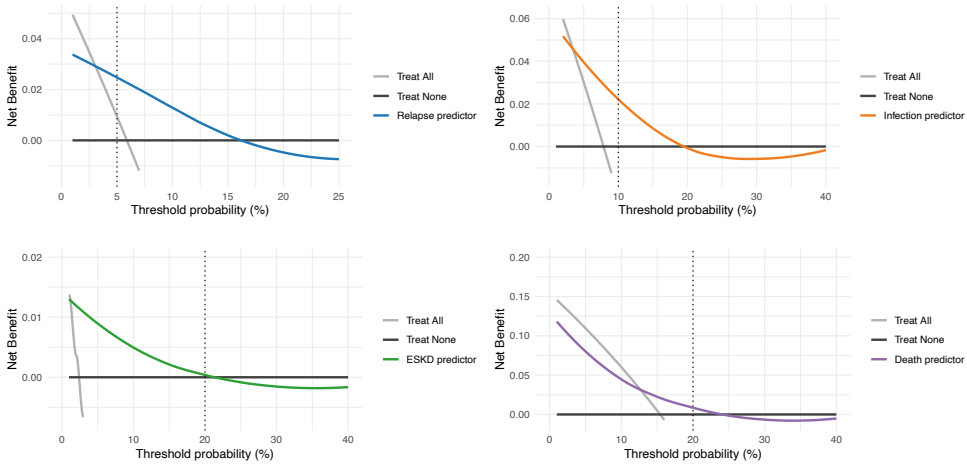


Figure 21. Decision curve analysis for five-year prediction of relapse, infection, end-stage kidney disease and mortality

Decision curve analysis calculates the net benefit by putting false positives (harm) on the same scale as true positives (benefit). False positive rates are multiplied by an exchange rate (how many false positives are worth one true positive) defined by a probability threshold. Our proposed thresholds, while arbitrary, for decisions are presented with dashed lines. Relapse: early discontinuation of immunosuppressive therapy below 5% five year risk. Infection: prolonged infection prophylaxis above 10% five year risk. End-stage kidney disease and mortality: general management guidelines above 20% five year risk. Treat All: All patients receive the intervention, Treat None: No patients receive the intervention, Outcome predictor: Our respective FAIRVASC model is used to guide the decision to intervene. Our thresholds are presented as dashed lines, and at these thresholds all models show net benefit.

Discussion

Data and registry interoperability to address fragmentation in rare disease research is a stated strategy by the European Union.[304] Here FAIRification (the implementation of the FAIR principles) is a key concept but what it entails, and how FAIR data is achieved, remains somewhat elusive. The original paper presented by Wilkinson and colleagues do not define the technical infrastructure, but GO FAIR, an initiative for the implementation of the FAIR principles from the group of authors behind the original paper, makes it clear that FAIRification is achieved through Semantic Web technologies.[288, 305]

As such, the work presented in Study I of this thesis aligns with GO FAIR and is an example of successful FAIRification. Through the implementation of a FAIR Data Point the FAIRVASC project allowed for interoperability of metadata (data about the data) which was not presented in this thesis, while the dedicated web interface presented here allowed for interoperability of the actual data.[a]

Prior to this, research data in vasculitis have been siloed. In Study I we present data of integrated non-siloed registries, but the model presented is also scalable. With the scope defined by competency questions, the model was designed primarily for the purpose of integrating the six participating registries. However, since the publication of Study I, one additional registry is queryable, the Italian vasculitis registry (Italvas), highlighting that the process is indeed extendable.

Yet, the technical implementation is not without limitations. Federated analytics, privacy-preserving data analytics over multiple remote datasets, is a rapidly evolving field. While there exist many levels of privacy, in Study I, we wanted to avoid subject level data leakage between the sites, and between the sites and the aggregation server. This was achieved through the blockade of subject level data and low cell-counts at query level. To facilitate the scalability of the project, and to ensure easy development, the infrastructure in Study I was built essentially using native SPARQL. SPARQL is a query language (largely analogous to SQL for relational databases) for the retrieval and manipulation of data, not statistical calculation. As such, there were considerable limitations to the federated analytics in this study. However, extensions to SPARQL, not used here, may facilitate more complex analyses of data.

[a] FAIRVASC FAIR Data Point: https://fairvasc-fdp.adaptcentre.ie/ (accessed: 2024-09-26)

Another potential limitation present in research is data quality. While high-quality data are essential to allow reliable decision-making, there are a paucity of formalised frameworks for the assessment of data quality, and no agreed thresholds recognised for what constitutes sufficient data quality. In Study I, we assessed data quality using the i~HD framework, seeing it being imperfect in all assessed domains. Structural data quality concerns (such as uniqueness, consistency, and to some extent completeness) can be technically addressed through data validation systems. Straightforward (e.g., limitations to the allowed data format or data range) validation systems are already in place in the participating registries, and quality for this type of data are excellent. When adding a level of complexity (e.g., is the data inherently coherent), the data quality worsen as such validation systems are not in place. Still, such data quality concerns are technically easy to address.

However, while missing data are inevitable in real-life datasets, the differences of the registry data and what is reported in the electronic health record (that is correctness) seen in Study I is concerning. These discrepancies may to some extent be explained poor variable definitions and be solved through making what input is requested clearer but may also simply reflect the human error (i.e., a mistake was made at data entry which did not violate the data validation barriers). Here automated input from electronic health records is the obvious solution, but then raises the question: Why do we need a registry in the first place?

While both registries and electronic health records systems are for collecting and storing data, their purpose are different: registries are for the collection of data regarding a specific patient population minimised for research, whereas the electronic health record are comprehensive systems for managing individual patient care. That said, this does not mean that electronic health records cannot be used for research (there is enormous potential in this largely untapped collection of data), but constraints around governance and patient-privacy are vast. Here registries serve a purpose, especially in rare diseases, where scarce data may be systematically organised. Registries may be used to address several questions of relevance in rare diseases, examples being related to the epidemiology, symptomatology, prognosis, real-world treatment data, and equity of care.

In Study I some of the above are addressed, primarily to show the feasibility of the approach. However, the results still provide insights, showcasing the to date, largest cohort in AAV. Firstly, the study demonstrates the considerable heterogeneity of AAV, and of AAV cohorts. AAV presents with a wide range of symptoms and severity levels, making the generalisability of studies highly sensitive to the inclusion criteria and recruitment setting. While the differences of observational cohorts have previously been described, this study convincingly demonstrates that the generalisability of any single centre or registry effort may be limited.

Secondly, Study I give indications of differences in care. Again, the comparison of is hampered by the above-described heterogeneity but also the incompleteness of

treatment data. However, some differences, in particularly regarding the use of intravenous methylprednisolone pulses, warrants discussion. The evidence-basis for its use is thin, but recommended in cases of organ- or life-threatening disease.[179] However, the use may be at the cost of higher rates of infection and diabetes.[306] While the difference is inadequately investigated for an extended discussion, it highlights how federated registries may be used for comprehensive benchmarking and improve equity of care.

Lastly, we studied the outcome of AAV patients regarding mortality and ESKD. Using a meta-analytical approach of random effects, we again saw considerable heterogeneity in the rates of both. This is likely partly due to under-reporting of outcome. The Skåne Vasculitis cohort, exhibiting the highest mortality rate, is also the only registry with complete mortality data ascertainment through linkage with official registers. Other issues may be inherent in the registry designs. Allowing for retrospective patient recruitment introduces both a potential survival bias (only those who survived long enough to be recruited are included) and an immortal time bias (the survival bias introduces an issue in the investigation of time-to-event analysis from the date of diagnosis, simply that a patient cannot experience the event of interest before the recruitment to the registry).

Many of the issues described, regarding for example completeness and heterogeneity, can (to some extent) be statistically addressed. Federated learning, that is the training of machine learning algorithms collaboratively without exchanging the data itself, harbours the potential to solve some of these problems.[307] While the field of federated learning is rapidly expanding, much of the implementation is still exploratory. As such, further investigations in this thesis used a centralised data-pool, to explore the phenotypic spectrum of AAV, and build prognostic models for adverse disease progression.

In Study II we explored and stratified patterns of symptoms in 3868 patients with AAV using model-based clustering. Exploring disease heterogeneity using unsupervised learning is not unprecedented in AAV, nor in systemic rheumatic diseases in general. In AAV, several studies have been made but only two have attempted to cover the full spectrum of disease at the time of diagnosis, one using multi-national RCT data (Mahr and colleagues), and one Japanese observational cohort study (Watanabe and colleagues).[41, 308] Of these, Mahr and colleagues' subclassification into five phenotypic groups (renal AAV with PR3-ANCA, renal AAV without PR3-ANCA, non-renal AAV, cardiovascular AAV, and gastrointestinal AAV) has gained most traction.[b]

Interestingly, both our study and Mahr and colleagues give limited importance of ANCA type in the stratification of patient symptomatology. Instead, much like in

[b] Searching Scopus, the work Mahr and colleagues is cited 201 times, as compared to five times for the work by Watanabe and colleagues. (accessed: 2024-09-01)

the model developed by Mahr and colleagues the main driver of separation in our model was kidney involvement. Here are differences in the rate of kidney involvement between the studies worth discussing, being 67% and 85% in ours and Mahr and colleagues' study, respectively. This reflects a general difference between real-world observational data and RCTs, described already in a comparative study by Pagnoux and colleagues.[309] The inclusion criteria for RCTs in AAV often require, or highly favour the inclusion patients with kidney involvement, explaining the difference seen. Evident is also, that severe kidney impairment may arise in both PR3-positive and MPO-positive (and even ANCA-negative) patients.[187] Given the significant impact of impaired kidney function on patient-outcome, a stratification based on kidney involvement has face validity.

The five identified clusters have distinct phenotypic and biochemical presentation, as well as disease outcome. Linking the five-cluster stratification to defining genetic, epigenetic, or other pathogenic precursors is hampered by the paucity of studies stratified by phenotypic expression. Naturally, investigations into the pathogenesis of AAV have been largely focused on the disease groups or ANCA patterns. The subclassification we present highlights the need for studies on the role of inter-organ vascular diversity, systemic inflammatory response, and ageing in phenotypic expression.

As opposed to supervised learning the purpose of unsupervised learning models is not primarily to be applied to new datasets, but to find structure in the data at hand. A model-based clustering model is fitted using an expectation-maximisation (EM) algorithm, consisting of two iterative steps, the E-step (where the probability of each data point belonging to each cluster is calculated using the current parameter estimates), and the M-step (where the model parameter estimates are updated to maximise the expected log-likelihood based on the probabilities from the E-step). To allow for application of the model to new data, we separated the E-step of the final model so that the probability of new datapoints belonging to each cluster could be given using the final parameter settings. Additionally, as probabilities will be given regardless how well the input fits the model, we implemented a flagging system for poorly fitting data.

As the primary function of the clustering algorithm is to find structure in the data at hand, the generalisability of the model is highly dependent on the source cohort. Here the multi-centre real-world setting is a considerable strength, with the cohort reflecting the full spectrum of AAV better than RCT data or any single-centre effort. However, more important is: Is the subclassification relevant?

Some assurance of the relevance of the subclassification may be given by studying and comparing patient outcome. We see distinct outcomes in the five clusters, and when comparing our subclassification with Mahr and colleagues, the clinical diagnosis, and the ANCA-pattern, our models have the highest predictive accuracy. With the estimation of prediction error penalising the complexity of the model, this

study challenges the view of being AAV adequately represented by a binary subclassification. While this notion is not novel, the design and presentation of scientific studies have persisted in a binary subclassification of AAV, potentially impeding our further understanding of the disease. As such, the results of Study II may have significant impact on how AAV is perceived, discussed and presented.

However, Study II is not without limitations. In the assembly of the large cohort from multiple registries there have been inevitable trade-offs between sample size and granularity. There is an absence of data of interest in the subclassification of disease, such as biomarkers, radiographic data (i.e., pulmonary fibrosis data), and histology. It should further be noted that this subclassification is based on a snapshot of data at the date of diagnosis. There is a considerable diagnosis delay in AAV, the extent of which is unknown in this study. Similarly, there is limited follow up data. As such, how the phenotypic expression and cluster affiliation evolve before and after diagnosis would be of interest to investigate further.

While Study II is an exploration of the symptomatology at diagnosis Study III concerns the trajectory of AAV. Here we built and assessed predictive models for adverse disease outcome (i.e., major relapse, serious infection, ESKD and mortality) in AAV, using clinical and laboratory data commonly collected and recorded at the time of diagnosis.

There is a paucity of prognostic models to guide clinical decision making in AAV. However, as a disease with heterogenous disease presentation, prognostication and risk factor identification using clinical and biochemical data available at the time of diagnosis have been repeatedly investigated, but rarely compiled to comprehensive models. Predictive models may lay foundations for new clinical management algorithms and risk profiles, which can drive healthcare savings and improve patient quality of life and ideally provide further understanding of the underlying disease processes. In the following sections we will discuss the four outcomes separately. It should however be noted that while we try to discuss how the predictors influence the model, the causality of relationships is not assessed.

As discussed in the *Introduction* of this thesis disease relapse is major concern in AAV, but the exact rate of relapse is highly variable between studies.[274] We here see a major relapse occurring in 24% within the first five years following diagnosis. We further note that we do not see a separation in the rate of relapse between PR3-positive and MPO-positive patients. Instead, the variable of associated with the greatest hazard ratio is haematuria. Persistent haematuria following induction treatment has previously been described as a predictor of relapsing disease, especially kidney relapse, likely reflecting smouldering disease activity.[310] However, when further investigating the hazard ratios associated with the included variables in the model, skin involvement, ear-nose throat involvement, better kidney function, advanced age and the absence of peripheral nerve involvement all show significant association, all (but the last mentioned) also previously described.[274, 311]

Our clinical prediction model is however flawed in terms of accuracy. Despite this, it outperforms the only existing clinical prediction model for relapse developed, highlighting the need for external validation in model development.[303]

An obvious limitation to our prediction model is the absence of treatment data, as the choice of induction and maintenance treatment influence the rate of relapse.[274] Another is the estimation of disease remission at six months following diagnosis. While not an uncommon definition for observational research, not all patients are in a stable remission at this time point.[312] To avoid the inclusion of smouldering disease activity we further included only major disease activity. Regardless, it is evident from our relapse prediction model, that any attempt at accurate prediction of relapse needs a longitudinal component, or a robust biomarker, that was not available in this study. However, while imperfect, the model in this study may provide some guidance for decision making.

The second outcome of interest, serious infection, is common in AAV but like in relapse the exact rate is cohort-dependent.[242] In Study III we see a hospital-care requiring infection in 26 % within the first five years following diagnosis. Again, several risk factors have been identified, but no comprehensive model for prediction of risk developed. Here we identify advanced age as the variable of most importance for the prediction of serious infectious events, which is in line with the existing literature.[241] The predictive accuracy of our model is acceptable, and we estimate a net-benefit to guide a decision-making at reasonable risk-thresholds. However, the prediction model has limitations. The model is far from perfectly accurate and tends to overestimate the risk of serious infection. Like for relapse, the absence of treatment data is a considerable limitation. The increased incidence of serious infections in AAV compared to the background population is logically a consequence of the immunosuppressive treatment, and today prophylaxis using trimethoprim–sulfamethoxazole is recommended.[179] In our cohort, we could, due to source registry design and the design FAIRVASC ontology, not separate those who received and not received prophylaxis. However, as a sensitivity analysis we could test the model in a subset known to have received prophylaxis, retaining predictive accuracy.

Unsurprisingly, the eGFR at diagnosis is of considerable importance for the prediction of ESKD.[313] While the absence of histopathological data prohibits any comparison to existing state of the art models for the prediction of kidney failure, our model based solely on clinical and laboratory data predicts time to ESKD with good accuracy. We further see a net-benefit for model-use in guidance of decision-making at reasonable risk-thresholds, however for this model the decision to be made (and threshold) is ill-defined beyond general guidance. As with the other models the absence treatment data is a considerable limitation. Currently, plasma exchange may be considered in high-risk patients (i.e., with a serum creatinine >300 µmol/L), with current evidence suggesting a reduced risk of progression to ESKD following its use.[179] Again, we could not separate those who were treated and

not treated with plasma exchange. However, as a sensitivity analysis we could test the model in a subset known to have undergone plasma exchange, seeing acceptable performance also in this subset.

As can be expected age is a strong predictor of mortality in AAV and is the main driver of our mortality prediction model. This model shows good predictive accuracy, outperforming existing mortality-prediction models such as both the original and updated Five Factor score. However, as a model for the prediction all-cause mortality its usefulness in guiding clinical decision-making is limited, as the intervention is arbitrary. It does, however, much like the Five Factor scores provide a formalised assessment of general risk. Other factors apart from age, significantly associated with time-to-death are, the absence of ear-nose-throat symptoms, a low eGFR, a recent deterioration of kidney function, cardiovascular involvement and severe lung involvement, all previously described in the literature.[229, 285, 287] We also see an increased hazard ratio with mucosal involvement, but the presentation is uncommon and consequently the confidence interval wide.

Study III was designed without sample size estimation and we did not penalise the complexity of any of the models. However, the sample size is a considerable strength, as is the real-world type and commonly collected data, all to some extent ensuring the generalisability of the models. While the limitations in terms of a paucity of longitudinal data and treatment information have already been described, the major limitation is the lack of external validation. For implementation in clinical practice the results should be validated in external cohort and the implications of implementation further assessed through thorough decision-analytic techniques.

Both in the separation of phenotypic expression (Study II) and the prediction of outcome (Study III) the diminished importance of the ANCA serotype is striking. The ANCA serotype is widely regarded as a major deciding factor both in subclassification of disease, and in the prognostication of outcome. While this of course may be inherent to the FAIRVASC cohort, the multi-registry design should rather increase the generalisability. However, neither studies presenting the ANCA type as a risk factor nor our studies rest on causal frameworks. Though hard to appropriately design, given our limited understanding of the underlying disease processes, causal investigations of exposures and outcome would be of considerable interest.

Herein lies the true potential of the integration of data and large sample sizes presented in this study. With the digitisation of research data, and the development of complex statistical or machine learning modelling, the potential to understand AAV has never been greater. Through the expansion of the technologies used here we may hopefully be closer to understand the *whys* of vasculitis and autoimmunity. *Why* do adverse events occur, or *why* does an individual present with a particular symptomatology, or ultimately *why* did that individual develop vasculitis? Somewhere, in the answers to these whys, we might also find the *hows* of vasculitis

and autoimmunity. *How* can we cure vasculitis, and ultimately *how* can we prevent it?

Conclusion and implications

We can here conclude that fragmented and siloed registries and cohorts in AAV can be integrated and successfully queried using a framework extendable also to other diseases suffering data fragmentation. However, we note that data quality is variable and needs to be systematically addressed to facilitate federated research. While the inherent differences of the included data somewhat hamper the interpretation and comparison of results, the successful technical integration demonstrate a possibility for a new privacy-preserving federated paradigm of rare disease research.

Still, there are currently technical and statistical limitations to data federation, necessitating traditional centralised data pooling. Exploring and stratifying the symptomatology of thousands of patients with AAV at the time of diagnosis, using unsupervised machine learning over such a data pool, we can further reinforce that AAV is beyond a binary construct. Instead, we suggest a data-driven subclassification based largely on the presence and severity of kidney involvement, and the extent of extra-renal manifestations to stratify patients into five subgroups, displaying distinct phenotypic expression and disease outcome. We further conclude that accurate prediction of key adverse outcome is possible in AAV, using data that is commonly collected at the time of diagnosis. However, the benefit of implementation for the guidance of clinical decision-making needs further evaluation.

While the true implications of these results are yet to be seen, the studies presented in this thesis give an indication of the new era in rare disease research in general, and AAV in particular. An era of interconnectivity. Here, the first wobbling steps of interconnected data are presented, but in the future access to large sample sizes may provide possibilities for the development of diagnostic and prognostic biomarkers, accelerate the development of new therapies, and improve health equity through benchmarking of patient care.

Acknowledgements

Behind all the numbers and models presented in this thesis are real people with ANCA-associated vasculitis. First and foremost, I would like to thank them. To describe a life as a date of birth, a date of diagnosis, and a date of death is of course folly. Between these dates a life happens, good and bad. But perhaps this thesis, somewhere in the long run, may help someone to have a few more of the good days.

In addition, there are many people without whom this work would never have been possible.

Aladdin J Mohammad, my supervisor. Six years ago, I knocked on the door of your office at the department of Rheumatology. I was struck immediately by your enthusiasm and knowledge. You helped me, a young man you had just met to fulfil a childhood dream of doing research at Cambridge. Working with you for the last four years, have meant me learning about vasculitis and epidemiology from the very best. You have opened the door for me to the vasculitis research community and helped me travel the world. Your generosity with time, money, and wisdom never seizes to amaze me. Thank you!

Mark A Little, my co-supervisor. A visionary and a renaissance man. If I could muster even a tenth of your curiosity and drive, I'd be well off.

Mårten Segelmark, my co-supervisor. Ever knowledgeable, from the smallest cytokine to the global epidemiology. What you don't know about vasculitis is not worth knowing.

Arthur White, my co-supervisor. What would I have done without you. Thank you! And you know what, I said I would never bother, but I'm great at writing for-loops now.

Everyone in the FAIRVASC consortium. I would also like to give a special thanks to Beyza Yaman, Dagmar Wandrei, Kris McGlinn, Matthew Rutherford, and Michelangelo Tesi. My digital company during the Covid-years, and the cogs that made this project turn. That we made this happen from our own bedrooms, kitchens, and living rooms is truly a feat.

Everyone in the Lund Vasculitis Epidemiology Research Group. I would like to give a special thanks to Jens Rathmann. A true inspiration as a person, researcher, and clinician. You're always warm and always have time. Your support and feedback have been so valuable. Raïssa de Boer, my closest colleague and company during

some rough and uncertain years. You taught me everything I know about research methodology and R, for that I am always grateful.

Everyone at the Rheumatology clinic in Lund. Thank you for letting me sneak around in the background and occasionally show up in scrubs. One day I will be back for real!

Och alla mina vänner. Jag lovade att jag skulle få med det någonstans, så här kommer det: Låt inte det bästa vara det möjligas fiende.

Och mamma, pappa, Sofia och Emily och alla som inte längre är med oss. Tack! Det är 215 år sedan en Gisslander senaste försvarade en avhandling vid Lunds universitet. Jag hoppas på ett bättre öde än stackars Alexander.

Och Sillen, så mycket som du promenerat på mitt tangentbord har du uppfyllt kraven för medförfattarskap. Trots detta får du nöja dig med ett omnämnande här. Jag hoppas du kan förlåta mig för det.

Och slutligen Linn. Min stora kärlek. Utan dig hade det här aldrig gått.

Populärvetenskaplig sammanfattning

Vaskuliter är sjukdomar som kännetecknas av inflammation i något eller några av kroppens blodkärl. Då det i kroppen finns både stora och små blodkärl delas vaskuliterna ofta upp beroende på de huvudsakligen drabbade kärlen, i stor- och småkärlsvaskuliter. Hos patienter med småkärlsvaskulit, uppvisar vissa en speciell sorts antikropp vid blodprovstagning. Den speciella typen av antikropp kallas för ANCA (en förkortning för anti-neutrofil cytoplasma-antikropp). Antikroppar är en del av kroppens immunförsvar och bekämpar i vanliga fall främmande ämnen. Just ANCA däremot, riktar sig mot strukturer på och i kroppens egna vita blodkroppar.

Eftersom det finns små blodkärl i alla delar av kroppen kan en såkallad ANCA-associerad vaskulit (AAV) yttra sig på många olika sätt. Inflammationen i blodkärlen gör att det finns en stor risk att det organ som drabbats skadas allvarligt och permanent, men även att den som drabbats av AAV dör. Därför måste AAV behandlas. Detta görs genom att dämpa inflammationen med en rad läkemedel såsom kortison, immunhämmande läkemedel, och låga doser cellgifter. Med modern behandling har prognosen förbättrats avsevärt men sjukdomen kan inte botas och behandlingen är ofta livslång. Det finns en betydande risk för att drabbas av återkommande sjukdomsskov och de immunhämmande läkemedlen ökar risken för allvarliga infektioner.

Som tur är, är AAV en sällsynt sjukdom. I Sverige idag drabbas cirka 30 personer per miljon vuxna och år. Att sjukdomen är sällsynt innebär dock att den är svår att studera. För att kunna dra säkra slutsatser krävs ofta stora patientunderlag, något som kan vara svårt att uppnå i ett enskilt land. I den här avhandlingen beskriver jag därför ett forskningsprojekt där patientdata från flera europeiska länder har sammanlänkats för att kunna studeras som grupp. Med hjälp av denna stora mängd data har jag sedan djupdykt i hur sjukdomen yttrar sig när den diagnosticeras, men även försökt bygga modeller som kan hjälpa sjukvården att utvärdera risken för att en patient ska drabbas allvarliga komplikationer av sin vaskulit.

Studie 1 i den här avhandlingen beskriver hur vi sammanfört patientdata från sex europeiska länder. Jag beskriver även hur kvaliteten på data är, genom att undersöka att det inte finns dubbletter, att data är i korrekt format, att det inte saknas relevant data och att uppgiven data överensstämmer med journaluppgifter. Jag granskar sedan med vilka typer av läkemedel patienter med AAV behandlas runtom i Europa,

samt hur många som drabbas av permanent och dialyskrävande njurskada och hur många som dör.

Att sammanlänka data från olika databaser är inte helt lätt, då variabler och definitioner kan variera kraftigt. För att säkerställa att de olika databaserna pratar om samma sak har det krävts ett extensivt harmoniseringsarbete. Den här studien beskriver ett arbetssätt för denna harmonisering som är överförbart till andra sjukdomstillstånd i behov sammanlänkning. Slutprodukten är en hemsida, från vilken en forskare kan ställa enkla och säkra frågeställningar utan att data lämnar den databas i vilken den befinner sig.

Jag noterade dock i studien vissa brister i datakvaliteten, speciellt att data i vissa fall inte överensstämde med uppgifter i patientens medicinska journal. Jag diskuterar sedan vissa metoder för att säkerställa högre datakvalitet. Totalt sammanlänkade jag i denna studie över 5000 patienter med AAV över sex patientregister. I studien noterade jag att det var stora skillnader gällande ålder vid insjuknande, symptombild, behandling och risk för njursvikt och död mellan de olika databaserna. Detta synliggör att behovet är stort av sammanlänkning för att fånga hela det sjukdomsspektrum som finns vid AAV, och att studier gjorda med data från endast en databas bör tolkas med försiktighet.

Eftersom det finns blodkärl överallt i kroppen varierar symptomen av AAV kraftigt från person till person. Vi vet idag inte varför någon drabbas av AAV, men inte heller varför den som drabbas, drabbas i just de blodkärlen och får den symptombilden som de får. Traditionellt klassificeras patienter med AAV i två subgrupper granulomatos med polyangiit (GPA) och mikroskopisk polyangiit (MPA), med olika symptombild. Denna uppdelning är av vikt då den används vid jämförelser i olika medicinska studier, men även då prognosen verkar skilja sig mellan grupperna. Vidare har patienter med AAV olika typer av ANCA, som riktar sig mot olika enzymer i de vita blodkropparna. Beroende på om en patient har antikroppar mot enzymet proteinas-3 (PR3-ANCA) eller myeloperoxidas (MPO-ANCA) verkar också symptom och prognos skilja sig.

I studie 2 använde jag mig av vårt stora patientunderlag för att undersöka hur AAV yttrar sig och med hjälp av maskininlärning gruppera symptombilder som ofta sammanfaller. Jag fann fem typer av patienter, som skiljde sig gällande symtom, blodprover och prognos. Dessa patientgrupper skiljer sig huvudsakligen baserad på graden av inflammation, antalet drabbade organ och påverkan på njurfunktionen.

Förhoppningen är att vi vidare ska kunna undersöka om dessa grupper skiljer sig gällande svar på behandling, sjukdomsmekanismer och genetik. Detta kräver dock tillgång till ytterligare data.

Prognosen för patienter med AAV skiljer sig kraftigt, men att förutsäga vem som kommer att drabbas av olika sjukdomsutfall är svårt. I studie 3 bygger vi modeller för att försöka förutsäga risken för att en patient med AAV ska drabbas av en

allvarlig infektion, ett allvarligt återfall i sin sjukdom, dialyskrävande kronisk njursvikt eller död inom fem år efter sin sjukdomsdebut. I studien drabbades totalt 26% av en allvarlig infektion, 24% av ett allvarligt sjukdomsåterfall, 12% av dialyskrävande kronisk njursvikt och 12% dog.

Jag noterade ett antal faktorer vid tiden för diagnos som var till hjälp för att förutsäga om en patient skulle drabbas av någon av de ovan nämnda utfallen och sammanställde dessa till sannolikhetsmodeller. Vi undersökte sedan statistiskt om modellerna skulle kunna vara till hjälp för exempelvis en läkare vid beslut kring uppföljning och behandling, och såg att så var fallet. Resultaten behöver dock säkerställas ytterligare innan användning i praktiken rekommenderas.

För att sammanfatta beskriver den här avhandlingen hur vi sammanlänkar olika europeiska databaser med uppgifter om patienter med den allvarliga sjukdomen AAV. Detta är av stor vikt då sjukdomen är sällsynt och antalet patientfall få i varje enskilt land. Stora patientunderlag kan ge oss nya insikter om hur AAV uppstår, yttrar sig och hur prognosen ser ut. Med hjälp av sammanlänkade data fokuserar den här avhandlingen sedan på hur symptombilden ser ut och kan grupperas. Vidare bygger jag modeller för att förutsäga risken för att drabbas av allvarliga komplikationer till följd av sjukdomen. Förhoppningen är att dessa studier ska verka som underlag för vår vidare förståelse av AAV, men också vara ett steg mot en mer individanpassad behandling.

References

1 Davies DJ, Moran JE, Niall JF, et al. Segmental necrotising glomerulonephritis with antineutrophil antibody: possible arbovirus aetiology? *Br Med J (Clin Res Ed)* 1982;**285**:606.

2 Ebbell B. *The Papyrus Ebers: the greatest egyptian medical document*. Copenhagen: Levin & Munksgaard 1937.

3 Rokitansky K. *Über einige der wichtigsten Krankheiten der Arterien*. Kaiserlich-Königliche Hof- und Staatsdruckerei 1852.

4 Kussmaul A, Maier R. Uber eine nicht bisher beschriebene eigenthumliche Arterienerkrankung (Periarteritis nodosa), die mit Morbus Brightii und rapid fortschreitender allgemeiner Muskelahmung einhergeht. *Deutsche Archiv Klinische Medizin* 1866;**1**:484-518.

5 Ferrari E. Ueber Polyarteritis acuta nodosa (sogenannte Periarteritis nodosa), und ihre Beziehungen zur Polymyositis und Polyneuritis acuta. *Beitr Pathol Anat* 1903;**34**:350-86.

6 Wohlwill F. Uber die nur mikroskopisch erkennbare Form der Periarteriitis. *Virchows Archiv für pathologische Anatomie und Physiologie und für klinische Medizin* 1923;**246**:377.

7 Klinger H. Grenzformen der Periarteritis nodosa. *Frankfurt Z Path* 1931;**29**:455-80.

8 Wegener F. Uber generalisierte, septische Gefasserkrankungen. *Verh Dtsch Ges Pathol* 1936;**29**:202-9.

9 Wegener F. Uber eine eigenartige rhinogene granulomatose mit besonderer beteiligung des arteriensystems und der nieren. *Beitr Path Anat* 1939;**102**:36.

10 Churg J, Strauss L. Allergic granulomatosis, allergic angiitis, and periarteritis nodosa. *The American journal of pathology* 1951;**27**:277.

11 Jennette JC, Falk RJ, Bacon PA, et al. 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. *Arthritis Rheum* 2013;**65**:1-11.

12 Falk RJ, Gross WL, Guillevin L, et al. Granulomatosis with polyangiitis (Wegener's): an alternative name for Wegener's granulomatosis. *J Am Soc Nephrol* 2011;**22**:587-8.

13 Jennette JC, Falk RJ, Andrassy K, et al. Nomenclature of Systemic Vasculitides. *Arthritis & Rheumatism* 1994;**37**:187-92.

14 Mahr A, Specks U, Jayne D. Subclassifying ANCA-associated vasculitis: a unifying view of disease spectrum. *Rheumatology* 2019;**58**:1707-9.

15 Aggarwal R, Ringold S, Khanna D, et al. Distinctions between diagnostic and classification criteria? *Arthritis Care Res (Hoboken)* 2015;**67**:891-7.
16 Zeek PM. Periarteritis Nodosa: A Critical Review. *American Journal of Clinical Pathology* 1952;**22**:777-90.
17 Alarcón-Segovia D. 1 - Classification of the Necrotizing Vasculitides in Man. *Clinics in Rheumatic Diseases* 1980;**6**:223-31.
18 Alarcon Segovia D, Brown AL, Jr. Classification and Etiologic Aspects of Necrotizing Angiitides: An Analytic Approach to a Confused Subject with a Critical Review of the Evidence for Hypersensitivity in Polyarteritis Nodosa. *Mayo Clin Proc* 1964;**39**:205-22.
19 Gilliam JN, Smiley JD. Cutaneous necrotizing vasculitis and related disorders. *Ann Allergy* 1976;**37**:328-39.
20 deShazo RD. The Spectrum of Systemic Vasculitis. *Postgraduate Medicine* 1975;**58**:78-82.
21 Fauci AS, Haynes BF, Katz P. The Spectrum of Vasculitis. *Annals of Internal Medicine* 1978;**89**:660-76.
22 Fries JF, Hunder GG, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of vasculitis. Summary. *Arthritis Rheum* 1990;**33**:1135-6.
23 Leavitt RY, Fauci AS, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of Wegener's granulomatosis. *Arthritis Rheum* 1990;**33**:1101-7.
24 Masi AT, Hunder GG, Lie JT, et al. The American College of Rheumatology 1990 criteria for the classification of Churg-Strauss syndrome (allergic granulomatosis and angiitis). *Arthritis Rheum* 1990;**33**:1094-100.
25 Watts R, Lane S, Hanslik T, et al. Development and validation of a consensus methodology for the classification of the ANCA-associated vasculitides and polyarteritis nodosa for epidemiological studies. *Ann Rheum Dis* 2007;**66**:222-7.
26 Liu LJ, Chen M, Yu F, et al. Evaluation of a new algorithm in classification of systemic vasculitis. *Rheumatology (Oxford)* 2008;**47**:708-12.
27 Rathmann J, Segelmark M, Englund M, et al. Stable incidence but increase in prevalence of ANCA-associated vasculitis in southern Sweden: a 23-year study. *RMD Open* 2023;**9**.
28 Berti A, Cornec D, Crowson CS, et al. The Epidemiology of Antineutrophil Cytoplasmic Autoantibody–Associated Vasculitis in Olmsted County, Minnesota. *Arthritis & Rheumatology* 2017;**69**:2338-50.
29 Abdulkader R, Lane SE, Scott DG, et al. Classification of vasculitis: EMA classification using CHCC 2012 definitions. *Ann Rheum Dis* 2013;**72**:1888.
30 Grayson PC, Ponte C, Suppiah R, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology Classification Criteria for Eosinophilic Granulomatosis with Polyangiitis. *Ann Rheum Dis* 2022;**81**:309-14.

31 Robson JC, Grayson PC, Ponte C, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for granulomatosis with polyangiitis. *Ann Rheum Dis* 2022;**81**:315-20.
32 Suppiah R, Robson JC, Grayson PC, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for microscopic polyangiitis. *Ann Rheum Dis* 2022;**81**:321-6.
33 Rathmann J, Segelmark M, Mohammad AJ. Evaluation of the ACR/EULAR 2022 criteria for classification of ANCA-associated vasculitis in a population-based cohort from Sweden. *Rheumatology* 2023.
34 Cornec D, Cornec-Le Gall E, Fervenza FC, et al. ANCA-associated vasculitis - clinical utility of using ANCA specificity to classify patients. *Nat Rev Rheumatol* 2016;**12**:570-9.
35 Lyons PA, Rayner TF, Trivedi S, et al. Genetically distinct subsets within ANCA-associated vasculitis. *N Engl J Med* 2012;**367**:214-23.
36 Lionaki S, Blyth ER, Hogan SL, et al. Classification of antineutrophil cytoplasmic autoantibody vasculitides: the role of antineutrophil cytoplasmic autoantibody specificity for myeloperoxidase or proteinase 3 in disease recognition and prognosis. *Arthritis Rheum* 2012;**64**:3452-62.
37 Mohammad AJ, Segelmark M. A Population-based Study Showing Better Renal Prognosis for Proteinase 3 Antineutrophil Cytoplasmic Antibody (ANCA)–associated Nephritis Versus Myeloperoxidase ANCA–associated Nephritis. *The Journal of Rheumatology* 2014;**41**:1366.
38 Unizony S, Villarreal M, Miloslavsky EM, et al. Clinical outcomes of treatment of anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis based on ANCA type. *Ann Rheum Dis* 2016;**75**:1166-9.
39 Lyons PA, Peters JE, Alberici F, et al. Genome-wide association study of eosinophilic granulomatosis with polyangiitis reveals genomic loci stratified by ANCA status. *Nature Communications* 2019;**10**:5120.
40 Iudici M, Puéchal X, Brigante A, et al. Randomized clinical trials in ANCA-associated vasculitis: a systematic analysis of the WHO - International Clinical Trials Registry Platform. *Orphanet J Rare Dis* 2020;**15**:130.
41 Mahr A, Katsahian S, Varet H, et al. Revisiting the classification of clinical phenotypes of anti-neutrophil cytoplasmic antibody-associated vasculitis: a cluster analysis. *Ann Rheum Dis* 2013;**72**:1003-10.
42 van der Woude FJ, Rasmussen N, Lobatto S, et al. Autoantibodies against neutrophils and monocytes: tool for diagnosis and marker of disease activity in Wegener's granulomatosis. *Lancet* 1985;**1**:425-9.
43 Falk RJ, Jennette JC. Anti-neutrophil cytoplasmic autoantibodies with specificity for myeloperoxidase in patients with systemic vasculitis and idiopathic necrotizing and crescentic glomerulonephritis. *N Engl J Med* 1988;**318**:1651-7.
44 Niles JL, McCluskey RT, Ahmad MF, et al. Wegener's granulomatosis autoantigen is a novel neutrophil serine proteinase. *Blood* 1989;**74**:1888-93.

45 Klebanoff SJ. Myeloperoxidase: friend and foe. *Journal of Leukocyte Biology* 2005;**77**:598-625.
46 Martin KR, Witko-Sarsat V. Proteinase 3: the odd one out that became an autoantigen. *Journal of Leukocyte Biology* 2017;**102**:689-98.
47 Olson SW, Arbogast CB, Baker TP, et al. Asymptomatic autoantibodies associate with future anti-glomerular basement membrane disease. *J Am Soc Nephrol* 2011;**22**:1946-52.
48 Bossuyt X, Cohen Tervaert J-W, Arimura Y, et al. Revised 2017 international consensus on testing of ANCAs in granulomatosis with polyangiitis and microscopic polyangiitis. *Nature Reviews Rheumatology* 2017;**13**:683-92.
49 Csernok E, Moosig F. Current and emerging techniques for ANCA detection in vasculitis. *Nat Rev Rheumatol* 2014;**10**:494-501.
50 Damoiseaux J, Csernok E, Rasmussen N, et al. Detection of antineutrophil cytoplasmic antibodies (ANCAs): a multicentre European Vasculitis Study Group (EUVAS) evaluation of the value of indirect immunofluorescence (IIF) versus antigen-specific immunoassays. *Ann Rheum Dis* 2017;**76**:647-53.
51 Savige J, Gillis D, Benson E, et al. International Consensus Statement on Testing and Reporting of Antineutrophil Cytoplasmic Antibodies (ANCA). *Am J Clin Pathol* 1999;**111**:507-13.
52 Rasmussen N, Damoiseaux J, Csernok E, et al. Individual values of antineutrophil cytoplasmic antibodies do not correspond between antigen-specific assays. *Clinical Chemistry and Laboratory Medicine (CCLM)* 2018;**56**:e39-e42.
53 Kitching AR, Anders HJ, Basu N, et al. ANCA-associated vasculitis. *Nat Rev Dis Primers* 2020;**6**:71.
54 Eisenberger U, Fakhouri F, Vanhille P, et al. ANCA-negative pauci-immune renal vasculitis: histology and outcome. *Nephrol Dial Transplant* 2005;**20**:1392-9.
55 Roth AJ, Ooi JD, Hess JJ, et al. Epitope specificity determines pathogenicity and detectability in ANCA-associated vasculitis. *J Clin Invest* 2013;**123**:1773-83.
56 Bautz DJ, Preston GA, Lionaki S, et al. Antibodies with dual reactivity to plasminogen and complementary PR3 in PR3-ANCA vasculitis. *J Am Soc Nephrol* 2008;**19**:2421-9.
57 Pendergraft WF, 3rd, Preston GA, Shah RR, et al. Autoimmunity is triggered by cPR-3(105-201), a protein complementary to human autoantigen proteinase-3. *Nat Med* 2004;**10**:72-9.
58 Suzuki K, Nagao T, Itabashi M, et al. A novel autoantibody against moesin in the serum of patients with MPO-ANCA-associated vasculitis. *Nephrology dialysis transplantation* 2014;**29**:1168-77.
59 Peschel A, Basu N, Benharkou A, et al. Autoantibodies to hLAMP-2 in ANCA-Negative Pauci-Immune Focal Necrotizing GN. *Journal of the American Society of Nephrology* 2014;**25**:455-63.
60 Simon A, Subra JF, Guilpain P, et al. Detection of Anti-Pentraxin-3 Autoantibodies in ANCA-Associated Vasculitis. *PLoS One* 2016;**11**:e0147091.

61 Mohammad AJ, Jacobsson LTH, Mahr AD, et al. Prevalence of Wegener's granulomatosis, microscopic polyangiitis, polyarteritis nodosa and Churg–Strauss syndrome within a defined population in southern Sweden. *Rheumatology* 2007;**46**:1329-37.
62 Cui Z, Zhao MH, Segelmark M, et al. Natural autoantibodies to myeloperoxidase, proteinase 3, and the glomerular basement membrane are present in normal individuals. *Kidney Int* 2010;**78**:590-7.
63 Calabresi P, Edwards EA, Schilling RF. Fluorescent antiglobulin studies in leukopenic and related disorders. *J Clin Invest* 1959;**38**:2091-100.
64 Wiesner O, Russell KA, Lee AS, et al. Antineutrophil cytoplasmic antibodies reacting with human neutrophil elastase as a diagnostic marker for cocaine-induced midline destructive lesions but not autoimmune vasculitis. *Arthritis Rheum* 2004;**50**:2954-65.
65 Zhao MH, Jones SJ, Lockwood CM. Bactericidal/permeability-increasing protein (BPI) is an important antigen for anti-neutrophil cytoplasmic autoantibodies (ANCA) in vasculitis. *Clin Exp Immunol* 1995;**99**:49-56.
66 McGrath MM, Isakova T, Rennke HG, et al. Contaminated cocaine and antineutrophil cytoplasmic antibody-associated disease. *Clin J Am Soc Nephrol* 2011;**6**:2799-805.
67 Xiao H, Heeringa P, Hu P, et al. Antineutrophil cytoplasmic autoantibodies specific for myeloperoxidase cause glomerulonephritis and vasculitis in mice. *J Clin Invest* 2002;**110**:955-63.
68 Little MA, Al-Ani B, Ren S, et al. Anti-proteinase 3 anti-neutrophil cytoplasm autoantibodies recapitulate systemic vasculitis in mice with a humanized immune system. *PLoS One* 2012;**7**:e28626.
69 Tan DS, Gan PY, O'Sullivan KM, et al. Thymic deletion and regulatory T cells prevent antimyeloperoxidase GN. *J Am Soc Nephrol* 2013;**24**:573-85.
70 Abdulahad WH, Stegeman CA, van der Geld YM, et al. Functional defect of circulating regulatory CD4+ T cells in patients with Wegener's granulomatosis in remission. *Arthritis & Rheumatism* 2007;**56**:2080-91.
71 Free ME, Bunch DOD, McGregor JA, et al. Patients With Antineutrophil Cytoplasmic Antibody–Associated Vasculitis Have Defective Treg Cell Function Exacerbated by the Presence of a Suppression-Resistant Effector Cell Population. *Arthritis & Rheumatism* 2013;**65**:1922-33.
72 Wilde B, Thewissen M, Damoiseaux J, et al. Regulatory B cells in ANCA-associated vasculitis. *Annals of the rheumatic diseases* 2013:annrheumdis-2012-202986.
73 Kuligowski MP, Kwan RYQ, Lo C, et al. Antimyeloperoxidase antibodies rapidly induce α4-integrin–dependent glomerular neutrophil adhesion. *Blood* 2009;**113**:6485-94.
74 Johnson PA, Alexander HD, McMillan SA, et al. Up-regulation of the granulocyte adhesion molecule Mac-1 by autoantibodies in autoimmune vasculitis. *Clinical and Experimental Immunology* 1997;**107**:513-9.

75 Tse WY, Nash GB, Hewins P, et al. ANCA-induced neutrophil F-actin polymerization: Implications for microvascular inflammation. *Kidney International* 2005;**67**:130-9.
76 Brinkmann V, Reichard U, Goosmann C, et al. Neutrophil extracellular traps kill bacteria. *Science* 2004;**303**:1532-5.
77 Kessenbrock K, Krumbholz M, Schönermarck U, et al. Netting neutrophils in autoimmune small-vessel vasculitis. *Nature Medicine* 2009;**15**:623-5.
78 Falk RJ, Terrell RS, Charles LA, et al. Anti-neutrophil cytoplasmic autoantibodies induce neutrophils to degranulate and produce oxygen radicals in vitro. *Proceedings of the National Academy of Sciences* 1990;**87**:4115-9.
79 Nakazawa D, Masuda S, Tomaru U, et al. Pathogenesis and therapeutic interventions for ANCA-associated vasculitis. *Nature Reviews Rheumatology* 2019;**15**:91-101.
80 Hoffman GS, Calabrese LH. Vasculitis: determinants of disease patterns. *Nature Reviews Rheumatology* 2014;**10**:454-62.
81 Müller A, Krause B, Kerstein-Stähle A, et al. Granulomatous Inflammation in ANCA-Associated Vasculitis. *Int J Mol Sci* 2021;**22**.
82 Gan P-Y, Chan A, Ooi JD, et al. Biologicals targeting T helper cell subset differentiating cytokines are effective in the treatment of murine anti-myeloperoxidase glomerulonephritis. *Kidney International* 2019;**96**:1121-33.
83 Chang J, Eggenhuizen P, O'Sullivan KM, et al. CD8+ T Cells Effect Glomerular Injury in Experimental Anti-Myeloperoxidase GN. *J Am Soc Nephrol* 2017;**28**:47-55.
84 McKinney EF, Lyons PA, Carr EJ, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nature Medicine* 2010;**16**:586-91.
85 McKinney EF, Lee JC, Jayne DRW, et al. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature* 2015;**523**:612-6.
86 Stegeman CA, Tervaert JWC, Sluiter WJ, et al. Association of Chronic Nasal Carriage of Staphylococcus aureus and Higher Relapse Rates in Wegener Granulomatosis. *Annals of Internal Medicine* 1994;**120**:12-7.
87 Alba MA, Flores-Suárez LF, Henderson AG, et al. Interstital lung disease in ANCA vasculitis. *Autoimmunity Reviews* 2017;**16**:722-9.
88 White JPE, Dubey S. Eosinophilic granulomatosis with polyangiitis: A review. *Autoimmunity Reviews* 2023;**22**:103219.
89 Rubenstein E, Maldini C, Vaglio A, et al. Cluster Analysis To Explore Clinical Subphenotypes Of Eosinophilic Granulomatosis With Polyangiitis (Churg-Strauss). *J Rheumatol* 2023.
90 Wechsler ME, Akuthota P, Jayne D, et al. Mepolizumab or Placebo for Eosinophilic Granulomatosis with Polyangiitis. *N Engl J Med* 2017;**376**:1921-32.

91 Wechsler ME, Nair P, Terrier B, et al. Benralizumab versus Mepolizumab for Eosinophilic Granulomatosis with Polyangiitis. *N Engl J Med* 2024;**390**:911-21.
92 Goronzy JJ, Weyand CM. Immune aging and autoimmunity. *Cell Mol Life Sci* 2012;**69**:1615-23.
93 Bloom JL, Pickett-Nairn K, Silveira L, et al. The Association Between Age at Diagnosis and Disease Characteristics and Damage in Patients With ANCA-Associated Vasculitis. *Arthritis Rheumatol* 2023;**75**:2216-27.
94 Chinar R, Antien LM, Daphne van H, et al. Genetic variants in ANCA-associated vasculitis: a meta-analysis. *Annals of the Rheumatic Diseases* 2016;**75**:1687.
95 Grumet FC, Coukell A, Bodmer JG, et al. Histocompatibility (HL-A) Antigens Associated with Systemic Lupus Erythematosus. *New England Journal of Medicine* 1971;**285**:193-6.
96 Matzaraki V, Kumar V, Wijmenga C, et al. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology* 2017;**18**:76.
97 Kawasaki A, Hasebe N, Hidaka M, et al. Protective Role of HLA-DRB1*13:02 against Microscopic Polyangiitis and MPO-ANCA-Positive Vasculitides in a Japanese Population: A Case-Control Study. *PLOS ONE* 2016;**11**:e0154393.
98 Mohammad A, Segelmark M. Primary systemic vasculitis with severe α1-antitrypsin deficiency revisited. *Scand J Rheumatol* 2014;**43**:242-5.
99 Segelmark M, Elzouki A-N, Wieslander J, et al. The PiZ gene of α1-antitrypsin as a determinant of outcome in PR3-ANCA-positive vasculitis. *Kidney International* 1995;**48**:844-50.
100 Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature* 2018;**558**:73-9.
101 Trivioli G, Marquez A, Martorana D, et al. Genetics of ANCA-associated vasculitis: role in pathogenesis, classification and management. *Nature Reviews Rheumatology* 2022;**18**:559-74.
102 Cárdenas-Roldán J, Rojas-Villarraga A, Anaya J-M. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. *BMC Medicine* 2013;**11**:73.
103 Gomes AM, Nery F, Ventura A, et al. Familial clusters of ANCA small-vessel vasculitis. *NDT Plus* 2008;**2**:34-5.
104 Knight A, Sandin S, Askling J. Risks and relative risks of Wegener's granulomatosis among close relatives of patients with the disease. *Arthritis & Rheumatism* 2008;**58**:302-7.
105 Knight A, Sandin S, Askling J. Increased Risk of Autoimmune Disease in Families with Wegener's Granulomatosis. *The Journal of Rheumatology* 2010;**37**:2553.
106 Jones BE, Yang J, Muthigi A, et al. Gene-Specific DNA Methylation Changes Predict Remission in Patients with ANCA-Associated Vasculitis. *J Am Soc Nephrol* 2017;**28**:1175-87.

107 Hewins P, Williams JM, Wakelam MJ, et al. Activation of Syk in neutrophils by antineutrophil cytoplasm antibodies occurs via Fcgamma receptors and CD18. *J Am Soc Nephrol* 2004;**15**:796-808.
108 Williams JM, Ben-Smith A, Hewins P, et al. Activation of the G(i) heterotrimeric G protein by ANCA IgG F(ab')2 fragments is necessary but not sufficient to stimulate the recruitment of those downstream mediators used by intact ANCA IgG. *J Am Soc Nephrol* 2003;**14**:661-9.
109 Reily C, Stewart TJ, Renfrow MB, et al. Glycosylation in health and disease. *Nature Reviews Nephrology* 2019;**15**:346-66.
110 Espy C, Morelle W, Kavian N, et al. Sialylation levels of anti–proteinase 3 antibodies are associated with the activity of granulomatosis with polyangiitis (Wegener's). *Arthritis & Rheumatism* 2011;**63**:2105-15.
111 Lardinois OM, Deterding LJ, Hess JJ, et al. Immunoglobulins G from patients with ANCA-associated vasculitis are atypically glycosylated in both the Fc and Fab regions and the relation to disease activity. *PLOS ONE* 2019;**14**:e0213215.
112 Xu P-C, Gou S-J, Yang X-W, et al. Influence of variable domain glycosylation on anti-neutrophil cytoplasmic autoantibodies and anti-glomerular basement membrane autoantibodies. *BMC Immunology* 2012;**13**:10.
113 Specks U. What you should know about PR3-ANCA. Conformational requirements of proteinase 3 (PR3) for enzymatic activity and recognition by PR3-ANCA. *Arthritis Res* 2000;**2**:263-7.
114 Kemna MJ, Schlumberger W, van Paassen P, et al. The avidity of PR3-ANCA in patients with granulomatosis with polyangiitis during follow-up. *Clin Exp Immunol* 2016;**185**:141-7.
115 Chacko BK, Scott DW, Chandler RT, et al. Endothelial surface N-glycans mediate monocyte adhesion and are targets for anti-inflammatory effects of peroxisome proliferator-activated receptor γ ligands. *J Biol Chem* 2011;**286**:38738-47.
116 Scott DW, Vallejo MO, Patel RP. Heterogenic endothelial responses to inflammation: role for differential N-glycosylation and vascular bed of origin. *J Am Heart Assoc* 2013;**2**:e000263.
117 Ercolini AM, Miller SD. The role of infections in autoimmune disease. *Clinical and Experimental Immunology* 2008;**155**:1-15.
118 Scott J, Hartnett J, Mockler D, et al. Environmental risk factors associated with ANCA associated vasculitis: A systematic mapping review. *Autoimmunity Reviews* 2020;**19**:102660.
119 Salmela A, Rasmussen N, Tervaert JWC, et al. Chronic nasal Staphylococcus aureus carriage identifies a subset of newly diagnosed granulomatosis with polyangiitis patients with high relapse rate. *Rheumatology* 2017;**56**:965-72.
120 Lamprecht P, Fischer N, Huang J, et al. Changes in the composition of the upper respiratory tract microbial community in granulomatosis with polyangiitis. *Journal of Autoimmunity* 2019;**97**:29-39.

121 Wagner J, Harrison EM, Martinez Del Pero M, et al. The composition and functional protein subsystems of the human nasal microbiome in granulomatosis with polyangiitis: a pilot study. *Microbiome* 2019;**7**:137.
122 Rhee RL, Lu J, Bittinger K, et al. Dynamic Changes in the Nasal Microbiome Associated With Disease Activity in Patients With Granulomatosis With Polyangiitis. *Arthritis Rheumatol* 2021;**73**:1703-12.
123 Miyauchi E, Shimokawa C, Steimle A, et al. The impact of the gut microbiome on extra-intestinal autoimmune diseases. *Nature Reviews Immunology* 2023;**23**:9-23.
124 Sun L, Zhang B. The digestive system and autoimmunity. *BMC Immunology* 2023;**24**:36.
125 Najem CE. Characterizing the Gut and Plasma Metabolomes in Patients with Anca-Associated Vasculitis. *2018 ACR/ARHP Annual Meeting*: ACR 2018.
126 Zycinska K, Wardyn KA, Zielonka TM, et al. Co-trimoxazole and prevention of relapses of PR3-ANCA positive vasculitis with pulmonary involvement. *European Journal of Medical Research* 2009;**14**:265.
127 Stegeman CA, Cohen Tervaert JW, de Jong PE, et al. Trimethoprim–Sulfamethoxazole (Co-Trimoxazole) for the Prevention of Relapses of Wegener's Granulomatosis. *New England Journal of Medicine* 1996;**335**:16-20.
128 Rathmann J, Stamatis P, Jönsson G, et al. Infection is associated with increased risk of MPO- but not PR3-ANCA-associated vasculitis. *Rheumatology* 2022;**61**:4817-26.
129 Houen G, Trier NH. Epstein-Barr Virus and Systemic Autoimmune Diseases. *Frontiers in Immunology* 2021;**11**.
130 Izci Duran T, Turkmen E, Dilek M, et al. ANCA-associated vasculitis after COVID-19. *Rheumatol Int* 2021;**41**:1523-9.
131 Kitamoto K, Tanaka Y, Kuboyama T, et al. Newly diagnosed ANCA-associated vasculitis after COVID-19 infection: a case report. *Journal of Medical Case Reports* 2023;**17**:366.
132 McGonagle D, Bridgewood C, Ramanan AV, et al. COVID-19 vasculitis and novel vasculitis mimics. *The Lancet Rheumatology* 2021;**3**:e224-e33.
133 Shakoor MT, Birkenbach MP, Lynch M. ANCA-Associated Vasculitis Following Pfizer-BioNTech COVID-19 Vaccine. *American Journal of Kidney Diseases* 2021;**78**:611-3.
134 Stassen PM, Sanders J-SF, Kallenberg CGM, et al. Influenza vaccination does not result in an increase in relapses in patients with ANCA-associated vasculitis. *Nephrology Dialysis Transplantation* 2007;**23**:654-8.
135 Watanabe T. Vasculitis Following Influenza Vaccination: A Review of the Literature. *Current Rheumatology Reviews* 2017;**13**:188-96.
136 Knight A, Sandin S, Askling J. Occupational risk factors for Wegener's granulomatosis: a case–control study. *Annals of the Rheumatic Diseases* 2010;**69**:737.

137 Gómez-Puerta JA, Gedmintas L, Costenbader KH. The association between silica exposure and development of ANCA-associated vasculitis: systematic review and meta-analysis. *Autoimmunity reviews* 2013;**12**:1129-35.
138 Farquhar H, McGettigan B, Chapman P, et al. Incidence of anti-neutrophil cytoplasmic antibody-associated vasculitis before and after the February 2011 Christchurch Earthquake. *Internal medicine journal* 2017;**47**:57-61.
139 Takeuchi Y, Saito A, Ojima Y, et al. The influence of the Great East Japan earthquake on microscopic polyangiitis: A retrospective observational study. *PLoS One* 2017;**12**:e0177482.
140 Yashiro M, Muso E, Itoh-Ihara T, et al. Significantly high regional morbidity of MPO-ANCA-related angitis and/or nephritis with respiratory tract involvement after the 1995 great earthquake in Kobe (Japan). *Am J Kidney Dis* 2000;**35**:889-95.
141 Lane SE, Watts RA, Bentham G, et al. Are environmental factors important in primary systemic vasculitis?: a case–control study. *Arthritis & Rheumatism* 2003;**48**:814-23.
142 Koldingsnes W, Nossent H. Epidemiology of Wegener's granulomatosis in northern Norway. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 2000;**43**:2481-7.
143 Draibe J, Rodo X, Fulladosa X, et al. Seasonal variations in the onset of positive and negative renal ANCA-associated vasculitis in Spain. *Clinical kidney journal* 2018;**11**:468-73.
144 Watts R, Gonzalez-Gay M, Lane S, et al. Geoepidemiology of systemic vasculitis: comparison of the incidence in two regions of Europe. *Annals of the rheumatic diseases* 2001;**60**:170.
145 O'Donnell J, Stevanovic V, Frampton C, et al. Wegener's granulomatosis in New Zealand: evidence for a latitude-dependent incidence gradient. *Internal medicine journal* 2007;**37**:242-6.
146 Gatenby PA, Lucas RM, Engelsen O, et al. Antineutrophil cytoplasmic antibody–associated vasculitides: Could geographic patterns be explained by ambient ultraviolet radiation? *Arthritis Care & Research* 2009;**61**:1417-24.
147 Hagenau T, Vest R, Gissel TN, et al. Global vitamin D levels in relation to age, gender, skin pigmentation and latitude: an ecologic meta-regression analysis. *Osteoporos Int* 2009;**20**:133-40.
148 Kälsch A-I, Peters A, Buhl B, et al. Retinoid X receptor beta polymorphisms do not explain functional differences in vitamins D and A response in Antineutrophil cytoplasmic antibody associated vasculitis patients. *Autoimmunity* 2009;**42**:467-74.
149 Kemna M, Tervaert J, Broen K. Timmermans SAMEG, Van Paassen P, Damoiseaux JGMC: Seasonal influence on the risk of relapse at a rise of antineutrophil cytoplasmic antibodies in vasculitis patients with renal involvement. *J Rheumatol* 2017;**44**:473-81.

150 Scott J, Havyarimana E, Navarro-Gallinad A, et al. The association between ambient UVB dose and ANCA-associated vasculitis relapse and onset. *Arthritis Res Ther* 2022;**24**:147.
151 Wada N, Mukai M, Kohno M, et al. Prevalence of serum anti-myeloperoxidase antineutrophil cytoplasmic antibodies (MPO-ANCA) in patients with Graves' disease treated with propylthiouracil and thiamazole. *Endocr J* 2002;**49**:329-34.
152 Nakazawa D, Tomaru U, Suzuki A, et al. Abnormal conformation and impaired degradation of propylthiouracil-induced neutrophil extracellular traps: implications of disordered neutrophil extracellular traps in a rat model of myeloperoxidase antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum* 2012;**64**:3779-87.
153 Gao Y, Zhao MH. Review article: Drug-induced anti-neutrophil cytoplasmic antibody-associated vasculitis. *Nephrology (Carlton)* 2009;**14**:33-41.
154 Gill C, Sturman J, Ozbek L, et al. Cocaine-induced granulomatosis with polyangiitis—an under-recognized condition. *Rheumatology Advances in Practice* 2023;**7**.
155 Mohammad AJ. An update on the epidemiology of ANCA-associated vasculitis. *Rheumatology* 2020;**59**:iii42-iii50.
156 World Health Organization. Official Updates to ICD-10. 2005. https://www.who.int/publications/m/item/icd-10-updates-2004 (accessed: 2024-09-26)
157 Fujimoto S, Uezono S, Hisanaga S, et al. Incidence of ANCA-associated primary renal vasculitis in the Miyazaki Prefecture: the first population-based, retrospective, epidemiologic survey in Japan. *Clin J Am Soc Nephrol* 2006;**1**:1016-22.
158 Fairweather D, Rose NR. Women and autoimmune diseases. *Emerg Infect Dis* 2004;**10**:2005-11.
159 Mossberg M, Segelmark M, Kahn R, et al. Epidemiology of primary systemic vasculitis in children: a population-based study from southern Sweden. *Scandinavian Journal of Rheumatology* 2018;**47**:295-302.
160 Schlieben DJ, Korbet SM, Kimura RE, et al. Pulmonary-renal syndrome in a newborn with placental transmission of ANCAs. *Am J Kidney Dis* 2005;**45**:758-61.
161 European Parliament and Council of the European Union. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. 2000. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000R0141 (accessed: 2024-09-26)
162 Quartuccio L, Treppo E, Valent F, et al. Healthcare and economic burden of ANCA-associated vasculitis in Italy: an integrated analysis from clinical and administrative databases. *Intern Emerg Med* 2021;**16**:581-9.
163 Degli Esposti L, Dovizio M, Perrone V, et al. Profile, Healthcare Resource Consumption and Related Costs in ANCA-Associated Vasculitis Patients: A Real-World Analysis in Italy. *Advances in Therapy* 2023;**40**:5338-53.

164 Ahn SS, Lim H, Lee CH, et al. Secular Trends of Incidence, Prevalence, and Healthcare Economic Burden in ANCA-Associated Vasculitis: An Analysis of the 2002-2018 South Korea National Health Insurance Database. *Front Med (Lausanne)* 2022;**9**:902423.
165 Thorpe CT, Thorpe JM, Jiang T, et al. Healthcare utilization and expenditures for United States Medicare beneficiaries with systemic vasculitis. *Seminars in Arthritis and Rheumatism* 2018;**47**:507-19.
166 Liu GY, Ventura IB, Achtar-Zadeh N, et al. Prevalence and Clinical Significance of Antineutrophil Cytoplasmic Antibodies in North American Patients With Idiopathic Pulmonary Fibrosis. *Chest* 2019;**156**:715-23.
167 Kagiyama N, Takayanagi N, Kanauchi T, et al. Antineutrophil cytoplasmic antibody-positive conversion and microscopic polyangiitis development in patients with idiopathic pulmonary fibrosis. *BMJ Open Respir Res* 2015;**2**:e000058.
168 Ando M, Miyazaki E, Ishii T, et al. Incidence of myeloperoxidase anti-neutrophil cytoplasmic antibody positivity and microscopic polyangitis in the course of idiopathic pulmonary fibrosis. *Respir Med* 2013;**107**:608-15.
169 Kariv R, Sidi Y, Gur H. Systemic vasculitis presenting as a tumorlike lesion. Four case reports and an analysis of 79 reported cases. *Medicine* 2000;**79**:349-59.
170 Pearce FA, Hubbard RB, Grainge MJ, et al. Can granulomatosis with polyangiitis be diagnosed earlier in primary care? A case-control study. *Qjm* 2018;**111**:39-45.
171 Sreih AG, Cronin K, Shaw DG, et al. Diagnostic delays in vasculitis and factors associated with time to diagnosis. *Orphanet Journal of Rare Diseases* 2021;**16**:184.
172 Taimen K, Mustonen A, Pirilä L. The Delay and Costs of Diagnosing Systemic Vasculitis in a Tertiary-Level Clinic. *Rheumatol Ther* 2021;**8**:233-42.
173 Houben E, Groenland SL, van der Heijden JW, et al. Relation between duration of the prodromal phase and renal damage in ANCA-associated vasculitis. *BMC Nephrol* 2017;**18**:378.
174 Takala JH, Kautiainen H, Malmberg H, et al. Wegener's granulomatosis in Finland in 1981–2000: clinical presentation and diagnostic delay. *Scandinavian Journal of Rheumatology* 2008;**37**:435-8.
175 Houben E, Bax WA, van Dam B, et al. Diagnosing ANCA-associated vasculitis in ANCA positive patients: A retrospective analysis on the role of clinical symptoms and the ANCA titre. *Medicine (Baltimore)* 2016;**95**:e5096.
176 Luqmani RA, Bacon PA, Moots RJ, et al. Birmingham Vasculitis Activity Score (BVAS) in systemic necrotizing vasculitis. *Qjm* 1994;**87**:671-8.
177 Mukhtyar C, Lee R, Brown D, et al. Modification and validation of the Birmingham Vasculitis Activity Score (version 3). *Ann Rheum Dis* 2009;**68**:1827-32.

178 Stone JH, Hoffman GS, Merkel PA, et al. A disease-specific activity index for Wegener's granulomatosis: Modification of the Birmingham Vasculitis Activity Score. *Arthritis & Rheumatism* 2001;**44**:912-20.
179 Hellmich B, Sanchez-Alamo B, Schirmer JH, et al. EULAR recommendations for the management of ANCA-associated vasculitis: 2022 update. *Annals of the Rheumatic Diseases* 2024;**83**:30.
180 Emmi G, Bettiol A, Gelain E, et al. Evidence-Based Guideline for the diagnosis and management of eosinophilic granulomatosis with polyangiitis. *Nature Reviews Rheumatology* 2023;**19**:378-93.
181 Benedek TG. History of the development of corticosteroid therapy. *Clin Exp Rheumatol* 2011;**29**:S-5-12.
182 Fahey JL, Leonard E, Churg J, et al. Wegener's Granulomatosis. *The American Journal of Medicine* 1954;**17**:168-79.
183 Fauci AS, Katz P, Haynes BF, et al. Cyclophosphamide Therapy of Severe Systemic Necrotizing Vasculitis. *New England Journal of Medicine* 1979;**301**:235-8.
184 Fauci AS, Haynes BF, Katz P, et al. Wegener's Granulomatosis: Prospective Clinical and Therapeutic Experience With 85 Patients for 21 Years. *Annals of Internal Medicine* 1983;**98**:76-85.
185 Jayne D, Rasmussen N, Andrassy K, et al. A Randomized Trial of Maintenance Therapy for Vasculitis Associated with Antineutrophil Cytoplasmic Autoantibodies. *New England Journal of Medicine* 2003;**349**:36-44.
186 De Groot K, Rasmussen N, Bacon PA, et al. Randomized trial of cyclophosphamide versus methotrexate for induction of remission in early systemic antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum* 2005;**52**:2461-9.
187 Jayne DR, Gaskin G, Rasmussen N, et al. Randomized trial of plasma exchange or high-dosage methylprednisolone as adjunctive therapy for severe renal vasculitis. *J Am Soc Nephrol* 2007;**18**:2180-8.
188 de Groot K, Harper L, Jayne DR, et al. Pulse versus daily oral cyclophosphamide for induction of remission in antineutrophil cytoplasmic antibody-associated vasculitis: a randomized trial. *Ann Intern Med* 2009;**150**:670-80.
189 Hiemstra TF, Walsh M, Mahr A, et al. Mycophenolate mofetil vs azathioprine for remission maintenance in antineutrophil cytoplasmic antibody-associated vasculitis: a randomized controlled trial. *Jama* 2010;**304**:2381-8.
190 Stone JH, Merkel PA, Spiera R, et al. Rituximab versus Cyclophosphamide for ANCA-Associated Vasculitis. *New England Journal of Medicine* 2010;**363**:221-32.
191 Jones RB, Cohen Tervaert JW, Hauser T, et al. Rituximab versus Cyclophosphamide in ANCA-Associated Renal Vasculitis. *New England Journal of Medicine* 2010;**363**:211-20.
192 Guillevin L, Pagnoux C, Karras A, et al. Rituximab versus Azathioprine for Maintenance in ANCA-Associated Vasculitis. *New England Journal of Medicine* 2014;**371**:1771-80.

193 Karras A, Pagnoux C, Haubitz M, et al. Randomised controlled trial of prolonged treatment in the remission phase of ANCA-associated vasculitis. *Ann Rheum Dis* 2017;**76**:1662-8.
194 Charles P, Terrier B, Perrodeau É, et al. Comparison of individually tailored versus fixed-schedule rituximab regimen to maintain ANCA-associated vasculitis remission: results of a multicentre, randomised controlled, phase III trial (MAINRITSAN2). *Ann Rheum Dis* 2018;**77**:1143-9.
195 Jones RB, Hiemstra TF, Ballarin J, et al. Mycophenolate mofetil versus cyclophosphamide for remission induction in ANCA-associated vasculitis: a randomised, non-inferiority trial. *Ann Rheum Dis* 2019;**78**:399-405.
196 Charles P, Perrodeau É, Samson M, et al. Long-Term Rituximab Use to Maintain Remission of Antineutrophil Cytoplasmic Antibody-Associated Vasculitis: A Randomized Trial. *Ann Intern Med* 2020;**173**:179-87.
197 Walsh M, Merkel PA, Peh C-A, et al. Plasma Exchange and Glucocorticoids in Severe ANCA-Associated Vasculitis. *New England Journal of Medicine* 2020;**382**:622-31.
198 Jayne DRW, Merkel PA, Schall TJ, et al. Avacopan for the Treatment of ANCA-Associated Vasculitis. *New England Journal of Medicine* 2021;**384**:599-609.
199 Specks U, Merkel PA, Seo P, et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. *N Engl J Med* 2013;**369**:417-27.
200 Walsh M, Collister D, Zeng L, et al. The effects of plasma exchange in patients with ANCA-associated vasculitis: an updated systematic review and meta-analysis. *BMJ* 2022;**376**:e064604.
201 Sanders J-SF, de Joode AAE, DeSevaux RG, et al. Extended versus standard azathioprine maintenance therapy in newly diagnosed proteinase-3 anti-neutrophil cytoplasmic antibody-associated vasculitis patients who remain cytoplasmic anti-neutrophil cytoplasmic antibody-positive after induction of remission: a randomized clinical trial. *Nephrology Dialysis Transplantation* 2016;**31**:1453-9.
202 Pearce FA, McGrath C, Sandhu R, et al. Outcomes and compliance with standards of care in anti-neutrophil cytoplasmic antibody-associated vasculitis-insights from a large multiregion audit. *Rheumatol Adv Pract* 2018;**2**:rky025.
203 Anghel LA, Farcaş AM, Oprean RN. Medication adherence and persistence in patients with autoimmune rheumatic diseases: a narrative review. *Patient Prefer Adherence* 2018;**12**:1151-66.
204 Garbe N, Schäfer C, Pilz A, et al. The impact of a structured one–day seminar on disease–specific knowledge, lifestyle habits and disease impairment in ANCA–associated vasculitis. Results of a randomized, controlled study. *Scandinavian Journal of Rheumatology* 2023;**52**:69-76.
205 Tremain AC, Wallace RP, Lorentz KM, et al. Synthetically glycosylated antigens for the antigen-specific suppression of established immune responses. *Nature Biomedical Engineering* 2023;**7**:1142-55.

206 Tsai S, Shameli A, Yamanouchi J, et al. Reversal of autoimmunity by boosting memory-like autoregulatory T cells. *Immunity* 2010;**32**:568-80.
207 Mackensen A, Müller F, Mougiakakos D, et al. Anti-CD19 CAR T cell therapy for refractory systemic lupus erythematosus. *Nature Medicine* 2022;**28**:2124-32.
208 Dörte L, Maria Z, Mario B, et al. CD19-targeting CAR T cells protect from ANCA-induced acute kidney injury. *Annals of the Rheumatic Diseases* 2024:ard-2023-224875.
209 Mueller F, Taubmann J, Voelkl S, et al. CD19-Targeted CAR-T Cells in Refractory Systemic Autoimmune Diseases: A Monocentric Experience from the First Fifteen Patients. *Blood* 2023;**142**:220.
210 Verdun N, Marks P. Secondary Cancers after Chimeric Antigen Receptor T-Cell Therapy. *New England Journal of Medicine* 2024;**390**:584-6.
211 Chalkia A, Flossmann O, Jones R, et al. Avacopan for ANCA-associated vasculitis with hypoxic pulmonary haemorrhage. *Nephrology Dialysis Transplantation* 2024.
212 Walton EW. Giant-cell granuloma of the respiratory tract (Wegener's granulomatosis). *Br Med J* 1958;**2**:265-70.
213 Merkel PA, Aydin SZ, Boers M, et al. The OMERACT core set of outcome measures for use in clinical trials of ANCA-associated vasculitis. *J Rheumatol* 2011;**38**:1480-6.
214 Quinn KA, Monti S, Christensen R, et al. Developing a composite outcome tool to measure response to treatment in ANCA-associated vasculitis: A mixed methods study from OMERACT 2020. *Semin Arthritis Rheum* 2021;**51**:1134-8.
215 Sacristán JA, Dilla T, Díaz-Cerezo S, et al. Patient-physician discrepancy in the perception of immune-mediated inflammatory diseases: rheumatoid arthritis, psoriatic arthritis and psoriasis. A qualitative systematic review of the literature. *PLoS One* 2020;**15**:e0234705.
216 Herlyn K, Hellmich B, Seo P, et al. Patient-reported outcome assessment in vasculitis may provide important data and a unique perspective. *Arthritis Care & Research* 2010;**62**:1639-45.
217 Robson JC, Jayne D, Merkel PA, et al. Systemic vasculitis and patient-reported outcomes: how the assessment of patient preferences and perspectives could improve outcomes. *Patient Relat Outcome Meas* 2019;**10**:37-42.
218 Monti S, Delvino P, Klersy C, et al. Factors influencing patient-reported outcomes in ANCA-associated vasculitis: correlates of the Patient Global Assessment. *Seminars in Arthritis and Rheumatism* 2022;**56**:152048.
219 van Eeden C, Osman MS, Cohen Tervaert JW. Fatigue in ANCA-associated vasculitis (AAV) and systemic sclerosis (SSc): similarities with Myalgic encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). A critical review of the literature. *Expert Review of Clinical Immunology* 2022;**18**:1049-70.

220 Basu N, Jones GT, Fluck N, et al. Fatigue: a principal contributor to impaired quality of life in ANCA-associated vasculitis. *Rheumatology (Oxford)* 2010;**49**:1383-90.
221 Basu N, McClean A, Harper L, et al. Explaining fatigue in ANCA-associated vasculitis. *Rheumatology (Oxford)* 2013;**52**:1680-5.
222 Hajj-Ali RA, Wilke WS, Calabrese LH, et al. Pilot study to assess the frequency of fibromyalgia, depression, and sleep disorders in patients with granulomatosis with polyangiitis (Wegener's). *Arthritis Care Res (Hoboken)* 2011;**63**:827-33.
223 Harper L, Hewitt CA, Litchfield I, et al. Management of fatigue with physical activity and behavioural change support in vasculitis: a feasibility study. *Rheumatology (Oxford)* 2021;**60**:4130-40.
224 Strand V, Jayne DRW, Horomanski A, et al. The impact of treatment with avacopan on health-related quality of life in antineutrophil cytoplasmic antibody-associated vasculitis: a post-hoc analysis of data from the ADVOCATE trial. *The Lancet Rheumatology* 2023;**5**:e451-e60.
225 Basu N, McClean A, Harper L, et al. Markers for work disability in anti-neutrophil cytoplasmic antibody-associated vasculitis. *Rheumatology (Oxford)* 2014;**53**:953-6.
226 Heron V, Gingold M, Kitching AR, et al. The impact of antineutrophil cytoplasmic antibody-associated vasculitis on employment and work disability in an Australian population. *Int J Rheum Dis* 2021;**24**:904-11.
227 Wallace ZS, Fu X, Harkness T, et al. All-cause and cause-specific mortality in ANCA-associated vasculitis: overall and according to ANCA type. *Rheumatology* 2019;**59**:2308-15.
228 Houben E, Penne EL, Voskuyl AE, et al. Cardiovascular events in anti-neutrophil cytoplasmic antibody-associated vasculitis: a meta-analysis of observational studies. *Rheumatology (Oxford)* 2018;**57**:555-62.
229 Sánchez Álamo B, Moi L, Bajema I, et al. Long-term outcomes and prognostic factors for survival of patients with ANCA-associated vasculitis. *Nephrology Dialysis Transplantation* 2023;**38**:1655-65.
230 Tabakovic D, Smith R, Jayne D, et al. High risk of stroke in ANCA-associated vasculitis—a population-based study. *Rheumatology* 2022;**62**:2806-12.
231 Borgas Y, Gisslander K, Mohammad MA, et al. Myocardial infarction in ANCA-associated vasculitis- a population-based cohort-study. *Zenodo* 2022: 252.
232 Leeuw Kd, Sanders JS, Stegeman C, et al. Accelerated atherosclerosis in patients with Wegener's granulomatosis. *Annals of the Rheumatic Diseases* 2005;**64**:753.
233 Filer AD, Gardner-Medwin JM, Thambyrajah J, et al. Diffuse endothelial dysfunction is common to ANCA associated systemic vasculitis and polyarteritis nodosa. *Annals of the Rheumatic Diseases* 2003;**62**:162.

234 Liapi M, Jayne D, Merkel PA, et al. Venous thromboembolism in ANCA-associated vasculitis: a population-based cohort study. *Rheumatology* 2021;**60**:4616-23.
235 Faurschou M, Obel N, Baslund B. High Risk of Pulmonary Embolism and Deep Venous Thrombosis but Not of Stroke in Granulomatosis With Polyangiitis (Wegener's). *Arthritis Care & Research* 2014;**66**:1910-4.
236 Merkel PA, Lo GH, Holbrook JT, et al. Brief communication: high incidence of venous thrombotic events among patients with Wegener granulomatosis: the Wegener's Clinical Occurrence of Thrombosis (WeCLOT) Study. *Ann Intern Med* 2005;**142**:620-6.
237 Kronbichler A, Leierer J, Leierer G, et al. Clinical associations with venous thromboembolism in anti-neutrophil cytoplasm antibody-associated vasculitides. *Rheumatology (Oxford)* 2017;**56**:704-8.
238 Huang Y-M, Wang H, Wang C, et al. Promotion of Hypercoagulability in Antineutrophil Cytoplasmic Antibody–Associated Vasculitis by C5a-Induced Tissue Factor–Expressing Microparticles and Neutrophil Extracellular Traps. *Arthritis & Rheumatology* 2015;**67**:2780-90.
239 Salmela A, Ekstrand A, Joutsi-Korhonen L, et al. Activation of endothelium, coagulation and fibrinolysis is enhanced and associates with renal anti-neutrophil cytoplasmic antibody-associated vasculitis†. *Nephrology Dialysis Transplantation* 2014;**30**:i53-i9.
240 Hilhorst M, Winckers K, Wilde B, et al. Patients with Antineutrophil Cytoplasmic Antibodies Associated Vasculitis in Remission Are Hypercoagulable. *The Journal of Rheumatology* 2013;**40**:2042.
241 Rathmann J, Jayne D, Segelmark M, et al. Incidence and predictors of severe infections in ANCA-associated vasculitis: a population-based cohort study. *Rheumatology* 2020;**60**:2745-54.
242 Vassilopoulos A, Vassilopoulos S, Kalligeros M, et al. Incidence of serious infections in patients with ANCA-associated vasculitis receiving immunosuppressive therapy: A systematic review and meta-analysis. *Front Med (Lausanne)* 2023;**10**:1110548.
243 Odler B, Riedl R, Gauckler P, et al. Risk factors for serious infections in ANCA-associated vasculitis. *Ann Rheum Dis* 2023;**82**:681-7.
244 Lionaki S, Hogan SL, Jennette CE, et al. The clinical course of ANCA small-vessel vasculitis on chronic dialysis. *Kidney Int* 2009;**76**:644-51.
245 Charlier C, Henegar C, Launay O, et al. Risk factors for major infections in Wegener granulomatosis: analysis of 113 patients. *Ann Rheum Dis* 2009;**68**:658-63.
246 Alberici F, Smith RM, Jones RB, et al. Long-term follow-up of patients who received repeat-dose rituximab as maintenance therapy for ANCA-associated vasculitis. *Rheumatology (Oxford)* 2015;**54**:1153-60.
247 Tieu J, Smith RM, Gopaluni S, et al. Rituximab Associated Hypogammaglobulinemia in Autoimmune Disease. *Front Immunol* 2021;**12**:671503.

248 Kronbichler A, Kerschbaum J, Gopaluni S, et al. Trimethoprim–sulfamethoxazole prophylaxis prevents severe/life-threatening infections following rituximab in antineutrophil cytoplasm antibody-associated vasculitis. *Annals of the Rheumatic Diseases* 2018;**77**:1440.
249 Heijl C, Harper L, Flossmann O, et al. Incidence of malignancy in patients treated for antineutrophil cytoplasm antibody-associated vasculitis: follow-up data from European Vasculitis Study Group clinical trials. *Annals of the Rheumatic Diseases* 2011;**70**:1415.
250 Hoffman GS, Kerr GS, Leavitt RY, et al. Wegener granulomatosis: an analysis of 158 patients. *Ann Intern Med* 1992;**116**:488-98.
251 Shang W, Ning Y, Xu X, et al. Incidence of Cancer in ANCA-Associated Vasculitis: A Meta-Analysis of Observational Studies. *PLOS ONE* 2015;**10**:e0126016.
252 Mahr A, Heijl C, Le Guenno G, et al. ANCA-associated vasculitis and malignancy: current evidence for cause and consequence relationships. *Best Pract Res Clin Rheumatol* 2013;**27**:45-56.
253 van Daalen EE, Rizzo R, Kronbichler A, et al. Effect of rituximab on malignancy risk in patients with ANCA-associated vasculitis. *Annals of the Rheumatic Diseases* 2017;**76**:1064.
254 Heijl C, Westman K, Höglund P, et al. Malignancies in Patients with Antineutrophil Cytoplasmic Antibody-associated Vasculitis: A Population-based Cohort Study. *The Journal of Rheumatology* 2020;**47**:1229.
255 Faurschou M, Sorensen IJ, Mellemkjaer L, et al. Malignancies in Wegener's granulomatosis: incidence and relation to cyclophosphamide therapy in a cohort of 293 patients. *The Journal of Rheumatology* 2008;**35**:100.
256 Robson JC, Doll H, Suppiah R, et al. Damage in the anca-associated vasculitides: long-term data from the European Vasculitis Study group (EUVAS) therapeutic trials. *Annals of the Rheumatic Diseases* 2015;**74**:177.
257 Mohammad AJ, Bakoush O, Sturfelt G, et al. The extent and pattern of organ damage in small vessel vasculitis measured by the Vasculitis Damage Index (VDI). *Scand J Rheumatol* 2009;**38**:268-75.
258 Exley AR, Bacon PA, Luqmani RA, et al. Development and initial validation of the vasculitis damage index for the standardized clinical assessment of damage in the systemic vasculitides. *Arthritis & Rheumatism* 1997;**40**:371-80.
259 Hellmich B, Flossmann O, Gross WL, et al. EULAR recommendations for conducting clinical studies and/or clinical trials in systemic vasculitis: focus on anti-neutrophil cytoplasm antibody-associated vasculitis. *Ann Rheum Dis* 2007;**66**:605-17.
260 Hogan SL, Nachman PH, Poulton CJ, et al. Understanding Long-term Remission Off Therapy in Antineutrophil Cytoplasmic Antibody-Associated Vasculitis. *Kidney Int Rep* 2019;**4**:551-60.
261 Moran SM, Scott J, Clarkson MR, et al. The Clinical Application of Urine Soluble CD163 in ANCA-Associated Vasculitis. *J Am Soc Nephrol* 2021;**32**:2920-32.

262 O'Reilly VP, Wong L, Kennedy C, et al. Urinary Soluble CD163 in Active Renal Vasculitis. *J Am Soc Nephrol* 2016;**27**:2906-16.
263 Pepper RJ, Draibe JB, Caplin B, et al. Association of Serum Calprotectin (S100A8/A9) Level With Disease Relapse in Proteinase 3-Antineutrophil Cytoplasmic Antibody-Associated Vasculitis. *Arthritis Rheumatol* 2017;**69**:185-93.
264 Tervaert JW, Huitema MG, Hené RJ, et al. Prevention of relapses in Wegener's granulomatosis by treatment based on antineutrophil cytoplasmic antibody titre. *Lancet* 1990;**336**:709-11.
265 Mehta P, Balakrishnan A, Phatak S, et al. Diagnostic accuracy of antineutrophil cytoplasmic antibodies (ANCA) in predicting relapses of ANCA-associated vasculitis: systematic review and meta-analysis. *Rheumatology International* 2023;**43**:437-48.
266 Tomasson G, Grayson PC, Mahr AD, et al. Value of ANCA measurements during remission to predict a relapse of ANCA-associated vasculitis--a meta-analysis. *Rheumatology (Oxford)* 2012;**51**:100-9.
267 Fussner LA, Hummel AM, Schroeder DR, et al. Factors Determining the Clinical Utility of Serial Measurements of Antineutrophil Cytoplasmic Antibodies Targeting Proteinase 3. *Arthritis Rheumatol* 2016;**68**:1700-10.
268 Kemna MJ, Damoiseaux J, Austen J, et al. ANCA as a predictor of relapse: useful in patients with renal involvement but not in patients with nonrenal disease. *J Am Soc Nephrol* 2015;**26**:537-42.
269 Yamaguchi M, Ando M, Kato S, et al. Increase of Antimyeloperoxidase Antineutrophil Cytoplasmic Antibody (ANCA) in Patients with Renal ANCA-associated Vasculitis: Association with Risk to Relapse. *J Rheumatol* 2015;**42**:1853-60.
270 McClure ME, Wason J, Gopaluni S, et al. Evaluation of PR3-ANCA Status After Rituximab for ANCA-Associated Vasculitis. *J Clin Rheumatol* 2019;**25**:217-23.
271 von Borstel A, Land J, Abdulahad WH, et al. CD27+CD38hi B Cell Frequency During Remission Predicts Relapsing Disease in Granulomatosis With Polyangiitis Patients. *Frontiers in Immunology* 2019;**10**.
272 Venhoff N, Niessen L, Kreuzaler M, et al. Reconstitution of the peripheral B lymphocyte compartment in patients with ANCA-associated vasculitides treated with rituximab for relapsing or refractory disease. *Autoimmunity* 2014;**47**:401-8.
273 King C, Druce KL, Nightingale P, et al. Predicting relapse in anti-neutrophil cytoplasmic antibody-associated vasculitis: a Systematic review and meta-analysis. *Rheumatology Advances in Practice* 2021;**5**.
274 He P, Hu J-P, Tian X-J, et al. Prevalence and risk factors of relapse in patients with ANCA-associated vasculitis receiving cyclophosphamide induction: a systematic review and meta-analysis of large observational studies. *Rheumatology* 2020;**60**:1067-79.

275 Brix SR, Noriega M, Tennstedt P, et al. Development and validation of a renal risk score in ANCA-associated glomerulonephritis. *Kidney Int* 2018;**94**:1177-88.
276 Berden AE, Ferrario F, Hagen EC, et al. Histopathologic classification of ANCA-associated glomerulonephritis. *J Am Soc Nephrol* 2010;**21**:1628-36.
277 Berti A, Cornec-Le Gall E, Cornec D, et al. Incidence, prevalence, mortality and chronic renal damage of anti-neutrophil cytoplasmic antibody-associated glomerulonephritis in a 20-year population-based cohort. *Nephrology Dialysis Transplantation* 2018;**34**:1508-17.
278 Sethi S, D'Agati VD, Nast CC, et al. A proposal for standardized grading of chronic changes in native kidney biopsy specimens. *Kidney Int* 2017;**91**:787-9.
279 KDIGO 2024 Clinical Practice Guideline for the Management of Antineutrophil Cytoplasmic Antibody (ANCA)-Associated Vasculitis. *Kidney Int* 2024;**105**:S71-s116.
280 Hruskova Z, Stel VS, Jayne D, et al. Characteristics and Outcomes of Granulomatosis With Polyangiitis (Wegener) and Microscopic Polyangiitis Requiring Renal Replacement Therapy: Results From the European Renal Association–European Dialysis and Transplant Association Registry. *American Journal of Kidney Diseases* 2015;**66**:613-20.
281 Little MA, Hassan B, Jacques S, et al. Renal transplantation in systemic vasculitis: when is it safe? *Nephrol Dial Transplant* 2009;**24**:3219-25.
282 Tang W, Bose B, McDonald SP, et al. The outcomes of patients with ESRD and ANCA-associated vasculitis in Australia and New Zealand. *Clinical journal of the American Society of Nephrology: CJASN* 2013;**8**:773.
283 Mohammad AJ, Jacobsson LTH, Westman KWA, et al. Incidence and survival rates in Wegener's granulomatosis, microscopic polyangiitis, Churg–Strauss syndrome and polyarteritis nodosa. *Rheumatology* 2009;**48**:1560-5.
284 Tan JA, Dehghan N, Chen W, et al. Mortality in ANCA-associated vasculitis: ameta-analysis of observational studies. *Annals of the Rheumatic Diseases* 2017;**76**:1566-74.
285 Heijl C, Mohammad AJ, Westman K, et al. Long-term patient survival in a Swedish population-based cohort of patients with ANCA-associated vasculitis. *RMD Open* 2017;**3**:e000435.
286 Guillevin L, Lhote F, Gayraud M, et al. Prognostic factors in polyarteritis nodosa and Churg-Strauss syndrome. A prospective study in 342 patients. *Medicine (Baltimore)* 1996;**75**:17-28.
287 Guillevin L, Pagnoux C, Seror R, et al. The Five-Factor Score revisited: assessment of prognoses of systemic necrotizing vasculitides based on the French Vasculitis Study Group (FVSG) cohort. *Medicine (Baltimore)* 2011;**90**:19-27.
288 Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016;**3**:160018.

289 McGlinn K, Rutherford MA, Gisslander K, et al. FAIRVASC: A semantic web approach to rare disease registry integration. *Computers in Biology and Medicine* 2022;**145**:105313.
290 Aerts H, Kalra D, Sáez C, et al. Quality of hospital electronic health record (EHR) data based on the international consortium for health outcomes measurement (ICHOM) in heart failure: pilot data quality assessment study. *JMIR medical informatics* 2021;**9**:e27842.
291 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of statistical software* 2010;**36**:1-48.
292 Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *BMC Bioinformatics* 2022;**23**:205.
293 Basagaña X, Barrera-Gómez J, Benet M, et al. A Framework for Multiple Imputation in Cluster Analysis. *American Journal of Epidemiology* 2013;**177**:718-25.
294 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011;**45**:1 - 67.
295 McParland D, Gormley I. Model Based Clustering for Mixed Data: clustMD. *Advances in Data Analysis and Classification* 2015;**10**.
296 O'Hagan A, Murphy TB, Gormley IC, et al. Clustering with the multivariate normal inverse Gaussian distribution. *Computational Statistics & Data Analysis* 2016;**93**:18-30.
297 Chang W, Cheng J, Allaire J, et al. shiny: Web Application Framework for R. 2023.
298 Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med* 2021;**385**:1737-49.
299 Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2011;**28**:112-8.
300 Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. *The Annals of Applied Statistics* 2008;**2**:841-60, 20.
301 Ishwaran H, Gerds TA, Kogalur UB, et al. Random survival forests for competing risks. *Biostatistics* 2014;**15**:757-73.
302 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565-74.
303 Samson M, Devilliers H, Thietart S, et al. Score to assess the probability of relapse in granulomatosis with polyangiitis and microscopic polyangiitis. *RMD Open* 2023;**9**.
304 European Commisson: Directorate-General for Health and Food Safety. Rare diseases. https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/rare-diseases_en (accessed: 2024-08-31).
305 GO FAIR. GO FAIR. https://www.go-fair.org/ (accessed: 2024-08-31).
306 Chanouzas D, McGregor JAG, Nightingale P, et al. Intravenous pulse methylprednisolone for induction of remission in severe ANCA associated

Vasculitis: a multi-center retrospective cohort study. *BMC Nephrology* 2019;**20**:58.
307 Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *npj Digital Medicine* 2020;**3**:119.
308 Watanabe H, Sada KE, Harigai M, et al. Exploratory classification of clinical phenotypes in Japanese patients with antineutrophil cytoplasmic antibody-associated vasculitis using cluster analysis. *Sci Rep* 2021;**11**:5223.
309 Pagnoux C, Carette S, Khalidi NA, et al. Comparability of patients with ANCA-associated vasculitis enrolled in clinical trials or in observational cohorts. *Clin Exp Rheumatol* 2015;**33**:S-77-83.
310 Benichou N, Charles P, Terrier B, et al. Proteinuria and hematuria after remission induction are associated with outcome in ANCA-associated vasculitis. *Kidney Int* 2023;**103**:1144-55.
311 Frumholtz L, Laurent-Roussel S, Aumaître O, et al. Clinical and pathological significance of cutaneous manifestations in ANCA-associated vasculitides. *Autoimmun Rev* 2017;**16**:1138-46.
312 Scott J, White A, Walsh C, et al. Computable phenotype for real-world, data-driven retrospective identification of relapse in ANCA-associated vasculitis. *RMD Open* 2024;**10**:e003962.
313 Sachez-Alamo B, Moi L, Bajema I, et al. Long-term outcome of kidney function in patients with ANCA-associated vasculitis. *Nephrology Dialysis Transplantation* 2024.