Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

UNIVERSITY OF ABERDEEN

Hamburg University of Applied Sciences

Faculty Life Sciences

Exploring healthcare utilisation in patients with ANCA-associated vasculitis in Scotland, UK

A longitudinal multicentre retrospective linked-data study

Master thesis

Master of Science Health Sciences

Submitted by

Laura Bruno

Enrolment number [redacted]

Hamburg

23rd of October 2018

1. Advisor: Prof. Dr. York Zöllner (HAW Hamburg)

2. Advisor: Dr. Neil Basu (University of Glasgow, formerly Aberdeen)

The master thesis was supervised and created in cooperation with the Epidemiology Group, within the Institute of Applied Health Sciences at the University of Aberdeen

Acknowledgements

This master thesis would not have been possible without the decisive support of a couple of people to whom I want to express my deep-felt gratitude.

First of all, thank you so much, Shifa, for all the endorsement during my internship with the Epidemiology Group and during my master thesis afterwards. Your statistical expertise and generosity in sharing your knowledge whenever needed, were extremely enriching and important for my personal development and the success of both, the internship project and the thesis. This deserves, especially under the circumstances of your own situation as a PhD student in the last year, special appreciation and I am very thankful.

I further want to thank you, Neil, for your guidance throughout the process and your ability to make very complex things very easy and hands-on. It was a pleasure to work with you and I am very grateful for this experience and all the efforts you made for having me in Aberdeen.

A lot of appreciation also goes to Lizzie Nicholson from eDRIS Scotland, thank you so much for the kind cooperation and support during the Safe Haven release processes, especially when files were needed within a short period of time.

Special thanks further go to my friends Christin, Dani and Toni. Thank you for all the days of shared desks, working together and trying to ease each other's frustrations. Thank you for all the coffees and conversations.

Last, but not least, thank you, York, for the very straight forward process during the thesis supervision and the uncomplicated handling of the upcoming issues. You have always been ready to help out with suggestions for solutions and sources, which is much appreciated.

Table of contents

List of Figures

List of Tables

Abbreviations

A&E ..... Accidents and Emergencies
AAV ..... ANCA-associated vasculitis
ANCA ..... Anti-neutrophil cytoplasmic antibody, Anti-neutrophil cytoplasmic antibody
COPD ..... Chronic Obstructive Pulmonary Disease
CT ..... Computer tomography
eDRIS ..... electronic Data Research and Innovation Service
EGPA ..... Eosinophilic granulomatosis with polyangiitis
GP ..... General practitioner
GPA ..... Granulomatosis with polyangiitis
HCHS ..... Hospital and Community Health Service
HERU ..... Health Economics Research Unit
ICD ..... International Classification of Diseases
ICU ..... Intensive Care Unit
ISD ..... Information Services Division
IQR ..... Interquartile range
IRR ..... Incidence Rate Ratio
MPO ..... Myeloperoxidase
MPA ..... Microscopic polyangiitis
NRS ..... National Records of Scotland
NHS ..... National Health Service
NOS ..... Newcastle-Ottawa-Scale
PR3 ..... Proteinase 3
RCTs ..... Randomised Controlled Trial
REC ..... Research Ethics Council
SIMD ..... Scottish Index of Multiple Deprivation
SMR00 ..... Scottish Morbidity Record on Outpatient Appointments and Attendances
SMR01 ..... Scottish Morbidity Record on General Acute Inpatient and Day Cases
UK ..... United Kingdom
US ..... United States (of America)
VIF ..... Variance Inflation Factor

Abstract

Objectives: ANCA-associated vasculitis (AAV) are a set of rare, chronic conditions, which have an increased health burden compared with the general population. The objective was to quantify this burden for Scottish patients with AAV by exploring healthcare utilisation in a well-defined cohort.

Methods: NHS Scotland provided longitudinal, historic multicentre data on routinely collected health records from seven centres across Scotland, UK. Included were 543 patients with AAV and up to five matches per patient from the general population (n=2671). Included healthcare utilisation parameters were inpatient hospitalisation, outpatient encounters, accidents and emergencies and associated costs. Poisson and linear regression models were conducted to identify driving factors.

Results: Patients with AAV showed significantly higher healthcare utilisation compared to their matches across all included parameters. This increase was sustainable over up to ten years of follow-up. The mean costs per person-year were 4.17 ($p<0.0001$) times higher in the AAV cohort than that in the general population. The incremental costs per person-year were £6,323.84 (95%CI=£1,727.82-£10,919.87) per person-year. Results of the regression analyses were contradictive and therefore not overly conclusive.

Conclusion: This study demonstrates that Scottish patients with AAV show increased healthcare utilisation compared with the general population. The regression analyses were inconclusive; further research is highly warranted.

1 Introduction

1.1 ANCA-associated vasculitis

Anti-Neutrophil Cytoplasmic Antibody (ANCA)-associated vasculitides (AAV) are a set of rare chronic conditions, which are characterised by necrotising inflammation of the walls of small vessels (McKinney, Willcocks, Broecker, & Smith, 2014). Patients with AAV typically show insufficient or lack of immune complex deposition in vessel walls (J. C. Jennette et al., 2012).

Figure 1 below shows the assignment of the different diseases comprised by systemic vasculitis according to their vessel involvement. The diagram was developed by the 2012 International Chapel Hill Consensus Conference on the Nomenclature of Systemic Vasculitides (Jennette et al. 2012).

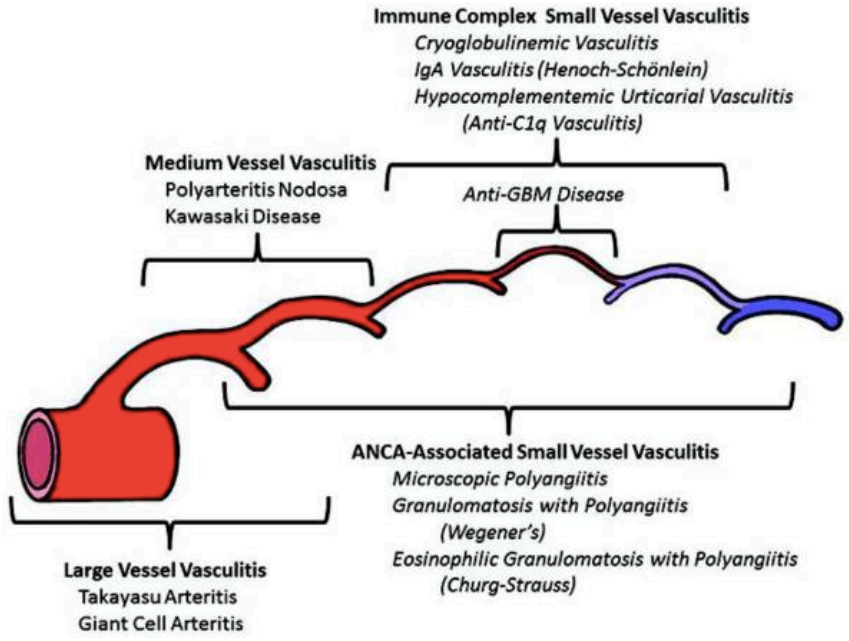


Figure 1 "Distribution of vessel involvement by large vessel vasculitis, medium vessel vasculitis and small vessel vasculitis [...]" (Jennette et al. 2012)

Accordingly, the umbrella term AAV comprises the following three diseases:

- Granulomatosis with polyangiitis (GPA), formerly known as Wegener's Granulomatosis
- Microscopic polyangiitis (MPA)
- Eosinophilic granulomatosis with polyangiitis (EGPA), formerly known as Churg-Strauss syndrome

Regarding the type of ANCA, patients either show ANCA, which are directed against proteinase 3 (PR3) or myeloperoxidase (MPO). The third group are AAV patients with ANCA negative status (Houben et al., 2016; McKinney et al., 2014; Morgan et al., 2017). The type of ANCA antibody often correlates with the type of AAV. Patients with GPA most often show PR3 antibodies, whereas patients with MPA show MPO (Jennette & Nachman, 2017; Kobayashi & Fujimoto, 2013). EGPA patients often are ANCA negative (Sokolowska et al., 2014).

The estimated incidence of AAV is 19.5 per million population. GPA is the most common subtype of AAV, with an estimated incidence of 11.3 per million as opposed to MPA with 5.9 per million population. The prevalence is estimated to be 255 per million population (Watts & Dharmapalaiah, 2012; Watts, Mooney, Skinner, Scott, & Macgregor, 2012).

The estimates for incidence and prevalence vary geographically. For example, GPA cases are more common in the UK than in Japan, whereas the reverse holds for MPA cases (Kobayashi & Fujimoto, 2013). Efforts to identify the cause for the geographic variations suggest that ethnicity may play a role with regard to the susceptibility for certain types of AAV (Bonatti, Reina, Neri, & Martorana, 2014). Other sources suggest a link between AAV type and vitamin-D levels (Cantorna & Mahon, 2004). The discussions are ongoing, no consensus has been reached.

The clinical presentation of AAV is broad and predominantly unspecific. The symptoms vary with the type of AAV. Nasal crusting, stuffiness, epistaxis, uveitis, symptoms of the upper respiratory tract and kidneys are typical symptoms of GPA patients. MPA patients often show severe renal involvement, with additional rash and neuropathy. For patient with EGPA, AAV is a multisystem disease, presented by asthma, nasal polyposis and peripheral blood eosinophilia (Watts & Dharmapalaiah, 2012; Yates & Watts, 2017).

Due to the ambiguity of symptoms, the diagnosis often occurs late, which leads to unfavourable outcomes regarding the quality of life and overall longevity of the patients. To date, there is no validated diagnostic system available (Houben et al., 2016; McKinney et al., 2014).

When left untreated, AAV can lead to death within a few months, which was found by Walton in 1958 already, when efficient therapy was not available yet (Walton, 1958). A similar picture was drawn almost thirty years later, when Fauci et al. showed that AAV lead to death within one year in 80% of cases (Fauci, Haynes, Katz, & Wolff, 1983). Since then, a lot has changed in the management of AAV.

The introduction of innovative treatment combinations resulted in an increased five-year survival rate of around 75% (Booth et al., 2003). The therapy is biphasic, encompassing the induction of remission, using immunosuppressants for quick control of disease activity in the first 3-6 months. This is followed by a maintenance phase of remission of at least 18 months. The medication is depending on the type of ANCA as well as the organ manifestation and disease severity, including especially cyclophosphamide, rituximab and the azathioprine group (Yates & Watts, 2017).

Notwithstanding the progress in five-year survival, the therapy is associated with substantial adverse effects and frequent relapses, demonstrating the need for less toxic and more effective medication (Schönermarck et al., 2014). Despite the treatment, patients with AAV are characterised by a 2.7-fold increased risk of death, compared to the general population, (Tan et al., 2017).

1.2 NHS Scotland

Following the National Health Service (Scotland) Act in 1947, NHS Scotland was established in 1948. It comprises 14 territorial health boards across the country and occupies approximately 140,000 staff (Scotland's Health on the Web, 2018).

Every inhabitant, regardless of nationality or duration of residence, is covered. The great majority of health services is tax funded and free for the patients. Accordingly, NHS Scotland is one of the largest public health care systems worldwide (Steel & Cylus, 2012). In

2017, this encompassed a population of 5.42 million, the highest ever (National Records of Scotland, 2018).

Just like Germany, Scotland is experiencing demographic changes, resulting in an increase of the population aged 65 or older from 13.1 to 16.8% between 1975 and 2010. Also, the age group of 80 or older has more than doubled in the same period (Steel & Cylus, 2012).

The healthcare expenditure per capita increased substantially from 2006 to 2010, as seen below. Compared to other countries in the UK, Scotlands healthcare spending is significantly higher. In 2011, 34% (11.68 billion) of the total budget of the Scottish government were provided for the healthcare system.

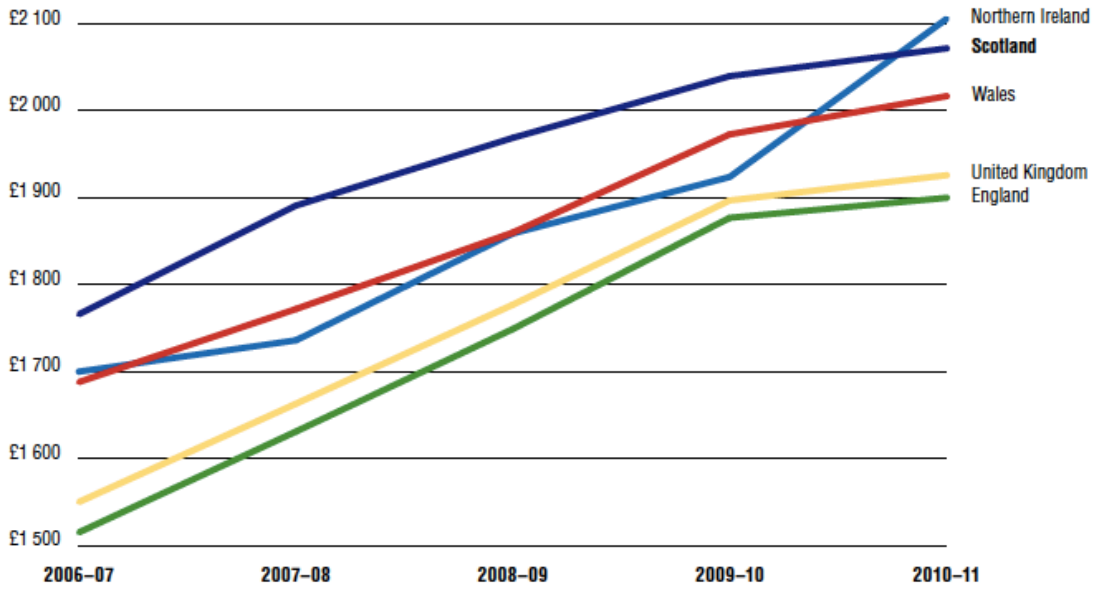


Figure 2 Health spending per capita in the United Kingdom, 2006/2007-2010/2011

Within NHS Scotland, each patient is assigned a ten digit CHI number. “The Community Health Index (CHI) is a population register, which is used in Scotland for health care purposes. The CHI number uniquely identifies a person on the index” (Information Services Division Scotland, 2018).

It was implemented in the 1970s and allows for the (non-identifiable) linkage of a multitude of administrative health records. These include inpatient, outpatient, prescribing and death

records among others (The University of Edinburgh, 2016). This infrastructure of administrative health records creates great opportunities for high quality research within NHS Scotland.

1.3 Research aim and objectives

AAV are among rare chronic conditions. Besides an increased health burden, chronic diseases show intensive healthcare use, which is partly due to the introduction of new technologies and drugs (Abegunde, Mathers, Adam, Ortegon, & Strong, 2007; Manuel, Schultz, & Kopec, 2002). However, studies analysing the economic burden of rare chronic diseases in for example cost of illness studies, are scarce (Lopez-Bastida et al., 2010). This holds true also for AAV.

Even though the health burden of patients with AAV has been widely researched (e.g. Basu et al., 2014; Dadonienė, Kumžaitė, Mačiulytė, & Miltinienė, 2017; Lee et al., 2014; Yates & Watts, 2017), a quantification of such was found to be missing, but highly warranted (Watts, Robson, & Pearce, 2017). Analyses comprising healthcare utilisation parameters, for example resource consumption regarding treatment, hospitalisation, corresponding costs for the health system, etc., are of high interest for policy makers, and physicians. Watts et al. call for "large-scale population-based studies […] to determine the direct and indirect costs to enable better resource allocation and to justify to health funders the introduction of novel biologic drugs" (Watts, Robson, & Pearce, 2017).

The present study serves as a contribution to fill this gap, with the overall research aim of:

Exploring healthcare utilisation in a well-defined cohort of Scottish patients with AAV

For this purpose, the definition of healthcare utilisation as well as information on the current state of research in this field were essential. This includes especially elevation methodologies, the *healthcare utilisation parameters*. Further, results of earlier projects need to be reviewed, in order to put the obtained results into context.

Supportive of the overall research aim, the following objectives were defined:

- To identify ways of measuring healthcare utilisation
- To find out about the availability of evidence on healthcare utilisation in adult patients with vasculitis with a specific focus on AAV
- Using the learnings from above, to assess healthcare utilisation in a well-defined AAV cohort in Scotland, UK
- To explore predicting factors for increased healthcare utilisation in AAV

This master thesis is a distinct work stream embedded in a programme of Scottish academic research led by Dr. Neil Basu. The gained insights shall inform local decision makers, with the aim of improving resource allocation to improve management of AAV patients.

2 Literature review

2.1 Methods

In September 2017, a systematic literature review was performed, searching Embase, Medline and Cochrane with the objective to identify evidence on ways of measuring healthcare utilisation. Another objective was to find existing literature on healthcare utilisation in adult patients with vasculitis. Search terms related to healthcare utilisation and systemic vasculitis and ANCA-associated vasculitis, respectively. The full list of included search terms can be found in Appendix I, showing the PICO framework as well as the complete summary of findings table.

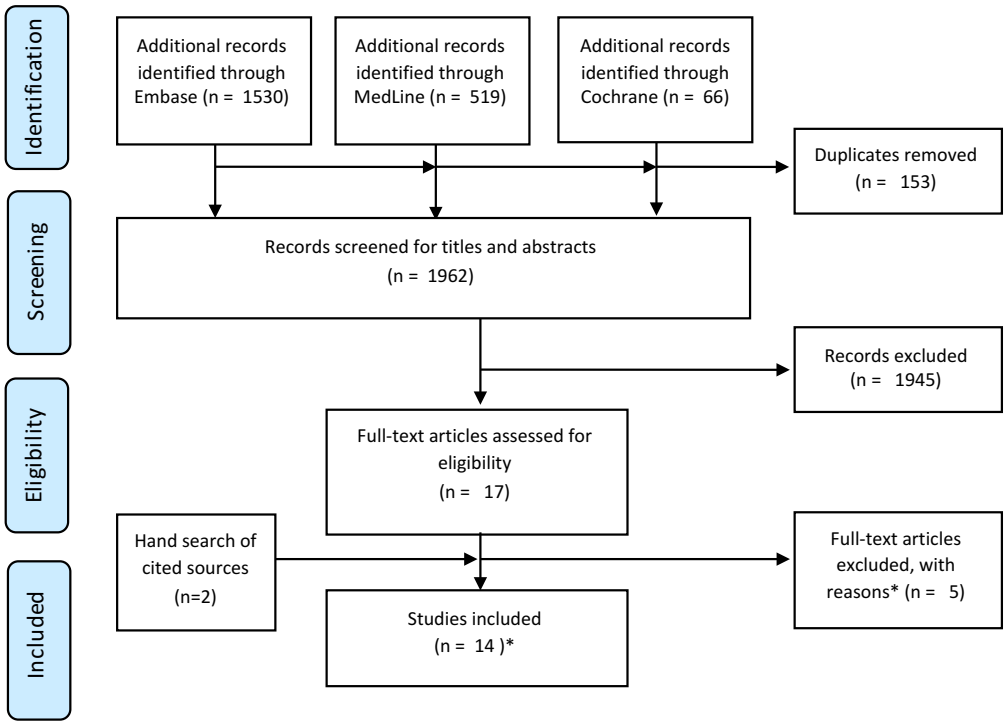
Ideally, sources were non-interventional, as the focus was mainly on the methods behind measuring healthcare utilisation. However, interventional studies were not excluded if they reported on useful aspects of the measurement of healthcare utilisation. All kinds of comparisons were included. Different variations of spelling and synonyms were taken into account. Reviews, case studies, and RCTs were excluded as well as studies with no abstract and in languages other than English and German. Results were extracted and duplicates removed using Refworks 2017.

The quality was assessed using "The Newcastle-Ottawa Scale (NOS) for assessing the quality in nonrandomised studies in meta-analyses". This tool is suitable for case control and cohort studies, working with a star system. A maximum of 9 stars can be assigned for the three main categories "selection of the study groups; the comparability of the groups; and the ascertainment […] of outcome of interest for […] cohort studies respectively" (Wells GA O'connell D., 2011).

2.2 Results

Of the initial 1962 results excluding duplicates, 1941 further studies were excluded in the screening phase, leaving 17 sources in the eligibility phase for the full-text assessment. Of these, five sources were excluded, the reasons for this procedure can be found in the detailed PRISMA flow diagram (Moher, Liberati, Tetzlaff, & Altman, 2014) below. By hand-searching the references of the included 12 studies, three new sources were identified. To sum up,

a total of 15 studies were included after full-text assessment. Ten of these were peer-reviewed journal articles, whereas five represent grey literature publications, including poster presentation abstracts, and conference abstracts.



*

1. **Cotch, M.F. (2000)**. The socioeconomic impact of vasculitis. Current Opinion in Rheumatology 2000, 12:20-23.
This publication is a review
2. **Putrik, P. et al. (2016)**. Socio-economically deprived patients have a higher likelihood for having any type of rheumatic and musculoskeletal diseases and have higher healthcare costs. Results from a population-based administrative database including 1.9 million persons (Basque Country, Spain). College of Rheumatology/ Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP http://dx.doi.org/10.1136/annrheumdis-2016-eular.1866
There are no results on the outcome of interest stated in the published abstract
3. **Raimundo, K. et al. (2015)**. Clinical and economic burden of patients with microscopic polyangiitis (MPA) in the United States. Annals of the Rheumatic Diseases. http://dx.doi.org/ 10.1136/annrheumdis-2015-eular.3205
The results of this published abstract are comprised in the later journal article
4. **Thorpe, C. T. et al. (2017).** Healthcare utilization and expenditures for United States Medicare beneficiaries with systemic vasculitis. Healthcare use and costs in vasculitis, Seminars in Arthritis and Rheumatism. http://dx.doi.org/10.1016/j.semarthrit.2017.08.005
The results of this accepted manuscript are comprised in the later journal article
5. **Wallace, Y. (2016)**. Nationwide trends in hospitalization and in-hospital mortality associated with ANCA-associated vasculitis (AAV). http://dx.doi.org/10.1136/annrheumdis-2016-eular.5833
The results of this accepted manuscript are comprised in the later journal article

Following: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Figure 3 PRISMA Flow Diagram following Moher, Liberati, Tetzlaff, & Altman, 2014

Of the total 15, eight sources (five journal articles and three grey literature publications) focussed on the population of main interest, patients with a form of AAV. Three further sources (two journal articles and one grey literature publication) used broader terms like *small-*, *medium-*, and *large vessel vasculitis*, as well as *systemic vasculitis* for the definition of their study populations, which also included AAV amongst others, but complicated the interpretation of results (Foocharoen et al., 2012; McCormick & Marra, 2015; Carolyn T. Thorpe et al., 2018). Other forms of vasculitis that were dealt with in the literature included *Giant Cell Arteritis*, *Takayasu Arteritis*, and *Behcet's syndrome*.

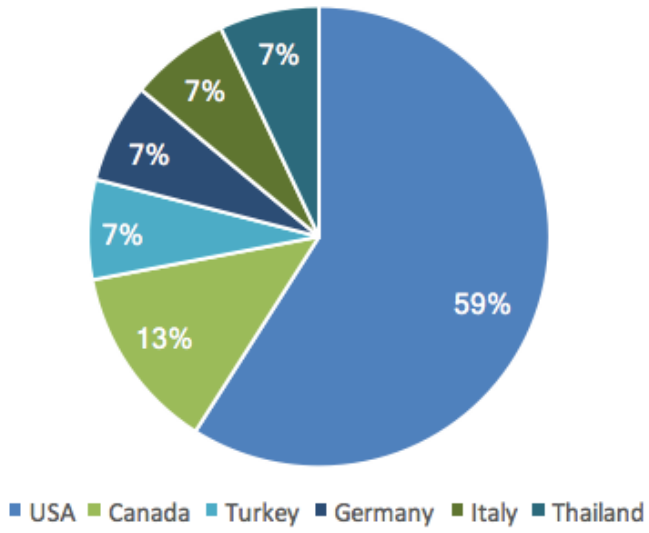


Figure 1 Study locations of sources included in the literature review

The majority of studies was conducted in the USA (n=9). This was also true for the studies focussing on the target population (n=7). Except for two studies from Italy and Germany, of which only the latter focussed on patients with AAV specifically (Reinhold-Keller, 2002), no studies researching healthcare utilisation parameters in patients with AAV were conducted in central Europe. The distribution of study locations can be seen in the pie chart on the left-hand side.

The majority, 66% (n=10,) of the papers were *population-based* cohort studies. The study durations ranged from 0.25 to 25 years. The power of the identified literature varied greatly, given the vastly different sample sizes, ranging from 44 to 176,498, where mentioned.

The study comprising the largest cohort compared the healthcare utilisation and expenditure for beneficiaries of the US healthcare provider Medicare with and without systemic vasculitis, based on claims data (Thorpe, 2017). However, their data comprised only one year of follow-up and did not distinguish between the different diseases, summarised by the term systemic vasculitis.

Wallace et al. (2016) further evaluated nationwide trends in hospitalisation and in-hospital mortality in GPA including all US-patients covered by Medicare, Medicaid, private insurance and no insurance based on publicly available inpatient data.With 18 years (1993-2011) of data included, this study also had the largest data base of the identified literature. However, their study did not state the total number of patients included in data analysis.

The majority of the studies based their results on national or large administrative databases with de-identified, linked data on medical records. These included claims and billing data as well as data on hospitalisations and mortality. Other sources were commercially used data from market research and self-administered or standardised questionnaires, which were more common in smaller, hospital-based studies.

Most of the data analyses were average-based and used either Stata, SAS, SPSS or R, to conduct mainly non-parametric tests, e.g. Mann-Whitney-U-, Wilcoxon-rank-sum-, McNemar-, Chi-square- or Kruskal-Wallis-test among others. No sources were found to identify driving factors instead of differences between groups.

Five of the included studies claimed to be population-based (Cotch et al., 1996; Foocharoen et al., 2012; McCormick et al., 2012; McCormick & Marra, 2015; Michet et al., 2015). In a response letter to the BMJ, Laupland highlights the importance of a clear definition of "population based". He says only studies may be entitled as such, if they include "all cases of disease occuring in an entire region" (Laupland, 2003). Szklo's description of the concept of population-based studies adds that the main purpose for conducting population-based studies is their great external validity. According to his definition, population-based cohort studies can include "any well-defined population", which "encompass(es) those that are defined by geographic boundaries […] (or) other criteria, such as membership in health maintenance organizations" (Szklo, 1998).

The most common healthcare utilisation component that could be identified was hospitalisation (n=11). This included the total number of hospital admissions and hospitalisation rate per year. Readmissions, on the other hand, were rarely captured (n=2) as most of the databases could not differentiate between the two. Length of stay, given as average or median, was the subject of 7 of the remaining papers.

The second most common measurement of healthcare utilisation (n=9) were the costs associated with the disease under study. These measures were reported as absolute, mean, incremental, annual and per-patient costs, as well as ratios compared to the general public or the per-patient expenditure before diagnosis.

Other forms of healthcare utilisation measurement used in the remaining papers (n=7) included the number and type of prescriptions and drugs dispensed, number of accidents and emergencies (A&E) as well as outpatient encounters with GPs, specialists, and other resource consumption indicators, such as number of CTs, laboratory tests, instrumental examinations, days spent in day-hospitals, ambulatory surgery, anaesthesia, dialysis, and imaging among others.

Only two studies compared the respective patients to the general population (Michet et al., 2015; C. T. Thorpe et al., 2008). All other studies either described healthcare utilisation in the cohorts or compared it over time, as seen in Raimundo et al. They compared the healthcare utilisation in patients with GPA and MPA in the first year after diagnosis to the second year following AAV diagnosis. Healthcare utilisation was measured included costs, associated with inpatients and outpatient health services, including visits to the emergency room.

Consistent over all included sources was the tendency of patients with systemic vasculitis to be characterised by a higher consumption of healthcare resources compared to the general population. Part of the literature indicated a magnitude of patients with AAV and systemic vasculitis to be roughly twice as costly for the health system as patients without the disease (Thorpe 2017; Raimundo 2015).

This observation was independent of the ways of measurement and definition of healthcare utilisation, for example whether it included prescription costs or not and whether the variables were elevated continuously or categorically. It also did not depend on the types of systemic vasculitis, which were examined, whether the definition was broad or specifically focussing on AAV. What did depend on both determinants, however, was the magnitude of the exceedance.

In short, when assigning the variety of researched diseases to the groups of small-, medium- and large-vessel vasculitis, there was a a gradient, indicating higher healthcare utilisation in small-vessel vasculitis compared to large-vessel vasculitis. Patients with large vessel vasculitis accordingly demonstrated only slightly elevated healthcare utilisation compared to the general population (Krulichova, Gamba, Ricci, & Garattini, 2004; Michet, Achenbach, Crowson, & Matteson, 2015). The largest differences were observed in patients with small-vessel vasculitis, with up to 56,642 USD per patient in a one-year period of follow-up (Cotch et al., 1996; Raimundo et al., 2015; Thorpe et al., 2018).

2.3 Limitations

There are multiple limitations regarding the informative value of the identified body of literature. First of all, five results (33.33% of total) represented grey literature and only consisted of abstracts without references, explained methodical approach and systematic derivation of results. Secondly, not all studies stated the total numbers of patients, whose data were included, which complicates the evaluation of results in terms of their informative value. In particular, this related to the largest identified study, conducted by Wallace et al., as they were investigating the nationwide trends in hospitalisation in the US (Wallace, 2016).

Additionally, all of the large cohort studies relied on claims data. Claims data rely on ICD-coded diagnoses when identifying the patients, which can introduce bias due to misclassification (Spencer, Mahtani, Brassey, & Heneghan, 2018). The problem there lies in the ICD-9 coding, which was seen in four studies. The ICD-9 classification of diseases did not have a separate code for patients with MPA, which inhibits the comparability of the results with other studies which considered MPA.

The use of claims data in healthcare research was discussed by Ferver et al., who weighed the pros and cons of studies using claims databases. Their review found that claims data were generally an appropriate foundation for the calculation of costs in for example cost-effectiveness studies, but coding was shown to be varying in quality (Ferver, Burton, & Jesilow, 2009).

Identified literature was further limited by the study locations with the majority of the studies coming from the USA. This also brings along another problem regarding the insurance situation of the patients, which differs vastly from European Health Systems. Some of the studies focus on patients insured by specific healthcare providers, for example the US federal health insurance program Medicare, rather than aiming for an all-encompassing approach. This in turn raises issues concerning the generalisability of results, as many patient populations of certain insurance providers do not represent the total population. Medicare is a good example in this regard, as its insurees are either older than 65 and without disability or younger than 65 with disability or end-stage renal disease (usa.gov).

Lastly, the quality assessment of the identified body of literature using the NOS resulted in on average 4.73 out of 9 possible stars, ranging from 3 to 6, indicating rather poor quality. This, in most of the studies, was due to a lack of "comparability of the cohorts on the basis of the design or analysis" (Wells GA O'connell D., 2011), meaning that it was controlled for important factors. Also, the point evaluating the adequacy of follow-up of cohorts was pitfalls for a lot of studies, as statements on such were often missing.

2.4 Conclusions

The results of the systematic literature review demonstrated the scarcity of research in the field of healthcare utilisation of patients with vasculitis in general, and especially of research focussing on patients with AAV.

The identified body of literature forms a good base for the capture of methods on how to measure healthcare utilisation. Among these, hospitalisation and costs were identified to be the most common types, given as mean, median, and rates compared to the general public or patient records prior to diagnosis.

Consistent over all studies was further the tendency of higher healthcare utilisation in patients with systemic vasculitis, when compared with the general population. The increase was highest in patients with small-vessel vasculitis and lowest in patients with large-vessel vasculitis.

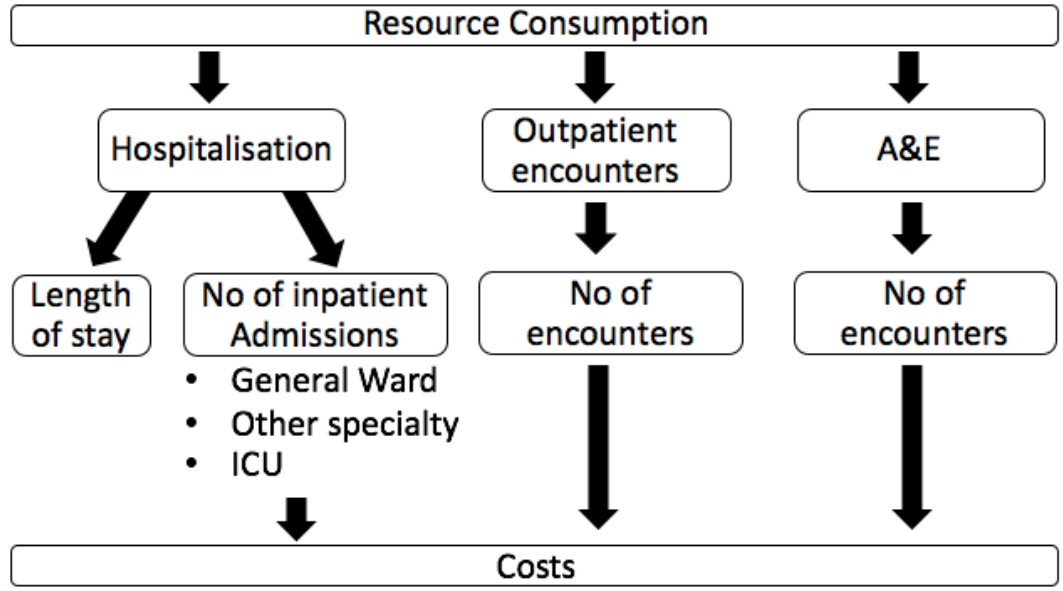
However, the included sources were characterised by numerous limitations regarding the study locations, included population as well as the disease classification and data base amongst others. Further research is therefore needed to assess healthcare utilisation in patients with AAV in Scotland.

3 Linked-data study

3.1 Methods

3.1.1 Outcomes of interest

Following the results of the literature review, the primary outcome healthcare utilisation was defined as the combination of health resource consumption and associated costs. Included parameters were hospitalisation, complemented by outpatient encounters and A&E, as depicted in the diagram below.



(ICU=Intensive Care Unit; A&E=Accidents & Emergencies)

Figure 4 Definition of Healthcare Utilisation, own representation

Hospitalisation encompasses the number of inpatient hospital admissions as well as the respective length of stay. It is distinguished between episodes on Intensive Care Units (ICU) and general wards, which include all specialties other than ICU.

Outpatient encounters include appointments with consultants, nurse led clinics and allied health professionals. They were captured as counts per patient and per year. A&E incidents were measured the same way.

The derivation of costs associated with healthcare utilisation serves two purposes. On one hand, costs summarise all the included healthcare utilisation parameters per person in a single variable. On the other hand, costs are the key figure, based on which healthcare can be planned by local decision makers. The translation of the healthcare utilisation parameters into costs on a per-patient-level is therefore not only an indicator for the total health economic burden of AAV but also is essential for the application of the results on public health level.

Lastly, it is of high interest to find out about driving factors for increased healthcare utilisation, to allow for better resource allocation. The identification of characteristics of high-risk patients in this regard is helpful for the alignment of patient care, favouring patient-relevant outcomes.

3.1.2 Data

For the purpose of the linked data study, NHS Scotland provided historic, administrative, non-identifiable data on routinely collected health records from seven centres across Scotland in collaboration with the Farr Institute for Health Informatics Research Scotland. Included were 543 Scottish AAV and up to five matches per patient from the general population (n=2671).

The primary data sources include the following:

- General Acute Inpatient and Day Cases – Scottish Morbidity Record (SMR01), including the period from 1995 to 2017
- Outpatient Appointments and Attendances – Scottish Morbidity Record (SMR00), including the period from 1997 to 2017
- A&E, including the period from 2007-2017
- Death records from National Records of Scotland (NRS), including the period from 1995-2017

Data were linked by the Information Services Division (ISD) Scotland, and matched based on sex, age (±2 years) and postal code. Those under the age of 16 were excluded from the study as well as patients whose diagnosis were found after death. Date of study entry was the individual date of diagnosis of the AAV-patients, which was also the entry date for their respective matches. Follow-up was until death or the 28.02.2017, whichever came first.

Data was accessible via the National Safe Haven, which, according to NHS Scotland, is "a secure environment supported by trained staff and agreed processes whereby health data can be processed and linked with other health data (and/or non-health related data) and made available in a de-identified form for analysis to facilitate research. It is a safeguard for confidential information which is being used for research purposes. Any researchers applying for access to health data must adhere to the Safe Haven principles" (NHS Research Scotland, 2018).

This includes a complex procedure regarding the release of results of any kind, required to be approved by the electronic Data Research and Innovation Service (eDRIS) within the ISD Scotland. As data on absolute patient counts may potentially be identifiable when it comes to very small numbers, results of this sort could not in all cases be disclosed for use outside of the Epidemiology Group. A (fictitious) example would be "3 male patients with AAV from the NHS Grampian were hospitalised in year 1998". This also affects parts of the assumption checking of the statistical analyses, for example scatter plots. Hence, these results can only be described, but not graphically presented. However, there were no major problems pertaining to the disclosure process for this thesis, due to the high number of patients and admissions involved in the study.

Ethics approval for the use of the non-identifiable data was granted by the NHS Research Ethics Committee, the Research & Development departments from all NHS health boards as well as from the Public Benefit and Privacy Panel for Health & Social Care. The application processes were managed by Shifa Sarica, PhD student at the Epidemiology Group, Aberdeen.

3.1.3 Statistical analyses

3.1.3.1 Descriptive statistics

All data analyses were performed with STATA version 14 (StataCorp, 2015). Descriptive analyses were used to obtain baseline characteristics, such as median age at index and follow-up time. Sex and deaths during follow-up in both cohorts were collected as percentage shares. Additionally, the represented types of AAV and ANCA were collected within the AAV cohort.

Healthcare utilisation parameters were calculated as rates per 1000 person-years included, from individual date of study entry until ten years of follow-up. The results were presented graphically in quarterly intervals. The respective STATA 14 code was developed in collaboration with Shifa Sarica, shared within the Epidemiology Group at the University of Aberdeen.

A discrete-time analysis was performed to investigate rate ratios over time by summarising the quarterly intervals to three periods; comprising the first 9 months, the following 4 years and the remaining 5.25 years until 10 years of follow-up. Due to very few incidents near the end of follow-up, results on A&E rate ratios could only be disclosed up until 8 years of follow-up. 95%CIs for the incident rate ratios were calculated using OpenEpi version 3.01 (Dean, Sullivan, & Soe, 2013). The rate ratios of the three periods and according 95%CIs are displayed underneath the resulting graphs as well as the numbers of patients at risk per year of study.

The NHS Scottish Costs Book was used to obtain annual tariffs, which were multiplied with the resource consumption, captured as counts of the respective healthcare utilisation parameter for each year of study (Information Services Division Scotland, 2017). Tariffs were inflated to 2016 values using the Hospital and Community Health Service (HCHS) Index. Inaccessible data on tariffs of years earlier than 2002 were estimated by using the latest known tariff (2002) as reference for deflation. Support was provided by the Health Economics Research Unit (HERU) of the University of Aberdeen.

The absolute costs were calculated per person-year per cohort as well as the incremental cost per person-year. A cost ratio was calculated, comparing patients with AAV to the general population. 95%CIs were computed using Fieller's theorem in GraphPads QuickCalcs tool (Graphpad Software, 2018). Kruskal-Wallis test was applied to test the mean differences in costs between the two cohorts.

3.1.3.2 Regression Analyses

For the purpose of identifying driving factors for elevated healthcare utilisation parameters post diagnosis on one hand and elevated costs post diagnosis on the other hand, assumptions of different models were tested. The purpose of all regression models developed in this study

was description rather than prediction. This is important to denote, as it had an impact on the choice of methods and solutions in case of violated assumptions.

Predictor variables of all regression analyses included the following:

- Age (interval scale)
- Sex (dichotomous: male, female)
- AAV-type (categorical: GPA, MPA, EGPA)
- ANCA-status (categorical: PR3, MPO, ANCA negative)
- Socio-economic status (Scottish Index of Multiple Deprivation (SIMD) quintiles, categorical: ranging from Q1 "most deprived" to Q5 "most affluent")
- NHS health boards the patients are assigned to geographically (categorical: NHS Grampian, NHS Lothian, NHS Greater Glasgow & Clyde, NHS Tayside, NHS Highland, NHS Fife)

The univariable analyses served as precursors of the multivariable models. The purpose for this was not the exclusion of predictors in the multivariable analysis, which were non-significant in the univariable analysis. Rather, the aim was to understand the general direction and link between dependent and independent variables and to control for confounders. This procedure was found in several other health economic studies of various medical fields (Barkun, Adam, Martel, & Bardou, 2013; Bloudek et al., 2012; Nguyen & Gordon, 2015; Urueña et al., 2015; Westerhout, Verheggen, Schreder, & Augustin, 2012).

Poisson Regression

Poisson regression describes "the sampling distribution of the number of occurences" (Kirkwood & Sterne, 2003). It is useful for large cohorts and when the "data […] is expressed as events per person-years of observation" (Dupont, 2009). Hence, Poisson regression was the method of choice for the identification of driving factors for each of the healthcare utilisation parameters *incident hospital admissions (general ward)*, *length of stay on a general ward*, *ICU episodes*, *length of stay on ICU*, *outpatient encounters* and *A&E*.

The counts of each parameter were calculated for patients with AAV over the total follow-up including up to 22 years per patient. Model assumptions were reviewed, based on the recommendations of the Institute for Digital Research and Evaluation (UCLA: Statistical

Consulting Group, 2018a). Accordingly, data must be count data and independent variables shall be either continuous, ordinal or nominal including dichotomous variables.

Further, observations must be independent, and the distribution of the count data must follow the typically right-skewed Poisson distribution, which was tested for using histograms (Bali, 2016). As a result of the Poisson distribution, the mean and the variance must be equal, which was accounted for by analysing descriptive statistics (Haight, 1967). Pearson goodness-of-fit statistics was further used to assess model fit for the data. A p-value of higher than 0.05 was interpreted as well-fitting.

In the following, univariable and multivariable Poisson regression models were developed for each healthcare utilisation parameter. The predictor variables for the multivariable regression were selected backwards and automatically, targeting a significance level of $p<0.05$. Robust standard errors were used to adjust for heteroscedasticity (Gujarati, 2018).

Linear Regression

The assumptions for the linear regression analyses were checked based on the recommendations of the Institute for Digital Research and Evaluation (UCLA: Statistical Consulting Group, 2018b).

Accordingly, residuals must be normally distributed. Using the actual sample observations, this assumption was heavily violated. Therefore, the dependent variable *costs post diagnosis* was log-transformed, using the natural logarithm. Log-transformation is known to better meet linear regression assumptions, in especially the normal distribution of residuals (Curran-Everett, 2018).

Secondly, homoscedasticity is an important assumption, which was tested for using Breusch-Pagan test. Eventually, robust standard errors were used, just as in the Poisson regression, because they are "asymptotically […] valid in the presence of any form of heteroscedasticity as well as homoscedasticity" (Gujarati, 2018).

As a means of testing for multicollinearity, variance inflation factors (VIF) were determined for the included predictor variables. A rule of thumb suggests VIFs>10 to be worrisome. However, other authors raise concern, when VIFs are higher than 2.5 (Williams, 2015). Clinical evidence was consulted in such a case of VIFs between 2.5 and 10, in order to decide for the inclusion or exclusion of a predictor variable.

Univariable and multivariable linear regression models were created to investigate influencing factors on costs following an AAV diagnosis. Predictor variables were manually selected via backwards elimination, based on a significance level of $p<0.05$. The general rule of thumb of a 10% change in effect size was applied to control for confounding. This means that non-significant predictors were kept in the model, if they changed the effect sizes by more than 10% (Hernán, Hernández-díaz, Werler, & Mitcheil, 2002).

3.2 Results

3.2.1 Baseline characteristics and follow-up

A total number of 543 patients with AAV and 2672 matches from the general population were included in the linked data study. The included patients came from six different NHS health boards, NHS Grampian, NHS Lothian, NHS Greater Glasgow and Clyde, NHS Tayside, NHS Highland and NHS Fife. Figure 2 below shows the geographic distribution of these health boards.

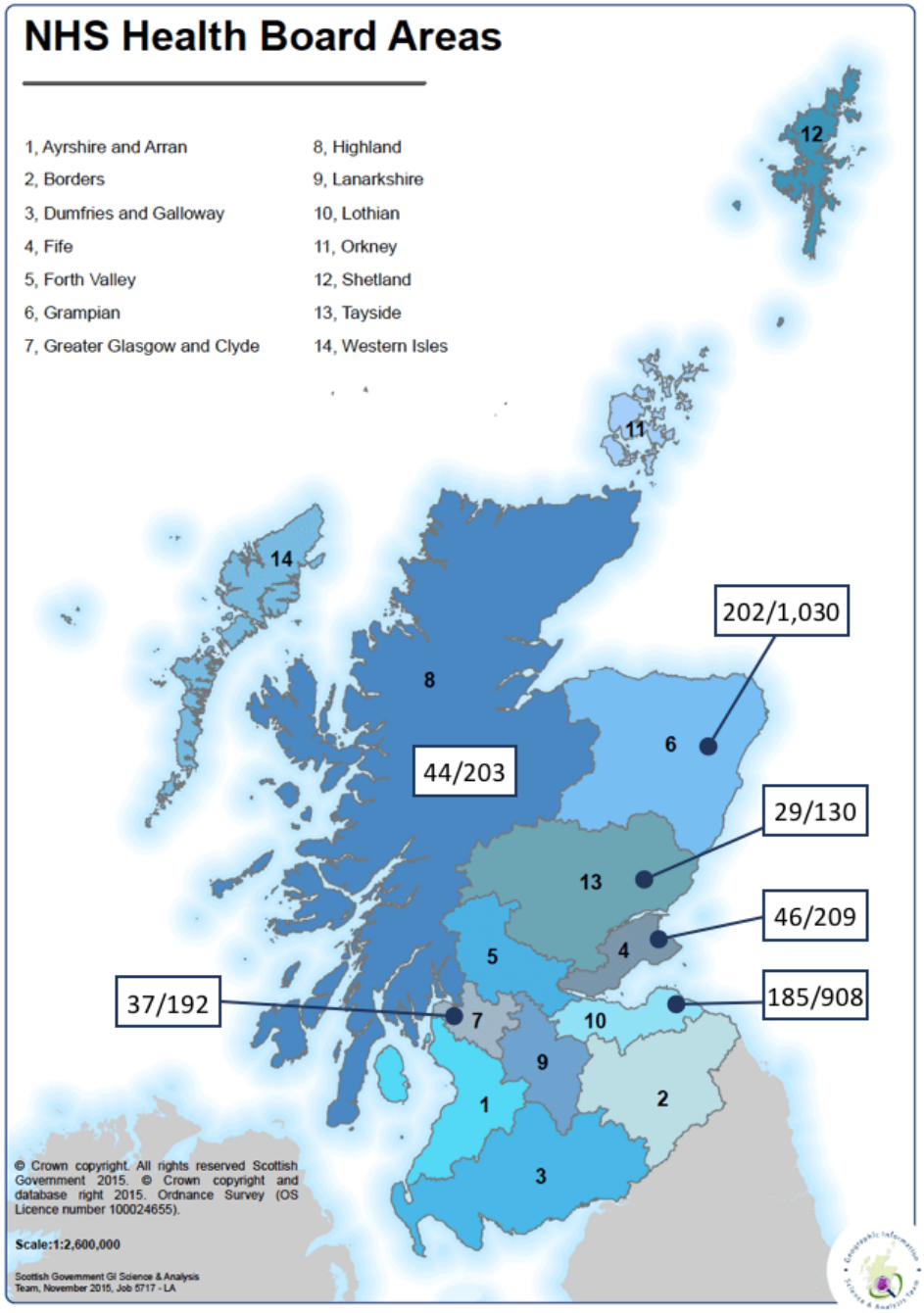


Figure 2 NHS Health boards of study participants (n, AAV/Non-AAV), based on Scottish Government, 2018

The earliest date of study entry was 01.01.1995, the latest entry date was 28.11.2016. Of the AAV cases, 53.59% were male, compared to 53.67% in the matches. Median follow-up time was 5.06 years in the AAV cohort (IQR 2.47-9.35 years), and 5.16 years (IQR 2.53-9.46) in the matched cohort.

Median age at study entry was 58.74 (IQR 48.93-67.99) in the AAV cohort, and 58.68 (IQR 48.98-67.92) in the matches. In the AAV-cohort, 18.93% of cases died during follow-up, as did 16.62% in the general population cohort.

Regarding socio-economic status, roughly half of the patients with AAV (50,83%; n=276) and matches (43.12%; n=1,152) were assigned to the quintiles 4 and 5 of the Scottish Index of Multiple Deprivation, which are the most affluent quintiles. Another 20.81% (n=113) of AAV cases and 16.06% (n=429) of matches account for the middle class, quintile 3. 27.44% (n=149) of AAV cases and 23,43% (n=626) general population matches, respectively, were assigned to quintiles 1 and 2, being the most deprived.

Among AAV cases, 58.2% had GPA, followed by MPA (28.91%) and EGPA (12.52%). Regarding the types of anti-bodies, PR3 was found in 52.67% of cases, 34.62% of cases showed MPO antibodies and 11.97% were ANCA negative. Percentage shares can be found in the pie charts on the right-hand side.

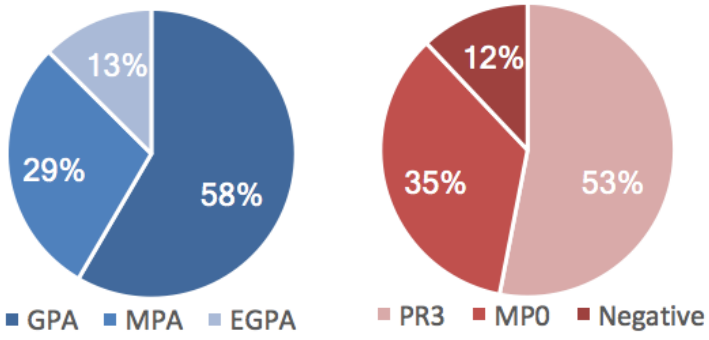


Figure 3 Percentage shares of AAV types (left) and ANCA in the AAV cohort

3.2.2 Rate ratios of healthcare utilisation

The following sections present the comparison of the healthcare utilisation parameters *incident hospital admissions*, *length of stay on a general ward*, *incident ICU episodes*, *length of stay on ICU*, *outpatient encounters* and *A&E* between patients with AAV and their matches from the general population.

The graphs show rate ratios per 1000 person-years included, in quarterly intervals. The tables underneath the graphs show the number of patients at risk per year of follow-up as well as the rate ratios from the discrete time analysis. Interval zero represents the individual date of diagnosis for patients with AAV, hence, date of study entry.

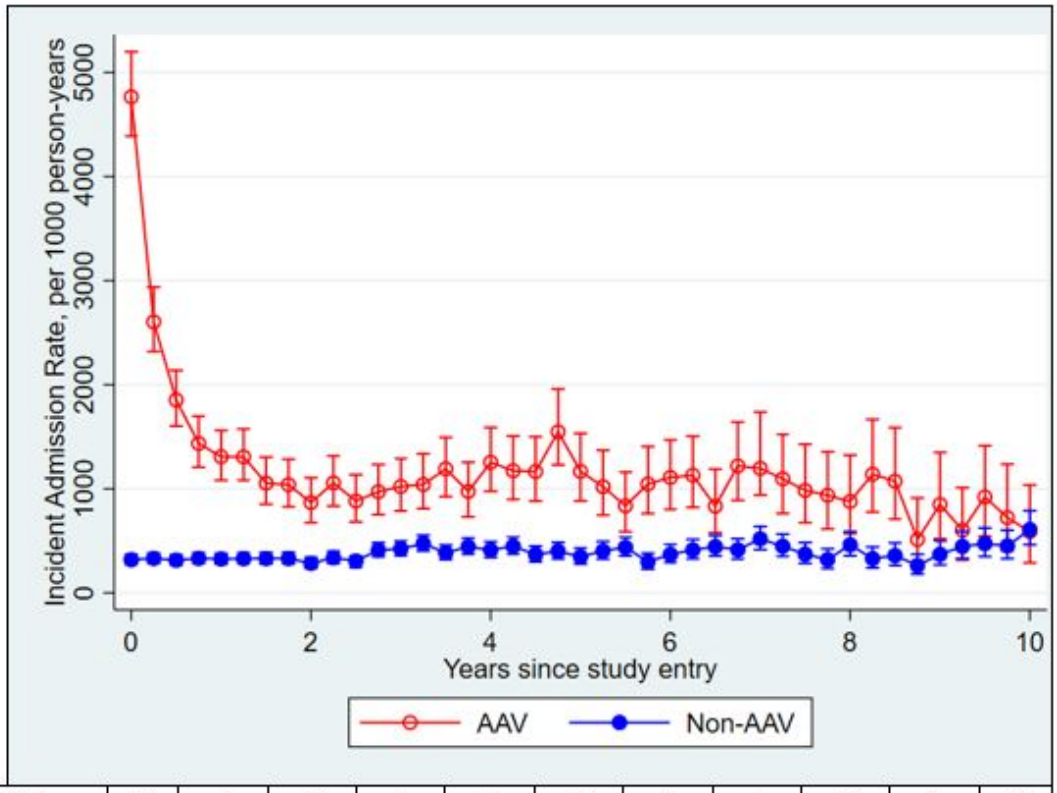
Incident hospital admissions and length of stay (general ward)

Post diagnosis, patients with AAV had on average 6.81 (SD=7.51) incident hospital admissions, ranging from 0 to 95. Median length of stay was 20.5 (IQR=8-43, Mean=36.82, SD=57,10).

Figure 4 below compares the incident admissions of patients with AAV compared to their general population matches, given as a quarterly rate per 1000 person-years included. Over the course of the study, the incident admission rate remained consistently higher in the AAV cohort, with statistical significance.

The first 9 months after diagnosis are particularly pronounced, with a rate ratio of 8.51 (95%CI=7.73-9.34) compared to the matches. In interval zero, patients with AAV had close to 4800 incident admissions per 1000 person-years, or rather 4.8 admissions per person-year included. The rate ratio decreases remarkably strong within the first two years after diagnosis, and then increases again, stagnating at about one admission per person-year included in patients with AAV.

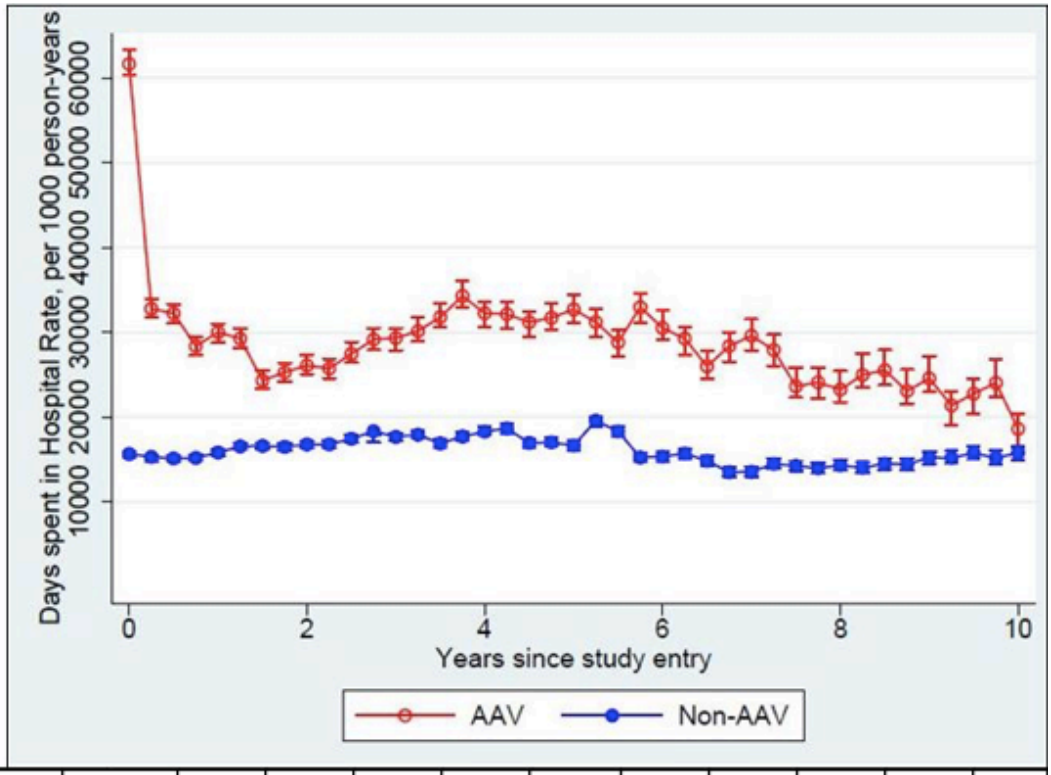
In the last 5 to 10 years of follow-up, AAV-patients still showed 2.44 (95%CI=2.22-2.68) times more inpatient hospital admissions than their matches.



Year		0	1	2	3	4	5	6	7	8	9	10
Patients at risk	AAV	543	501	439	375	318	273	233	207	176	146	117
	Non-AAV	2672	2493	2198	1868	1590	1362	1161	1027	888	732	590
Rate Ratio		8.51	3.02				2.44					
95% CI		7.73-9.34	8.82-3.24				2.22-2.68					

Figure 4 Incident hospital admission rate (general ward), per 1000 person-years

Additionally, it was found that AAV-patients are not only admitted more frequently, but also stayed in the hospital for alonger duration than the general population. In the first 9 months, patients with AAV spent 2.58 (95%CI=2.54-2.63) times more days in hospital than their matches, with over 60 days spent in hospital per person-year included. This rate ratio decreased to 1.77 (95%CI=1.74-1.8) in 5 to 10 years of follow-up, as can be seen in Figure 5 below. Likewise the incident admissions graph, the rate on length of stay on a general ward increases after about two years of follow-up, with a second peak at approximately four years of follow-up.

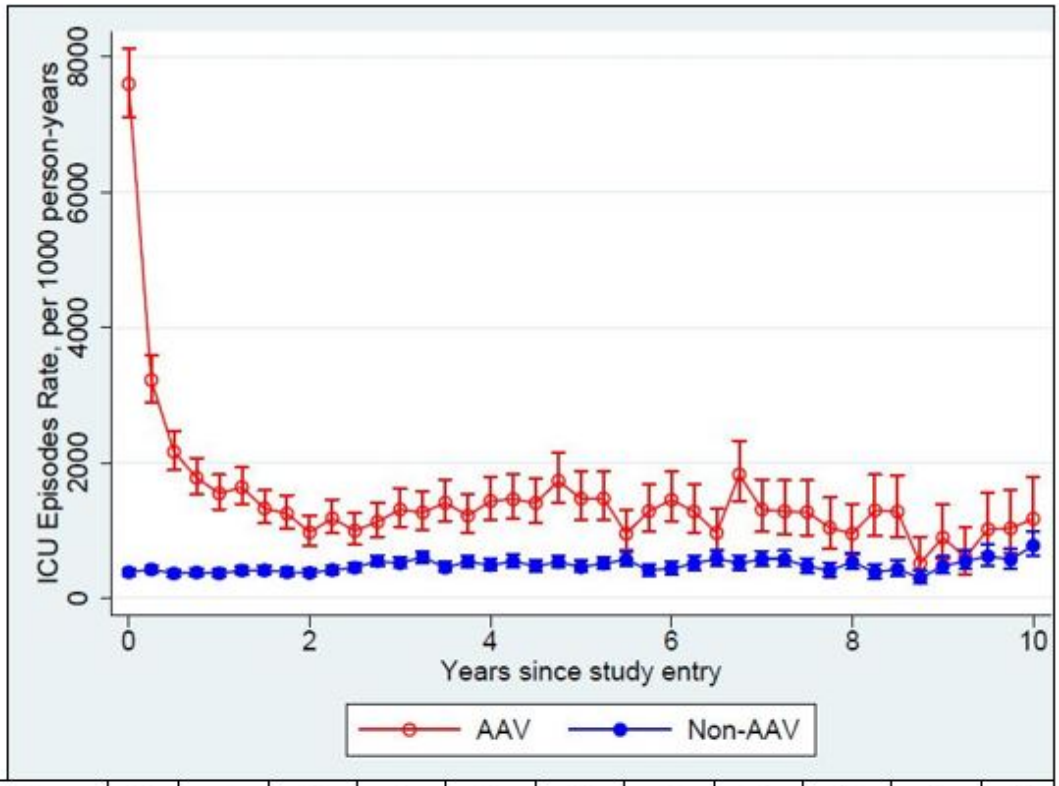


Year	0	1	2	3	4	5	6	7	8	9	10
Rate Ratio	2.58	1.69				1.77					
95% CI	2.54-2.63	1.67-1.71				1.74-1.8					

Figure 5 Days spent on a general ward, per 1000 person-years

Episodes and length of stay on ICU

The rate on incident ICU episodes shows a picture similar to that of the incident hospital admissions on a general ward, as can be seen in Figure 6 below.



Year		0	1	2	3	4	5	6	7	8	9	10
Patients at risk	AAV	543	501	439	375	318	273	233	207	176	146	117
	Non-AAV	2672	2493	2198	1868	1590	1362	1161	1027	888	732	590
Rate Ratio		9.96	2.89				2.4					
95% CI		9.15-10.82	2.71-3.08				2.21-2.62					

Figure 6 ICU Episodes Rate, per 1000 person-years

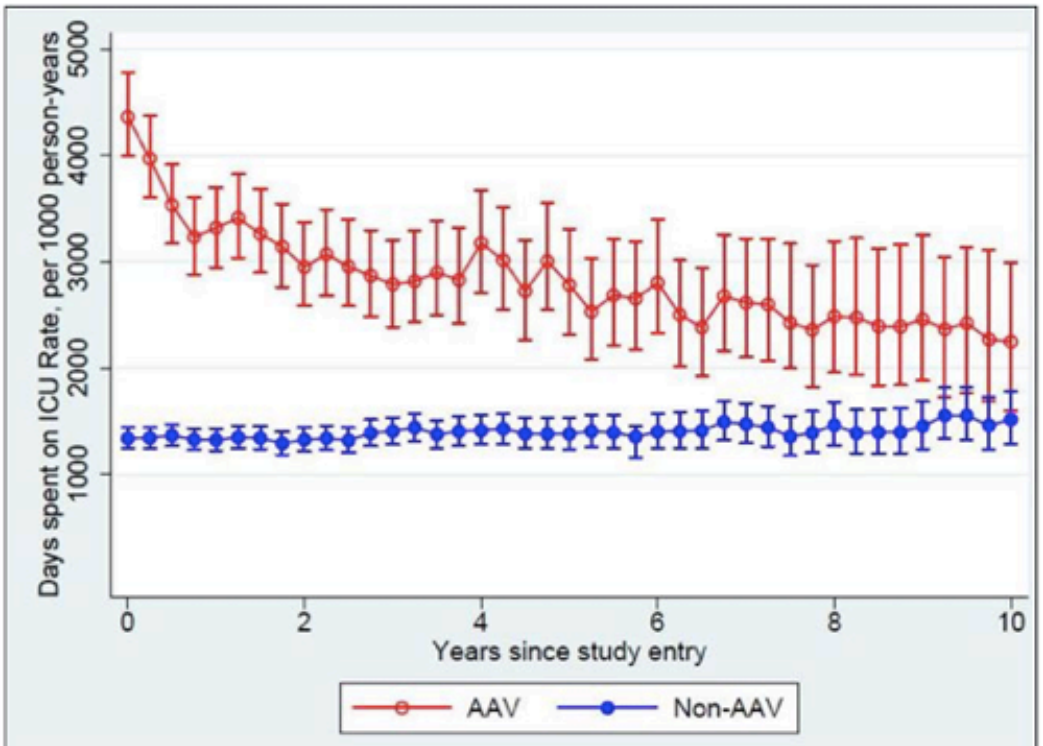
In the first 9 months, AAV-cases had close to 10 times more episodes on an ICU, compared to their matches, showing 7.6 episodes on ICU per person-year included at the time of diagnosis.

Likewise the incident admission rate on general wards, the rate in patients with AAV increases again after two years, and then stagnates at about 1.2 episodes per person-year included. In the last 5 to 10 years of follow-up, there is still a 2.4 (95%CI=2.21-2.62) times higher rate on incident ICU episodes in AAV-patients compared to the general population.

Similar to the analysis on length of stay on a general ward, patients with AAV also show longer stays on ICU, with 2.83 (95%CI=2.67-3.01) times longer stays on ICU in the first 9 months, compared to their matches. See Figure 7 below. The mean length of stay on an ICU was 2.21 (SD=6.04) days in patients with AAV. Interval zero demonstrates 4.3 days spent

on ICU per person-year included in patients with AAV, compared to 1.3 days in the general population matches. over the course of follow-up. Over time, the rate ratios decrease, still demonstrating that the patients with AAV show 1.78 (95%CI=1.69-1.88) times increased length of stay on ICU.

Only very slightly, but clearly visibly, is the increasing course of the rate on length of stay on ICU in the general population. This is probably explicable simply by the increasing age at the end of follow-up. The matches accordingly showed 1.3 (95%CI=1.2-1.4) days on ICU per person-year included in the first interval, compared to 1.5 (95%CI=1.3-1.8) days per person-year in the last interval.



Year	0	1	2	3	4	5	6	7	8	9	10
Rate Ratio	2.83	2.23				1.78					
95% CI	2.67-3.01	2.14-2.33				1.69-1.88					

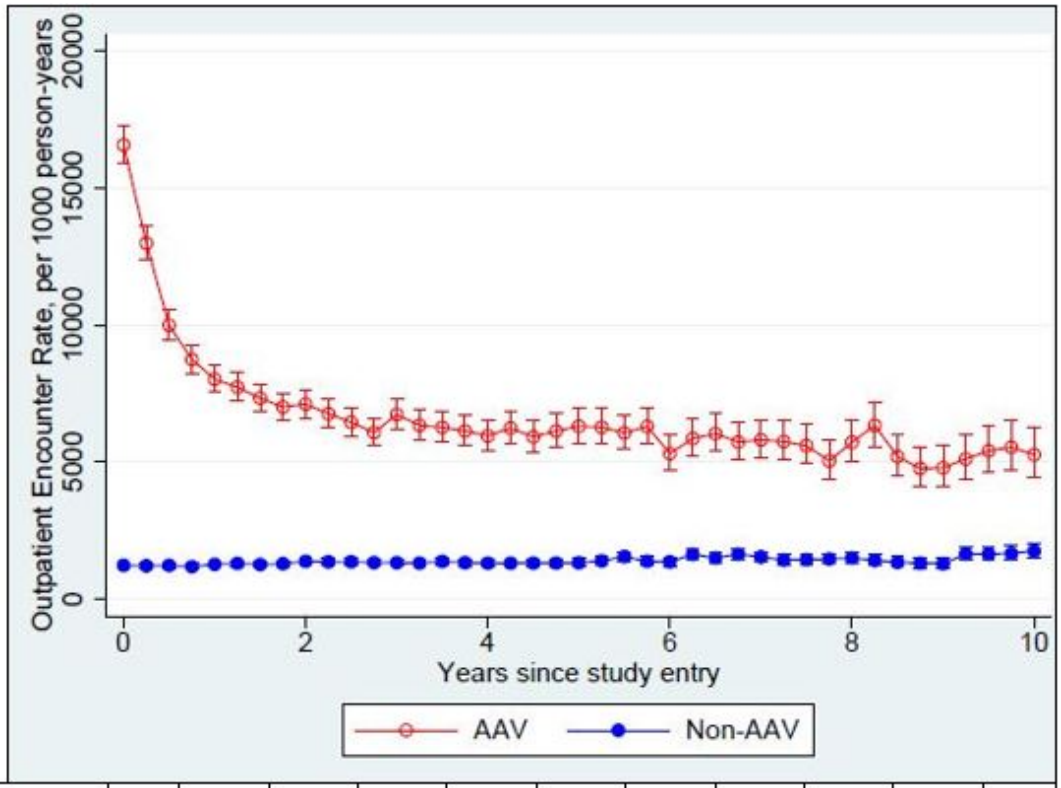
Figure 7 Days spent in Intensive Care Unit (ICU)

Outpatient encounters

Patients with AAV had a mean of 43.10 (SD=32.73) outpatient encounters during follow-up, ranging from 0-266. The largest difference between patients with and without AAV was observed in the rate of outpatient encounters. Again, the first nine months after diagnosis were especially pronounced.

Following Figure 3 below, the AAV cohort had 10.01 (95%CI=9.57-10.48) times more outpatient encounters than their matches in the first period of follow-up. At the time of diagnosis and study entry, patients with AAV had 16.6 outpatient encounters per person-year included. In comparison, only 1.2 outpatient encounters per person-year included were shown in the general population at that time.

The rate ratio decreased over time, and stagnated after about four years of follow-up. However, even in the final period of 5 to 10 years of follow-up, the rate ratio remains high, showing an increase in the number of outpatient encounters by a factor of 3.89 (95%CI=3.74-4.04) in patients with AAV. This translates to approximately 5 outpatient encounters per person-year included.



Year		0	1	2	3	4	5	6	7	8	9	10
Patients at risk	AAV	533	491	429	365	308	263	223	197	166	136	107
	Non-AAV	2622	2443	2148	1818	1540	1312	1111	977	838	682	540
Rate Ratio		10.01	5.13				3.89					
95% CI		9.57-10.48	4.98-5.28				3.74-4.04					

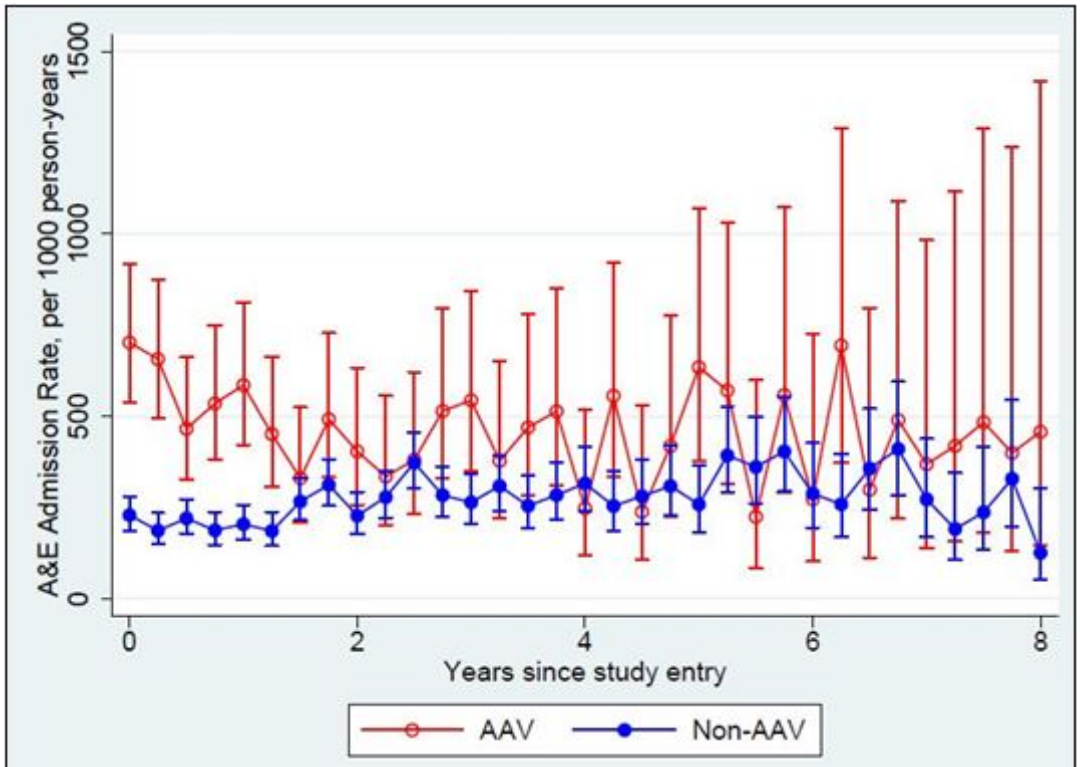
Figure 8 Outpatient Encounter Rate, per 1000 person-years

Accidents and emergencies

On average, patients with AAV had 0.12 (SD=0.47) A&E episodes in the period prior to diagnosis and 1.93 (SD=2.58) episodes after diagnosis. Due to the scarcity of events, the results on rate ratios between the cohorts could only be released for eight years of follow-up, as opposed to the ten years included in the other parameters. Figure 9 below shows the results, which are less easy to interpret than the results of the other healthcare utilisation parameters.

In the first period of follow-up, the number of A&E is significantly increased in the AAV-cohort with 0.7 (95%CI=0.54-0.92) A&E incidents per person-year included. This results in a rate ratio of 2.58 (95%CI=2.54-2.63).

The individual quarterly rates, however, are insignificant after 1.25 years of follow-up. Nonetheless, the discrete time analysis shows that over time, there is still a statistically significant rate ratio of 1.77 (95%CI=1.74-1.78) in the final follow-up period.



Year		0	1	2	3	4	5	6	7	8
Patients at risk	AAV	407	366	305	241	186	144	106	83	57
	Non-AAV	1999	1824	1532	1202	931	720	532	412	285
Rate Ratio		2.58	1.68				1.77			
95% CI		2.54-2.63	1.67-1.71				1.74-1.8			

Figure 9 A&E Admission Rate, per 1000 person-years

3.2.3 Costs

When translating healthcare utilisation parameter counts into costs, absolute costs for the cohort and per person-year costs are differentiated, to consider different time periods included in the data sets. The A&E dataset contributed 8 years of follow-up, compared to up to 18 years in the SMR00 and SMR01 datasets.

After diagnosis (and study entry), the absolute costs for inpatient and outpatient data amounted to £23,346,544.70 in patients with AAV, compared to £23,813,983.10 in the general population matches, including data from 2000-2017. Regarding A&E, the absolute costs in patients with AAV amounted to £103,216.27 post diagnosis, compared to £295,260.27 including data from 2007-2017. Hence, the absolute costs per cohort in the present study amounted to £23.449.760,97 in patients with AAV, compared to £24.109.244,07 in the general population matches, as seen in Table 1.

This table highlights the excess healthcare utilisation in the smaller cohort of patients with AAV, showing costs almost as high as that of the general population cohort, which comprises four times as many study participants.

Table 1 Costs per cohort due to healthcare utilisation (2000-2017*)

Cohort		AAV		General population
n (follow-up, years)		502 (2819,52)		2043 (12097,02)
One year prior to study entry				
LOS (general ward)	£	453.065,12	£	1.346.302,25
LOS (ICU)	£	802.567,23	£	922.227,60
Outpatient encounters	£	80.996,47	£	163.578,46
A&E*	£	6.448,73	£	17.096,88
Costs per cohort	£	1.343.077,55	£	2.449.205,19
Post study entry				
LOS (general ward)	£	10.187.997,10	£	13.729.308,50
LOS (ICU)	£	9.889.830,60	£	7.548.986,60
Outpatient encounters	£	3.268.717,00	£	2.535.688,00
A&E*	£	103.216,27	£	295.260,97
Costs per cohort	£	23.449.760,97	£	24.109.244,07

*A&E (2007-2017): AAV (n=469, follow-up=2745.69); General population (n=1832, follow-up=11582.26)

On person level, the costs accounting for healthcare utilisation parameters one year prior to study entry were £2,676.36 in the AAV cohort. Accordingly, patients with AAV were 2.23 times higher compared to the general population matches, as seen in Table 2.

After diagnosis, these costs increased by a factor of 3.11 in the AAV cohort, adding up to £8,317.91 per person-year, whereas the costs of the general population cohort amounted to £1,994.07 per person-year.

Table 2 Healthcare utilisation costs per patient-year (2000-2017*)

Cohort		AAV		General population
n (follow-up, years)		502 (2819,52)		2043 (12097,02)
One year prior to study entry				
LOS (general ward)	£	902,52	£	658,98
LOS (ICU)	£	1.598,74	£	451,41
Outpatient encounters	£	161,35	£	80,07
A&E*	£	13,75	£	9,33
Costs per person-year	£	2.676,36	£	1.199,79
Post study entry				
LOS (general ward)	£	3.613,38	£	1.134,93
LOS (ICU)	£	3.507,63	£	624,04
Outpatient encounters	£	1.159,32	£	209,61
A&E*	£	37,59	£	25,49
Costs per person-year	£	8.317,92	£	1.994,08

*A&E (2007-2017): AAV (n=469, follow-up=2745.69); General population (n=1832, follow-up=11582.26)

(A&E=Accidents & Emergencies; LOS=Length of stay)

After diagnosis, patients with AAV were accordingly 4.17 ($p<0.0001$) times more expensive compared to the matched cohort from the general population. The incremental costs per person-year were therefore £6,323.84 (95%CI=£1,727.82-£10,919.87). These are the additional costs per person-year of each patient with AAV, compared to the general population.

Inpatient stays on a general ward accounted for 85% (£5,362.04) of these extra costs, 39% (£2,478.45) for stays on general wards and 46% (£2,883.59) for days spent on ICU. Outpatient encounters accounted for 15% (£949.79) of the incremental costs, whereas A&E did not contribute as a percentage share (£12.10). Table 1 comprises the detailed costs per person-year for each of the healthcare utilisation parameters for patients with AAV and their general population matches.

To summarise, the descriptive cost data shows 4.17 ($p<0.0001$) times higher costs in patients with AAV compared to their matches from the general population over the course of up to 17 years of follow-up. Stays on ICU account for the largest percentage share of the incremental costs.

3.2.4 Regression analyses

3.2.4.1 Poisson regression analyses

The objective of the conducted Poisson regression analyses was to examine the existence and directions of possible relationships between the included predictor variables and the healthcare utilisation parameters.

Therefore, model assumptions were reviewed, beginning with the type of data included in the analyses. This assumption was met, as the dependent variables were measured in counts (UCLA: Statistical Consulting Group, 2018a) and the independent variables are either continuous (*age*) or categorical (*sex*, *AAV type*, *ANCA status*, *socio-economic status*, *NHS health board*).

Independence of observations can be assumed, as the cohort includes only single individuals. Further, the distribution of the data plays an important role, which was tested for using histograms (Bali, 2016). Figure 10 below exemplarily shows the histogram of length of stay on a general ward. It clearly shows a right-skewed distribution of counts of the number of days spent in hospital.

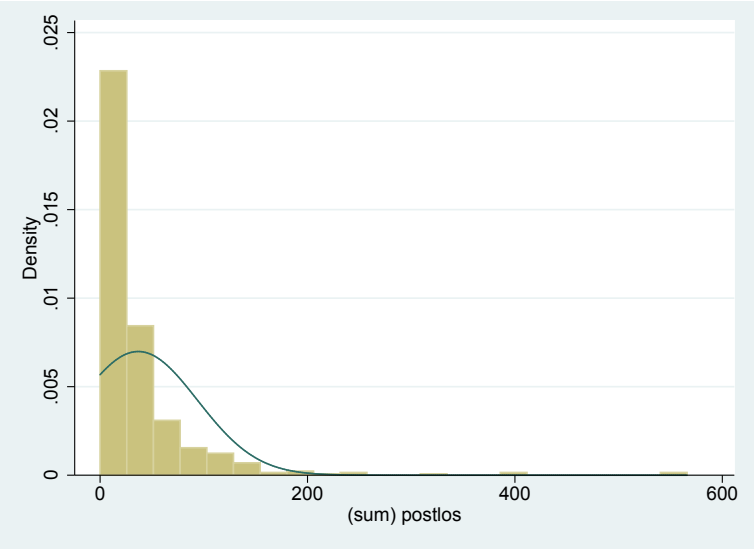


Figure 10 Histogram of days spent on a general ward, post diagnosis

The histograms of the remaining parameters show a similar distribution and can be found in appendix II. As a consequence of the Poisson distribution, the mean and the variance must be equal, which was tested for by analysing descriptive statistics. However, the means and the variances of the healthcare utilisation variables are not the same in any case, which indicates a poor fit of the model for the data (Haight, 1967).

The insufficient fit of the model was confirmed by Pearson goodness of fit statistics, showing $p<0.0001$ for all parameters. Nonetheless, the Poisson regression results are described in the following, exploring the existence and directions of the relationships between included predictors and healthcare utilisation parameters.

Inpatient admissions and length of stay (general ward)

Table 2 below shows the incidence rate ratios (IRR) of the univariable and multivariable Poisson regression, examining the influence of *age*, *sex*, *AAV type*, *ANCA status*, *socio-economic status* and *NHS health board* on inpatient hospital admissions.

The univariable analyses revealed that age is significantly associated with the number of incident hospital admissions, with an IRR of 1.01 (95%CI=1-1.02; $p<0.01$). Sex, however, was of minor importance, women had only slightly increased, but not statistically significant IRR compared to men (IRR=1.09; 95%CI=0.9-1.32; $p=0.4$).

Regarding the type of AAV, patients with EGPA had significantly less hospital admissions compared to patients with GPA (IRR=0.63; 95%CI=0.47-0.85; p<0.01). The results for MPA, as well as the ANCA statuses MPO and negative, compared to PR3, were inconclusive. Patients with higher socio-economic statuses had less incident hospital admissions compared to the most deprived SIMD quintiles, significantly, however, only for quintile 4 (IRR=0.63; 95%CI=0.48-0.84; p<0.01).

Looking at geographic differences, patients from the NHS Greater Glasgow and Clyde, (IRR=1.45; 95%CI=1.03-2.05; p=0.03) as well as from the NHS Fife (IRR=1.75; 95%CI=1.27-2.43; p<0.01) had statistically significantly higher IRR for inpatient hospital admissions than patients from the NHS Grampian. For all other health boards, the results were non-significant, but still indicating more hospital admissions in patients from Lothian and Tayside compared to Grampian, whereas patients from NHS Highland had slightly less admissions.

Table 3 Poisson regression analyses exploring incident hospital admissions (2000-2017)

			Univariate analyses		Multivariate analysis	
HCU Parameter	Predictor (Base)	Subgroup	IRR (95% CI)	p-value	IRR (95% CI)	p-value
Hospital admission (general ward) (n=494)	Age		1.01 (1-1.02)	<0.01	1.01 (1-1.02)	<0.01
	Sex (Male)	Female	1.09 (0.9-1.32)	0.4	-	-
	AAV Type (GPA)	MPA	1.11 (0.91-1.36)	0.29	-	-
		EGPA	0.63 (0.47-0.85)	<0.01	-	-
	ANCA Status (PR3)	MP0	0.94 (0.79-1.13)	0.54	-	-
		ANCA negative	1.01 (0.65-1.56)	0.98	-	-
	SES Quintiles (Q1)	Q2	0.81 (0.59-1.1)	0.18	-	-
		Q3	0.75 (0.57-1)	0.05	-	-
		Q4	0.63 (0.48-0.84)	<0.01	-	-
		Q5	0.8 (0.57-1.11)	0.19	-	-
	Health board (Grampian)	Lothian	1.21 (0.95-1.52)	0.12	1.21 (0.96-1.53)	0.11
		Glasgow and Clyde	1.45 (1.03-2.05)	0.03	1.64 (1.16-2.32)	<0.01
		Tayside	1.03 (0.78-1.37)	0.81	1.02 (0.78-1.34)	0.86
		Highland	0.94 (0.72-1.24)	0.68	0.98 (0.76-1.26)	0.89
		Fife	1.75 (1.27-2.43)	<0.01	1.83 (1.3-2.56)	<0.01

(ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; PR3=Proteinase 3)

The multivariable Poisson analysis included age and health board, with 494 observations. Included predictors were age and NHS health board. Accordingly, for each additional year of age, the number of inpatient hospital admissions increased by 1% (IRR=1.01; 95%CI=1-1.02; p<0.01).

Similar to the univariable analysis, results were significant only for the NHS health boards Greater Glasgow and Clyde and Fife, which showed increased inpatient hospital admissions of 64% (IRR=1.64; 95%CI=1.16-2.32; p<0.01) and 83% (IRR=1.83; 95%CI=1.3-2.56; p<0.01), respectively, compared to patients from the NHS Grampian.

The univariable as well as the multivariable Poisson analysis on length of stay on a general ward revealed a significant association between the age of AAV patients and their length of stay on a general ward. In both models, one year increase in age was associated with an increased length of stay by 4%. This is an interesting result, given that the median length of stay in AAV patients post diagnosis was 20.5 days (IQR=8-43, Mean=36.82, SD=57.10). A ten year increase in age would consequently amount to an increase in the length of stay of 8.2 days per AAV patient.

The univariable analysis further showed that patients with MPA stayed significantly longer on a general ward compared to patients with GPA (IRR=1.42; 95%CI=1.06-1.89; p=0.02), whereas patients with EGPA stayed less long, however, not significantly. None of the results regarding the ANCA status or the socio-economic status were statistically significant.

Looking at geographical differences, patients from the NHS Lothian (IRR=0.68; 95%CI=0.5-0.92; p=0.01) and Tayside (IRR=0.48; 95%CI=0.32-0.73; p<0.01) stayed significantly less long on a general ward, compared to patients from the NHS Grampian. The same results were confirmed in the multivariable analysis. The multivariable analysis on length of stay hence included the predictors age and NHS health board, just like the analysis on the incident hospital admissions, including 494 observations.

Table 4 Poisson regression analyses exploring length of stay on a general ward (2000-2017)

			Univariate analyses		Multivariate analysis	
HCU Parameter	Predictor (Base)	Subgroup	IRR (95% CI)	p-value	IRR (95% CI)	p-value
LOS (general ward) (n=494)	Age		1.04 (1.03-1.05)	<0.01	1.04 (1.03-1.05)	<0.01
	Sex (Male)	Female	1.28 (0.96-1.69)	0.09	-	-
	AAV Type (GPA)	MPA	1.42 (1.06-1.89)	0.02	-	-
		EGPA	0.67 (0.45-1)	0.05	-	-
	ANCA Status (PR3)	MP0	1.15 (0.86-1.54)	0.36	-	-
		ANCA negative	0.76 (0.53-1.1)	0.15	-	-
	SES Quintiles (Q1)	Q2	1.05 (0.66-1.65)	0.85	-	-
		Q3	1.1 (0.71-1.72)	0.67	-	-
		Q4	0.85 (0.55-1.31)	0.47	-	-
		Q5	0.83 (0.57-1.23)	0.35	-	-
	Health board (Grampian)	Lothian	0.68 (0.5-0.92)	0.01	0.67 (0.5-0.9)	<0.01
		Glasgow and Clyde	0.82 (0.5-1.33)	0.42	1.2 (0.75-1.9)	0.45
		Tayside	0.48 (0.32-0.73)	<0.01	0.48 (0.33-0.7)	<0.01
		Highland	0.7 (0.44-1.12)	0.14	0.76 (0.48-1.21)	0.25
		Fife	0.94 (0.55-1.58)	0.8	0.97 (0.58-1.65)	0.92

(ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; LOS=Length of stay; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; PR3=Proteinase 3)

ICU episodes and length of stay on ICU

Both, the number of ICU episodes and length of stay on ICU were modelled by ANCA status only, with 494 observations included in each model. The sole statistically significant result from the multivariable analyses stated that patients with negative ANCA status stayed significantly less long on ICU (IRR=0.36; 95%CI=0.17-0.75, p<0.01) compared to patients with PR3. The same result was shown in the univariable analysis (IRR=0.36; 95%CI=0.17-0.76, p<0.01)

Additionally, the univariable analyses revealed a significant relationship between the type of AAV and the number of ICU episodes. Accordingly, patients with MPA had 1.74 (95%CI=1.18-2.59, p<0.01) times more ICU episodes compared to patients with GPA. Patients with EGPA on the other hand experienced significantly less stays on ICU, with an IRR of 0.3 (95%CI=0.12-0.73, p<0.01).

In comparison to patients from the NHS Grampian, patients from the NHS Lothian had significantly increased episodes on ICU, with an IRR of 1.63 (95%CI=1.08-2.46, p=0.02), and also stay there longer, even if this result was non-significant (IRR=1.17; 95%CI=0.71-1.94, p=0.55).

Table 5 Poisson regression analyses exploring ICU episodes (2000-2017)

			Univariate analyses		Multivariate analysis	
HCU Parameter	Predictor (Base)	Subgroup	IRR (95% CI)	p-value	IRR (95% CI)	p-value
Hospital admission (ICU) (n=494)	Age		1.01 (1-1.02)	0.05	-	-
	Sex (Male)	Female	0.97 (0.67-1.42)	0.9	-	-
	AAV Type (GPA)	MPA	1.74 (1.18-2.59)	<0.01	-	-
		EGPA	0.3 (0.12-0.73)	<0.01	-	-
	ANCA Status (PR3)	MP0	1.37 (0.91-2.05)	0.13	1.38 (0.92-2.06)	0.12
		ANCA negative	0.59 (0.31-1.14)	0.12	0.59 (0.31-1.14)	0.12
	SES Quintiles (Q1)	Q2	0.84 (0.4-1.76)	0.65	-	-
		Q3	0.63 (0.32-1.27)	0.2	-	-
		Q4	0.63 (0.33-1.23)	0.17	-	-
		Q5	0.77 (0.4-1.5)	0.44	-	-
	Health board (Grampian)	Lothian	1.63 (1.08-2.46)	0.02	-	-
		Glasgow and Clyde	1.02 (0.4-2.59)	0.97	-	-
		Tayside	0.67 (0.29-1.6)	0.37	-	-
		Highland	1.4 (0.74-2.67)	0.3	-	-
		Fife	1.21 (0.53-2.76)	0.66	-	-

(ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; ICU=Intensive Care Unit; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; PR3=Proteinase 3)

Patients from the second most affluent socio-economic quintile stayed significantly less long on an ICU, compared to the most deprived patients. None of the other quintiles showed statistically significant results and socio-economic status does also not seem to be associated with the number of ICU episodes.

None of the other predictors showed significant associations with the length of stay on ICU, indicating that predictors other than the ones chosen in these analyses might be worth exploring.

Table 6 Poisson regression analyses exploring length of stay on ICU (2000-2017)

			Univariate analyses		Multivariate analysis	
HCU Parameter	Predictor (Base)	Subgroup	IRR (95% CI)	p-value	IRR (95% CI)	p-value
LOS (ICU) (n=494)	Age		1.01 (1-1.02)	0.46	-	-
	Sex (Male)	Female	1.19 (0.72-1.95)	0.5	-	-
	AAV Type (GPA)	MPA	1.52 (0.92-2.5)	0.11	-	-
		EGPA	0.14 (0.06-0.37)	<0.01	-	-
	ANCA Status (PR3)	MP0	1.12 (0.66-1.88)	0.68	1.21 (0.67-1.89)	0.67
		ANCA negative	0.36 (0.17-0.76)	<0.01	0.36 (0.17-0.75)	<0.01
	SES Quintiles (Q1)	Q2	0.62 (0.22-1.75)	0.37	-	-
		Q3	0.51 (0.2-1.28)	0.15	-	-
		Q4	0.39 (0.16-0.95)	0.04	-	-
		Q5	0.64 (0.27-1.55)	0.33	-	-
	Health board (Grampian)	Lothian	1.17 (0.71-1.94)	0.55	-	-
		Glasgow and Clyde	1.16 (0.22-6.03)	0.86	-	-
		Tayside	0.91 (0.19-4.39)	0.9	-	-
		Highland	0.94 (0.43-2.05)	0.88	-	-
		Fife	1.38 (0.52-3.68)	0.52	-	-

(ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; PR3=Proteinase 3)

Outpatient encounters

Table 6 below shows the univariable and multivariable Poisson regression results on outpatient encounters. The univariable analysis revealed that patients from socio-economic quintile 4 had significantly less outpatient encounters than the most deprived quintile (IRR=0.78; 95%CI=0.64-0.96; p=0.02), just alike the analyses on the number of inpatient hospital admissions to a general ward and the length of stay on ICU. It further showed that patients from the NHS Tayside had significantly less outpatient encounters compared to patients from the NHS Grampian (IRR=0.46; 95%CI=0.35-0.59; p<0.01).

The model resulting from the multivariable analysis encompassed 494 observations and included health board as the only predictor, showing the same result as the univariable, Tayside patients had less outpatient encounters than Grampian patients.

Table 7 Poisson regression analyses exploring outpatient encounters (2000-2017)

			Univariate analyses		Multivariate analysis	
HCU Parameter	Predictor (Base)	Subgroup	IRR (95% CI)	p-value	IRR (95% CI)	p-value
Outpatient encounters (n=494)	Age		1 (1-1.01)	0.77	-	-
	Sex (Male)	Female	1.09 (0.98-1.21)	0.12	-	-
	AAV Type (GPA)	MPA	1.08 (0.96-1.21)	0.2	-	-
		EGPA	0.95 (0.82-1.11)	0.52	-	-
	ANCA Status (PR3)	MP0	1.05 (0.94-1.18)	0.4	-	-
		ANCA negative	1.13 (0.96-1.32)	0.15	-	-
	SES Quintiles (Q1)	Q2	0.88 (0.71-1.08)	0.22	-	-
		Q3	0.81 (0.66-1)	0.05	-	-
		Q4	0.78 (0.64-0.96)	0.02	-	-
		Q5	0.95 (0.78-1.16)	0.62	-	-
	Health board (Grampian)	Lothian	0.99 (0.88-1.11)	0.81	0.99 (0.88-1.11)	0.87
		Glasgow and Clyde	1.27 (0.96-1.67)	0.09	1.27 (0.97-1.67)	0.09
		Tayside	0.46 (0.35-0.59)	<0.01	0.46 (0.36-0.59)	<0.01
		Highland	1.16 (0.95-1.43)	0.15	1.17 (0.95-1.43)	0.14
		Fife	1.15 (0.96-1.37)	0.12	1.17 (0.98-1.4)	0.09

(ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; PR3=Proteinase 3)

A&E

With regard to predictors of A&E incidents within the AAV cohort, the univariable analyses showed a significant link with age, AAV type, socio-economic status and NHS health board. Sex and ANCA status do not significantly predict A&E incidents. Age also seems to play a minor role, with an IRR of 1.01 (95%CI=1-1.02; p=0.02).

The social gradient that could earlier be discovered in the other healthcare utilisation parameters applies for A&E incidents as well, patients from higher socio-economic quintiles have significantly less A&E compared to the most deprived (Q3: IRR=0.6; 95%CI=0.38-0.97; p=0.04; Q4: IRR=0.57; 95%CI=0.36-0.91; p=0.02).

The highest IRR of all conducted Poisson regression analyses were obtained by exploring the NHS health board as a predictor for A&E incidents in both, the uni- and the multivariable analyses.

The model of A&E included age and health board as predictors and 464 observations. Accordingly, patients from the NHS Greater Glasgow and Clyde experienced A&E 4.5 (95%CI=2.78-7.27; p<0.01) times more often than patients from the NHS Grampian.

Table 8 Poisson regression exploring A&E incidents (2007-2017)

			Univariate analyses		Multivariate analysis	
HCU Parameter	Predictor (Base)	Subgroup	IRR (95% CI)	p-value	IRR (95% CI)	p-value
A&E (n=461)	Age		1.01 (1-1.02)	0.02	1.01 (1.01-1.02)	<0.01
	Sex (Male)	Female	1.11 (0.86-1.44)	0.41	-	-
	AAV Type (GPA)	MPA	1.44 (1.08-1.92)	0.01	-	-
		EGPA	1 (0.64-1.57)	1	-	-
	ANCA Status (PR3)	MP0	1.26 (0.95-1.67)	0.11	-	-
		ANCA negative	0.99 (0.63-1.57)	0.98	-	-
	SES Quintiles (Q1)	Q2	0.91 (0.57-1.43)	0.68	-	-
		Q3	0.6 (0.38-0.97)	0.04	-	-
		Q4	0.57 (0.36-0.91)	0.02	-	-
		Q5	0.65 (0.41-1.02)	0.06	-	-
	Health board (Grampian)	Lothian	3.03 (2.21-4.15)	<0.01	2.99 (2.19-4.07)	<0.01
		Glasgow and Clyde	3.94 (2.45-6.34)	<0.01	4.5 (2.78-7.27)	<0.01
		Tayside	1.49 (0.96-2.31)	0.08	1.44 (0.93-2.25)	0.11
		Highland	2.27 (1.45-3.58)	<0.01	2.35 (1.49-3.7)	<0.01
		Fife	2.42 (1.63-3.57)	<0.01	2.46 (1.64-3.68)	<0.01

(A&E=Accidents & Emergencies; ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; PR3=Proteinase 3)

3.2.4.2 Linear regression analyses

The purpose of the linear regression analyses was to explore influencing factors on the costs associated with healthcare utilisation, using the same prediction variables as in the Poisson regression analyses. Review of the assumptions showed not normally distributed residuals, when using the actual sample observations.

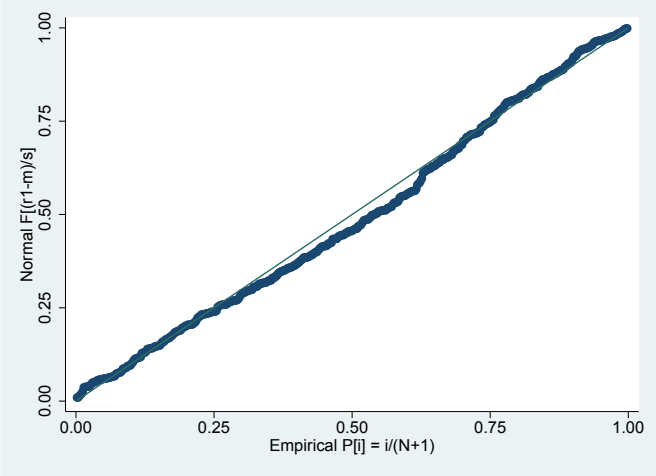


Figure 11 Q-Q-Plot on the distribution of residuals

Hence, log-transformation of the dependent variable was necessary. The resulting Quantile-Quantile-Plot below shows the distribution of residuals from the multiple linear regression model. The distribution partially deviated from normality, but was overall found to not wildly violate the normal distribution assumption.

Using the actual sample observations, the homoscedasticity assumption was likewise violated, as Breusch-Pagan/Cook-Weisberg test for heteroscedasticity showed p=0.0292. In order to adjust for this problem, robust standard errors were used.

Variance inflation factors (VIF) were determined for the included predictor variables to test for multicollinearity. Elevated VIFs were found for the predictors *AAV type* (VIF=4.39) and *ANCA status* (VIF=4.58). Also, there is clinical evidence suggesting collinearity between the AAV and ANCA status (Kobayashi & Fujimoto, 2013; Sokolowska et al., 2014). This, and the general research trend towards ANCA antibodies instead of AAV types, were reasons for why the variable *AAV type* was excluded as a predictor from the multivariable analysis (Jennette & Nachman, 2017).

Lastly, linearity between dependent variable and predictors was confirmed using scatterplots with fitted line[1]. As mentioned earlier, scatterplots could not be released due to the depiction of raw data, which could potentially be identifiable.

[1] The ISD Scotland could not release the resulting scatter plots for public use, due to patient confidentiality concerns. *See chapter 3.1.2 Data for an explanation.*

Table 9 below shows the untransformed beta coefficients resulting from the univariable and the multivariable linear regression analyses. Because of the log-transformation of the independent variable *costs post diagnosis*, the transformed exp(beta) coefficients are displayed on the right-hand side as well. All regression tables from the Stata output of both, the Poisson and the linear regression analyses, can be found in the appendices.

The final model included 494 observations and produced R-square=0.045 (F(11,482)=2.14; p=0.0163). Included predictors were status of ANCA, socio-economic status quintiles and NHS health board.

Table 9 Univariable and multivariable linear regression models exploring influences on costs

		Univariate Analysis		Final Model (n=494)		Univariate Analysis		Final Model (n=494)	
Predictor (Base)	Subgroup	Beta (95% CI)	p-value	Beta (95% CI)	p-value	exp(Beta) (95% CI)	p-value	exp(Beta) (95% CI)	p-value
Age		0.01 (0-0.01)	0.06	-	-	1.01 (1-1.01)	0.06	-	-
Sex (Male)	Female	0.07 (-0.12-0.26)	0.48	-	-	1.07 (0.88-1.3)	0.48	-	-
AAV Type (GPA)	MPA	0.04 (-0.17-0.26)	0.69	-	-	1.04 (0.84-1.3)	0.69	-	-
	EGPA	-0.47 (-0.77-(-0.17))	<0.01	-	-	0.62 (0.46-0.84)	<0.01	-	-
ANCA Status (PR3)	MP0	-0.15 (-0.36-0.06)	0.17	-0.18 (-0.4-0.03)	0.09	0.86 (0.7-1.07)	0.17	0.83 (0.67-1.03)	0.09
	ANCA negative	-0.31 (-0.62-0.01)	0.06	-0.36 (-0.68-(-0.05))	0.03	0.74 (0.54-1.01)	0.06	0.69 (0.51-0.96)	0.03
SES Quintiles (Q5)	Q1	0.07 (-0.3-0.43)	0.71	0.21 (-0.17-0.6)	0.28	1.07 (0.74-1.54)	0.71	1.24 (0.84-1.82)	0.28
	Q2	-0.13 (-0.44-0.17)	0.4	-0.09 (-0.4-0.22)	0.56	0.88 (0.64-1.19)	0.48	0.91 (0.67-1.24)	0.56
	Q3	-0.09 (-0.39-0.2)	0.52	-0.1 (-0.4-0.19)	0.49	0.91 (0.68-1.22)	0.52	0.9 (0.67-1.21)	0.49
	Q4	-0.17 (-0.45-0.1)	0.22	-0.16 (-0.44-0.11)	0.24	0.84 (0.64-1.11)	0.22	0.85 (0.64-1.12)	0.24
NHS Healthboard	Lothian	-0.16 (-0.4-0.07)	0.16	-0.19 (-0.43-0.04)	0.11	0.85 (0.68-1.07)	0.16	0.82 (0.65-1.04)	0.11
(Grampian)	Glasgow and Clyde	-0.55 (-0.95-(-0.15))	<0.01	-0.59 (-1.02-(-0.16))	<0.01	0.58 (0.39-0.86)	<0.01	0.55 (0.36-0.85)	<0.01
	Tayside	-0.51 (-0.95-(-0.08))	0.02	-0.58 (-1.04-(-0.12))	0.01	0.6 (0.39-0.92)	0.02	0.56 (0.35-0.88)	0.01
	Highland	-0.27 (-0.64-0.1)	0.15	-0.26 (-0.58-0.05)	0.1	0.76 (0.53-1.1)	0.15	0.77 (0.56-1.05)	0.1
	Fife	-0.3 (-0.66-0.05)	0.1	-0.41 (-0.84-0.03)	0.07	0.74 (0.52-1.06)	0.1	0.67 (0.43-1.03)	0.07

(ANCA=Antineutrophil Cytoplasmic Antibody; CI=Confidence Interval; EGPA=Eosinophilic Granulomatosis with Polyangiitis; exp(Beta)=exponential(Beta); GPA=Granulomatosis with Polyangiitis; MPA=Microscopic Polyangiitis; MPO=Myeloperoxidase; NHS=National Health Service; PR3=Proteinase 3)

Accordingly, patients with negative ANCA status showed statistically significantly lower costs compared to patients with PR3 (exp(Beta)=0.69; 95%CI=0.51-0.96; p=0.03). Patients with MPO also had lower costs, however, the results were not statistically significant.

The social gradient, which was discovered in the Poisson regression analyses, could partly be detected also in the linear regression. This was shown by the socio-economic quintile 1, stating that patients with the lowest socio-economic status have 24% higher costs (exp(Beta)=1.24; 95%CI=0.84-1.82; p=0.28) compared to the most affluent. Nonetheless, none of the results regarding socio-economic status were statistically significant and none of the other quintiles support the earlier discovered tendency.

Looking at the differences in costs due to the geographic location of the patients, all NHS health boards showed lower costs, compared to patients from the NHS Grampian. These results were significant for the NHS Greater Glasgow and Clyde (exp(Beta)=0.55; 95%CI=0.36-0.85; p<0.01), and NHS Tayside (exp(Beta)=0.56; 95%CI=0.35-0.88; p=0.01).

The univariable analyses further revealed significantly lower costs in patients with EGPA compared to GPA (exp(Beta)=0.62; 95%CI=0.46-0.84; p<0.01). MPA patients had slightly higher costs, but without statistical significance (exp(Beta)=1.04; 95%CI=0.84-1.3; p=0.69).

The results for age and sex were likewise not statistically significant, but generally compliant with the earlier results, as age exp(Beta) was 1.01 (95%CI=; p) and women showed slightly higher costs than men (exp(Beta)=1.07; 95%CI=0.88-0.13; p=0.48).

4 Discussion

4.1 Summary of findings

The overarching research aim of this thesis was to explore healthcare utilisation in Scottish patients with AAV. Supportive of this aim was the objective to identify ways of measuring healthcare utilisation and to find out about available evidence on healthcare utilisation in adult patients with vasculitis, specifically focussing on AAV. These results were then used to define healthcare utilisation, with the objective to assess healthcare utilisation in a well-defined AAV cohort in Scotland, UK. Lastly, it was of interest to explore predicting factors for increased healthcare utilization in AAV.

The following presents a detailed summary of findings meeting these aims.

Systematic literature review

The systematic literature review included a total number of 15 sources, dealing with ways of measuring and defining healthcare utilisation in patients with systemic vasculitis and AAV, respectively. Accordingly, the most common healthcare utilisation definitions included inpatient hospitalisation and costs. The NOS quality assessment indicated sources of rather poor quality, due to lacking comparability and the large proportion of grey literature. Further, none of the sources were found to deal with UK patients.

For the conducted linked-data study, the definition of healthcare utilisation following the literature review was complemented by the number of outpatient encounters as well as A&E incidents. Regarding inpatient hospitalisation, it was further differentiated between general wards and ICU.

Baseline characteristics and healthcare utilisation rates

The study was based on historic, administrative, non-identifiable multicentre data encompassing 543 patients with AAV and up to five matches per case (n=2671) from the general population. The median follow-up was approximately 5 years in both cohorts. GPA and PR3 were the most prevalent types of AAV and ANCA represented.

Compared to the general population cohort, patients with AAV showed substantially higher healthcare utilisation across all included parameters. This exceedance was statistically significant for all discrete time analysis periods, with up to 10 years of follow-up.

The first 9 months after diagnosis were especially pronounced, with rate ratios of up to 10.01 (95%CI=9.57-10.48), as seen in outpatient encounters. Over time, the rate ratios decreased. Nonetheless, the combined rate ratios for the last 5.25 years (4.25 for A&E, respectively) showed statistically significantly higher healthcare utilisation in patients with AAV across all parameters, even if single intervals were non-significant.

Description of costs

Translated into costs per person-year, healthcare utilisation amounted to £8,317.91 in patients with AAV post diagnosis, as opposed to £1,994.07 in the general population cohort. The resulting incremental costs of £6,323.84 (95%CI=£1,727.82-£10,919.87) demonstrate that patients with AAV were 4.17 times more expensive than their matches from the general population. A large proportion (89%) of these extra costs was due to inpatient hospitalisation.

Regression analyses

The Poisson regression analyses did not yield an all-encompassing result on driving factors of increased healthcare utilisation parameters, as assumptions were partially not met. Nonetheless, results may serve as a cautious initial assessment. Age was not a predictor of major importance, as the IRR revolved around 1 in all cases. The same holds for sex, women had slightly higher IRR across parameters compared to men, however, this was not significant.

Patients with MPA showed significantly higher healthcare utilisation compared to patients with GPA with regard to A&E incidents and ICU episodes. On the other hand, patients with EGPA showed significantly lower healthcare utilisation, which was significant for all inpatient hospital parameters.

The same tendency was seen in the type of ANCA status, patients with MPO showed higher healthcare utilisation in most of the parameters, although not statistically significant. Patients with ANCA negative status additionally showed lower healthcare utilisation compared to PR3 in most of the parameters, significantly, however, only for the length of stay on ICU.

A social gradient was discovered across all healthcare utilisation parameters, indicating that the most affluent patients with AAV showed the lowest healthcare utilisation. This result was, however, significant in most parameters only for quintile 4, the largest group (n=146).

The tendency of the social gradient was visible also in the linear regression, however, none of the results were statistically significant.

The linear regression results further were compliant with the Poisson regression results regarding age, sex, and also AAV type, but contradicted the results on ANCA status. Accordingly, both, patients with MPO and negative ANCA status showed lower costs compared to patients with PR3, even if not statistically significant.

Regarding NHS health boards, patients from the NHS Greater Glasgow and Clyde as well as Tayside showed only half the costs of patients from NHS Grampian, which seems to partially contradict the results from the Poisson analysis as well. There, Glasgow patients had significantly more A&E attendances, (IRR=3.94;95%CI=2.45-6.34; $p<0.01$).

These patients also showed more outpatient encounters, inpatient hospital admissions, ICU episodes and longer stays on ICU. However, these results were not statistically significant. Likewise, patients from NHS Lothian, Highland and Fife non-significantly showed lower costs compared to NHS Grampian patients.

Key findings

- Over many years of follow-up, patients with AAV showed significantly higher healthcare utilisation rate ratios across all included parameters, compared to the general population
- Patients with AAV were 4.17 times more expensive than the matched cohort, with absolute costs per person-year of £8,317.91 and incremental costs of £6,323.84
- Scottish women with AAV show significantly higher healthcare utilisation, age, conversely, does not seem to be of major importance in this regard
- Patients with MPA show significantly higher healthcare utilisation compared to GPA
- Poisson and linear regression analyses were not entirely conclusive, further research is warranted to clarify predictors for increased healthcare utilisation in Scottish patients with AAV

4.2 Discussion of findings

4.2.1 Methodology

The present examination of healthcare utilisation in patients with ANCA-associated vasculitis was based on historic, multicentre, non-identifiable data records, which were linked by the ISD Scotland and made accessible via the national Safe Haven. The use of electronic administrative health-data registries is generally recognised as an appropriate means of predicting costs in patients with varying predominant diseases (e.g. Asaria et al., 2016; Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014; Thorn et al., 2016).

Due to the relatively high number of included patients with AAV (n=543) as well as their geographic distributed across the country, the study participants were considered representative for the total AAV cohort in Scotland, UK.

The prevalence of the different subgroups of AAV and the respective types of ANCA furthermore matched epidemiological evidence (Houben et al., 2016; McKinney, Willcocks, Broecker, & Smith, 2014; Watts & Dharmapalaiah, 2012; Watts, Mooney, Skinner, Scott, & Macgregor, 2012).

Moreover, the results from the Poisson regression analyses demonstrated the same directions for the pairs of AAV type and ANCA status (MPA and MPO as well as GPA and PR3, respectively), which were previously found to be related (Jennette & Nachman, 2017; Kobayashi & Fujimoto, 2013).

4.2.2 Comparison with other studies

One of the main findings of the systematic literature review was the scarcity of evidence towards healthcare utilisation in patients with AAV. The identified sources correspondingly built the base for putting the results of the conducted linked-data study in context of current evidence in the field.

The graphs depicting the rate ratios on included healthcare utilisation parameters altogether showed very similar courses. The first nine months after diagnosis were in all cases particularly pronounced, indicating that the patients are sickest, when receiving their diagnosis.

This assumption is compliant with the literature, stating that the diagnosis in AAV patients often occurs late, when the disease has already progressed and hospitalisation is inevitable (Houben et al., 2016; McKinney et al., 2014). This draws back to the circumstance that the healthcare utilisation parameters included in this examination are mirroring the health burden of the patients.

After inception with very high healthcare utilisation, the course followed with a steep decrease in the rate ratios. This indicates the controlling of the disease with respective inpatient treatment. Nonetheless, patients with AAV showed higher healthcare utilisation compared to the general population, which was sustainable over many years of follow-up.

Translated into costs, this amounted to £8,317.91 per AAV-patient-year following diagnosis. Raimundo et al. found all-cause costs in US-American GPA patients amounting to 41,400USD (approximately £31,240[2]) per patient during 12-months following diagnosis. Patients with MPA further showed healthcare utilisation costs of 56,643USD (approximately £42,990) (Raimundo, Farr, Kim, & Duna, 2015).

McCormick et al. researched Canadian patients with systemic autoimmune rheumatic diseases (SARDS) (Lupus Erythematosus, systemic sclerosis, Sjörgen's disease, poly/dermatomyositis and systemic vasculitis) and found annual costs per patient of 8,901CAD (approximately £5,206) in the first year of follow-up. Differences may be explained by a broader definition of healthcare utilisation, as both studies included for example costs for prescriptions, which amounted to 2,909USD per patient in the American example.

Raimundo et al. further calculated the incremental costs between the period before and after diagnosis. Accordingly, costs increased by a factor of 1.88 in MPA patients, whereas costs were 3.11 times higher post diagnosis in the present study (Raimundo, Farr, Kim, & Duna, 2015). Thorpe et al. furthermore found annual healthcare expenditure in patients with systemic vasculitis to be twice as high as that of patients without systemic vasculitis (Thorpe et al., 2008). However, their definition also included many other diseases, such as Giant Cell Arteritis, Takayasu's disease, Polyarteritis Nodosa, Behcet's disease, and others.

[2] based on October 2018 exchange rate, taken from OANDA Currency Converter (OANDA, 2018)

Additionally, the US American and Canadian healthcare systems differ strongly from NHS Scotland. Having said that, the US health system is characterised by a large private sector, whereas NHS Scotland is almost an entirely public system. Also, the health expenditure in the US system are generally higher, compared to NHS Scotland (Schütte, Acevedo, & Flahault, 2018).

The heterogeneity of ways to measure healthcare utilisation is complicating the comparability across studies. This holds, for example, when count data is captured categorically, as seen in the survey data of Reinhold-Keller et al. (Reinhold-Keller et al., 2002). Instead of annual costs per patient, Cotch et al. further measured costs per admission, which were 12,023USD (approximately £9,125) in patients with GPA (Cotch et al., 1996).

As indicated in the summary of findings, the results of the Poisson and linear regression analyses were not consistent. Most striking differences were demonstrated looking at the type of ANCA as well as the NHS health boards. Defragmentation of which of the two conducted regression models shows more credibility appears to not be expedient, besides impossible. Both of the models showed methodological difficulties. The Poisson regression model showed insufficient goodness of fit, whereas the residuals of the linear regression model derived from normal distribution, despite log-transformation of the dependent variable.

Additionally, the results from the linear regression contradicted the expected outcomes based on clinical evidence. In a cohort study with US-American patients with AAV, Jennette et al. showed a significantly higher need for chronic dialysis or transplantation in patients with MPO-ANCA ($p<0.01$). Their data further demonstrated significantly more deaths in the MPO-cohort compared to PR3 ($p=0.03$) (Jennette & Nachman, 2017).

Worse renal survival in patients with MPO-ANCA compared to PR3-ANCA was also found in other studies, indicating that MPO-ANCA are linked to a much more acute clinical presentation of AAV patients, compared to PR3 (de Joode, Sanders, & Stegeman, 2013; Quintana et al., 2014).

End-stage renal disease with need for dialysis and or transplantations was furthermore estimated to cause healthcare costs of £23,426 per patient-year in an English cohort (including primary care and prescriptions) (Kerr, Bray, Medcalf, O'Donoghue, & Matthews, 2012). Because of the

high renal involvement associated with MPO-ANCA, the healthcare utilisation of these patients was presumed higher compared to that of PR3-ANCA patients. This result, however, was contradicted by the linear regression model.

On the other hand, patients with PR3 are known to experience relapse more often than MPO-patients (de Joode et al., 2013; Lionaki et al., 2012). Raimundo et al. found 2.6 times higher costs associated with relapse in the first year of follow-up in US-American GPA patients (88,065 vs. 30,682 USD; $p<0.0001$) (Raimundo, Farr, Kim, & Duna, 2015).

While there is a lot of discussion towards the predictive value of ANCA serotypes on clinical outcomes in patients with AAV (J. Charles Jennette & Nachman, 2017; Lionaki et al., 2012; McKinney et al., 2014), there is only little evidence for their influence on health economic outcomes. Consequently, the cost-intensity of ANCA serotypes and AAV types, respectively as well as the influence of the other included predictors of healthcare utilisation, remain inconclusive.

The results on healthcare utilisation rate ratios comparing AAV patients to the general population fit with the results of earlier studies. Irrespective of the detailed parameters included in the studies, patients with systemic vasculitis showed significantly higher healthcare utilisation compared to the general population.

Also, evidence supported the slope of the rate ratios to show especially pronounced healthcare utilisation in the first year following AAV-diagnosis, which decreased over time in most cases (Raimundo, Farr, Kim, & Duna, 2015). Wallace et al., mark an exemption, as they found hospitalisation rates to increase by 24% (to 6.3 per million) in GPA patients, over 8 years of follow-up (Wallace et al., 2017).

Notwithstanding the general compliance with earlier research, none of the data used for the development of rate ratios comparing AAV patients with the general population in earlier studies were longitudinal.

The results described in the present thesis are therefore firstly demonstrating significantly higher healthcare utilisation in patients with AAV compared to the general population. This increase was sustainable over many years of follow-up.

4.2.3 AAV in the context of rheumatological diseases

In the context of other rheumatological diseases, the healthcare utilisation in patients with AAV as defined in this study might be comparable with the one found in cases of COPD with medium to high disease severity.

Investigating a large German sample of COPD patients (n=2741) and lung-healthy matches (n=1537), Wacker et al. found healthcare costs to be increased by a factor of 2.4 to 5.5 in patients with COPD of different disease severity. However, their definition of healthcare utilisation was very broad, and included for instance treatment costs (Wacker et al., 2016).

Cortaredona et al. examined the costs due to comorbidity in France. Accordingly, healthcare utilisation in Scottish patients with AAV is comparable to the costs per capita of French patients with chronic kidney disease, without comorbidity (8,323€, approximately £7,311) or with major depression (9,694€) or patients with alcohol use disorders and chronic kidney disease as comorbidity (9,344€). The costs per capita in this study refer to the "weighted average estimates of costs in 2014 for prevalent cases […] and incident cases […]".

Most strikingly, cancers, including breast, lung, colorectal, stomach, liver, kidney, pancreatic, and oesophageal cancer, showed lower costs per capita in 2014. This holds also for cancers with comorbidities including stroke, heart disease, and diabetes. Only cancers in combination with chronic kidney failure or with cirrhosis were more expensive than to the Scottish AAV cohort (Cortaredona & Ventelou, 2017).

4.3 Strengths and limitations

Healthcare utilisation in patients with systemic vasculitis in general is a field, which has not been extensively researched. The scarcity of studies focussing on patients with AAV in particular, compared to the general population, was firmly demonstrated in the present examination.

This holds true, especially with regard to the availability of studies from the UK. The presented results therefore are assessing healthcare utilisation in Scottish patients with ANCA-associated vasculitis for the first time, over a period of up to ten years of follow-up.

Great strengths of the study include the longitudinal multicentre data, comprising a major part of the total Scottish AAV cohort. Matching and linking were performed independently by the

ISD Scotland, including up to five general population matches per case. The resulting cohort is considered representative for the total AAV cohort in Scotland, due to the number of included patients, their epidemiological characteristics (e.g. types of AAV and ANCA), but also because of their geographic location across the country. The AAV diagnoses were furthermore not based on ICD-10 coding, but verified by specialists, minimising misclassification bias. The data was routinely collected and not based on claims.

Because of the powerful data, it is highly unlikely that the results simply occurred by chance.

The results are limited to the data bases included, which encompass secondary and tertiary care only. Patients, who were never treated in an inpatient setting, but solely in primary care, are omitted. Still, given the severe development of the diseases, it can be assumed that the large majority of cases is included in the cohort.

Patients were assigned to either GPA, MPA or EGPA, comprised under the umbrella term AAV. However, the Chapel Hill Consensus Conference Nomenclature of Vasculitides entitles *single-organ AAV*, for example renal-limited AAV, as one of the major clinicopathologic variants as well. This type of AAV is not covered in the data, but is also not mentioned as a separate disease (J. C. Jennette et al., 2012). Still, it might be worth exploiting the clinicopathologic variants of AAV with regard to differences in healthcare utilisation, as the patients needs are likely to differ.

Another important limitation is the definition of healthcare utilisation. In this example, the investigation of healthcare utilisation included the parameters inpatient hospitalisation, including the number of admissions and the referring length of stay for both, general wards and intensive care settings. It further comprised outpatient encounters and A&E incidents.

Nonetheless, there are many other ways of measuring healthcare utilisation, which might be of interest for further investigation. Those may include the number and type of prescriptions as well as inpatient medication, which is known to be cost-intensive in patients with AAV (Casian & Jayne, 2011; Tesar, 2015). Also, certain inpatient and outpatient procedures, which typically occur in patients with AAV, may be of interest for further evaluation.

4.4 Implications

4.4.1 Public Health Relevance

The results presented in this thesis show the substantial economic burden, which is linked to an AAV diagnosis. Despite the rarity of the disease, with a prevalence of approximately 255 per million population, these patients have a considerable monetary impact on NHS Scotland. In absolute terms, each patient with AAV costs £6,323.84 more for the health system than patients without AAV. Having said this, the total economic burden is still an underestimation, because it does not yet include the treatment costs. Hence, the absolute cost ratio can be presumed even higher.

On public health level, these findings are interesting for politicians as well as clinicians. The objective is to inform local decision makers about the financial impact of the disease. As highlighted by Watts et al., a quantification of the health burden in patients with AAV is also a means of proving to funders the need for the introduction of new drugs and better resource allocation (Watts, Robson, & Pearce, 2017).

In the following, clinical pathways and the general supply of healthcare for these patients, may need to be re-adjusted. The analyses showed that the major part of the extra costs entailing with an AAV diagnosis are due to inpatient hospitalisation on general wards as well as on ICU.

According to the hypothesis of AAV patients to be at the most severe stages of the disease, when getting their diagnosis, the reasonable approach seems to be an intervention that is preventive to deterioration. An early diagnosis system would likely be most beneficial in this regard. That way, patients could be treated before severe progression of the disease. This may avoid lengthy stays on ICU and general wards as well as reduce the numbers of outpatient encounters. This, in turn, may reduce the costs per patient-year associated with healthcare utilisation, in favour of the healthcare system.

4.4.2 Future Research

Future research is highly warranted to confirm the descriptive results from the present study. Also, a complementation of the definition of healthcare utilisation is warranted, to yield an all-encompassing assessment of healthcare utilisation in patients with AAV.

Other parameters of interest for future research include certain inpatient and outpatient procedures, which are typical needs of patients with AAV, are of special interest. An example for this would be dialysis, among others. Another interesting aspect worth researching in the future would be the estimation of the indirect costs caused by the loss of productivity and ability to work.

Again, the results have yet to include costs due to medication. Novel biologic drugs, such as Rituximab, are highly researched for their clinical value as induction therapy in patients with AAV (Bajema et al., 2017). Clinical effectiveness of such treatment regimes will inevitably entail a decrease in the healthcare utilisation parameters covered in this examination, as they mirror the health burden of the patients. From a health economic perspective, cost-effectiveness studies would need to ensue, in order to create a more comprehensive picture of healthcare utilisation and the total economic burden of AAV for NHS Scotland.

Additionally, the investigation of driving factors for increased healthcare utilisation should be re-examined. Solutions for the methodological issues might include negative binomial regression instead of Poisson regression and a general linear model approach instead of linear regression.

It is of high importance that future research projects in this field are conducted in diverse places, given the geographic variations of disease characteristics as well as prevalence of combinations regarding AAV types and types of ANCA. Only then, healthcare utilisation in patients with ANCA-associated vasculitis can thoroughly be assessed, allowing for better informed decision making regarding the need of new drugs as well as the targeting of driving factors.

5 Conclusion

The results found in this study complement the scarce body of evidence towards healthcare utilisation in patients with AAV, with the novelty of firstly providing extensive, longitudinal data from the UK.

Despite the lacking explanatory power of the regression analyses, the descriptive analyses comprehensibly quantified the economic burden of patients with AAV. It was demonstrated that these patients are characterised by significantly higher healthcare utilisation compared to the general population. As a result, they are 4.17 times more expensive than their matches. The costs per person-year amounted to £8,317.91. This means that each patient with AAV costs the NHS Scotland £6,323.84 more per person-year than patients from the general population.

The findings shall inform decision makers about the substantial financial impact despite the rarity of the disease, in order to promote better resource allocation. It is important to consider that the results are still an underestimation of the total economic burden.

Future research should focus on a complementary definition of healthcare utilisation, comprising the cost-intensive medication, AAV-typical procedures like dialysis as well as indirect costs.

6 Bibliography

Abegunde, D. O., Mathers, C. D., Adam, T., Ortegon, M., & Strong, K. (2007). The burden and costs of chronic diseases in low-income and middle-income countries. *Lancet*. https://doi.org/10.1016/S0140-6736(07)61696-1

Asaria, M., Walker, S., Palmer, S., Gale, C. P., Shah, A. D., Abrams, K. R., … Sculpher, M. (2016). Using electronic health records to predict costs and outcomes in stable coronary artery disease. *Heart*, *102*(10), 755–762. https://doi.org/10.1136/heartjnl-2015-308850

Bali, J. (2016). *Basics of Biostatistics: A Manual for Medical Practitioners*. New Dehli, India.

Barkun, A. N., Adam, V., Martel, M., & Bardou, M. (2013). Cost-effectiveness analysis: stress ulcer bleeding prophylaxis with proton pump inhibitors, H2 receptor antagonists. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *16*(1), 14–22. https://doi.org/10.1016/j.jval.2012.08.2213

Basu, N., McClean, A., Harper, L., Amft, E. N., Dhaun, N., Luqmani, R. A., … Macfarlane, G. J. (2014). The characterisation and determinants of quality of life in ANCA associated vasculitis. *Annals of the Rheumatic Diseases*, *73*(1), 207–211. https://doi.org/10.1136/annrheumdis-2012-202750

Bloudek, L. M., Stokes, M., Buse, D. C., Wilcox, T. K., Lipton, R. B., Goadsby, P. J., … Martelletti, P. (2012). Cost of healthcare for patients with migraine in five European countries: Results from the International Burden of Migraine Study (IBMS). *Journal of Headache and Pain*, *13*(5), 361–378. https://doi.org/10.1007/s10194-012-0460-7

Bonatti, F., Reina, M., Neri, T. M., & Martorana, D. (2014). Genetic susceptibility to ANCA-associated vasculitis: State of the art. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2014.00577

Booth, A. D., Almond, M. K., Burns, A., Ellis, P., Gaskin, G., Neild, G. H., … Jayne, D. R. W. (2003). Outcome of ANCA-associated renal vasculitis: a 5-year retrospective study. *American Journal of Kidney Diseases : The Official Journal of the National Kidney Foundation*, *41*(4), 776–784. https://doi.org/10.1016/S0272-6386(03)00025-8

Cantorna, M. T., & Mahon, B. D. (2004). Mounting evidence for vitamin D as an environmental factor affecting autoimmune disease prevalence. *Experimental Biology and Medicine (Maywood, N.J.)*, *229*(11), 1136–1142. https://doi.org/229/11/1136 [pii]

Casian, A., & Jayne, D. (2011). Plasma exchange in the treatment of Wegener's granulomatosis, microscopic polyangiitis, Churg-Strauss syndrome and renal limited vasculitis. *Current Opinion in Rheumatology*, *23*(1), 12–17. https://doi.org/10.1097/BOR.0b013e32834120c1

Curran-Everett, D. (2018). Explorations in statistics : the log transformation. *Advances in Physiology Education*, *42*, 343–347. https://doi.org/10.1152/advan.00018.2018

Dadonienė, J., Kumžaitė, G., Mačiulytė, R., & Miltinienė, D. (2017). Clinical characteristics and long-term survival differences of the ANCA-associated vasculitis group: a cross-sectional study of 27 patients. *Acta Medica Lituanica*, *24*(2), 107–112. https://doi.org/10.6001/actamedica.v24i2.3491

Dean, A., Sullivan, K., & Soe, M. (2013). OpenEpi: Open Source Epidemiologic Statistics for Public Health. Retrieved February 23, 2018, from www.OpenEpi.com, updated 2013/04/06

Dupont, W. D. (2009). *Statistical modeling for biomedical researchers: A simple introduction to the analysis of complex data, second edition*. *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*. https://doi.org/10.1017/CBO9780511575884

Fauci, A. S., Haynes, B. F., Katz, P., & Wolff, S. M. (1983). Wegener's granulomatosis: Prospective clinical and therapeutic experience with 85 patients for 21 years. *Annals of Internal Medicine*. https://doi.org/10.7326/0003-4819-98-1-76

Graphpad Software. (2018). GraphPad QuickCalcs. Retrieved February 23, 2018, from http://graphpad.com/quickcalcs/ttest1/?Format=C

Gujarati, D. N. (2018). Linear Regression: A Mathematical Introduction. *Volume 177 of Quantitative Applications in the Social Sciences*, 272.

Haight, F. A. (1967). Handbook of the Poisson distribution. *Publications in Operations Research*, *11*(11), 168.

Hernán, M. A., Hernández-díaz, S., Werler, M. M., & Mitcheil, A. A. (2002). Hernán et al., 2002_ Am J Epidemiol_Causal Knowledge as a Prerequisite for Confounding Evaluation.pdf, *155*(2), 176–184.

Houben, E., Bax, W. A., van Dam, B., Slieker, W. A. T., Verhave, G., Frerichs, F. C. P., … Penne, E. L. (2016). Diagnosing ANCA-associated vasculitis in ANCA positive patients. *Medicine*, *95*(40), e5096. https://doi.org/10.1097/MD.0000000000005096

Information Services Division Scotland. (2017). Costs Archive. Retrieved September 18, 2018, from http://www.isdscotland.org/Health-Topics/Finance/Costs/costs-archive.asp

Kirkwood, B. B., & Sterne, J. (2003). *Essential medical statistics*. *Malden, MA: Blackwell Science*. https://doi.org/10.1002/sim.1961

Kobayashi, S., & Fujimoto, S. (2013). Epidemiology of vasculitides: Differences between Japan, Europe and North America. *Clinical and Experimental Nephrology*. https://doi.org/10.1007/s10157-013-0813-9

Lopez-Bastida, J., Oliva, J., Antoñanzas, F., García-Altés, A., Gisbert, R., Mar, J., & Puig-Junoy, J. (2010). Spanish recommendations on economic evaluation of health technologies. *Eur J Health Econ*, *11*. https://doi.org/10.1007/s10198-010-0244-4

Manuel, D. G., Schultz, S. E., & Kopec, J. A. (2002). Measuring the health burden of chronic disease and injury using health adjusted life expectancy and the Health Utilities Index. *Journal of Epidemiology and Community Health*, *56*(11), 843–850. https://doi.org/10.1136/jech.56.11.843

McKinney, E. F., Willcocks, L. C., Broecker, V., & Smith, K. G. C. (2014). The immunopathology of ANCA-associated vasculitis. *Seminars in Immunopathology*. https://doi.org/10.1007/s00281-014-0436-6

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2014). Preferred Reporting Items for Systematic Reviews and Meta-Analyses : *Annals of Internal Medicine*, *151*(2), 264–269. https://doi.org/10.1371/journal.pmed1000097

Morgan, M. D., Szeto, M., Walsh, M., Jayne, D., Westman, K., Rasmussen, N., … Harper, L. (2017). Negative anti-neutrophil cytoplasm antibody at switch to maintenance therapy is associated with a reduced risk of relapse. *Arthritis Research and Therapy*, *19*(1), 1–7. https://doi.org/10.1186/s13075-017-1321-1

Nguyen, K.-H., & Gordon, L. G. (2015). Cost-Effectiveness of Repetitive Transcranial Magnetic Stimulation versus Antidepressant Therapy for Treatment-Resistant Depression. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *18*(5), 597–604. https://doi.org/10.1016/j.jval.2015.04.004

NHS Research Scotland. (2018). Data Safe Haven. Retrieved September 12, 2018, from https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens-2

Raimundo, K., Farr, A. M., Kim, G., & Duna, G. (2015). Clinical and economic burden of antineutrophil cytoplasmic antibody-associated vasculitis in the United States. *Journal of Rheumatology*, *42*(12), 2383–2391. https://doi.org/10.3899/jrheum.150479

Schönermarck, U., Gross, W. L., de Groot, K., Schonermarck, U., Gross, W. L., & de Groot, K. (2014). Treatment of ANCA-associated vasculitis. *Nature Reviews Nephrology*, *10*(1), 25–36. https://doi.org/10.1038/nrneph.2013.225

Scottish Government. (2018). Scottish regional resilience partnerships' framework for exotic notifiable animal diseases contingency plans. Retrieved October 9, 2018, from https://beta.gov.scot/publications/scottish-regional-resilience-partnerships-framework-exotic-notifiable-animal-diseases-contingency/pages/7/

Sokolowska, B., Szczeklik, W., Wludarczyk, A., Kuczia, P., Jakiela, B., Gasior, J., … Musial, J. (2014). ANCA-positive and ANCA-negative phenotypes of eosinophilic granulomatosis with polyangiitis (EGPA): outcome and long-term follow-up of 50 patients from a single Polish center. *Clinical and Experimental Rheumatology*, *32*(3, Suppl. 82), 41–47.

Spencer, E., Mahtani, K., Brassey, J., & Heneghan, C. (2018). Misclassification bias. Retrieved October 7, 2018, from https://catalogofbias.org/biases/misclassification-bias/

StataCorp. (2015). Stata Statistical Software: Release 14. *2015*. https://doi.org/10.2307/2234838

Tan, J. A., Dehghan, N., Chen, W., Xie, H., Esdaile, J. M., & Avina-Zubieta, J. A. (2017). Mortality in ANCA-associated vasculitis: a meta-analysis of observational studies. *Annals of the Rheumatic Diseases*, annrheumdis-2016-210942. https://doi.org/10.1136/annrheumdis-2016-210942

Tesar, V. (2015). Moderator's view: Should all patients with ANCA-associated vasculitis be primarily treated with rituximab? *Nephrology Dialysis Transplantation*. https://doi.org/10.1093/ndt/gfv215

Thorpe, C. T., Devellis, R. F., Blalock, S. J., Hogan, S. L., Lewis, M. A., & Devellis, B. M. (2008). Patient perceptions about illness self-management in ANCA-associated small vessel vasculitis. *Rheumatology*, *47*(6), 881–886. https://doi.org/10.1093/rheumatology/ken126

UCLA: Statistical Consulting Group. (2018a). Poisson Regression. Stata Analysis Examples. Retrieved September 25, 2018, from https://stats.idre.ucla.edu/stata/dae/poisson-regression/

UCLA: Statistical Consulting Group. (2018b). Regression with STATA Chapter 1 - Simple and Multiple Regression.

Urueña, A., Pippo, T., Betelu, M. S., Virgilio, F., Hernández, L., Giglio, N., … Vizzotti, C. (2015). Cost-effectiveness analysis of rotavirus vaccination in Argentina. *Vaccine*, *33*(S1), A126–A134. https://doi.org/10.1016/j.vaccine.2014.12.074

Wacker, M. E., Jörres, R. A., Schulz, H., Heinrich, J., Karrasch, S., Karch, A., … Holle, R. (2016). Direct and indirect costs of COPD and its comorbidities: Results from the German COSYCONET study. *Respiratory Medicine*, *111*, 39–46. https://doi.org/10.1016/j.rmed.2015.12.001

Walton, E. W. (1958). Giant-cell granuloma of the respiratory tract (wegener's granulomatosis). *British Medical Journal*, *2*(5091), 265–270. https://doi.org/10.1136/bmj.2.5091.265

Watts, R. A., & Dharmapalaiah, C. (2012). Topical Reviews. *Reports on the Rheumatic Diseases*, 1–12. https://doi.org/10.1161/STROKEAHA.107.493502

Watts, R. A., Mooney, J., Skinner, J., Scott, D. G. I., & Macgregor, A. J. (2012). The contrasting epidemiology of granulomatosis with polyangiitis (Wegener's) and microscopic polyangiitis. *Rheumatology*, *51*(5), 926–931. https://doi.org/10.1093/rheumatology/ker454

Wells GA O'connell D., B. S. (2011). The Newcastle-Scale for assessing the quality of nonrandomised studies in meta-analyses. Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Westerhout, K. Y., Verheggen, B. G., Schreder, C. H., & Augustin, M. (2012). Cost effectiveness analysis of immunotherapy in patients with grass pollen allergic rhinoconjunctivitis in Germany. *Journal of Medical Economics*, *15*(5), 906–17. https://doi.org/10.3111/13696998.2012.688904

Williams, R. (2015). Multicollinearity. *University of Notre Dame, Http://Www3.Nd.Edu/~rwilliam/*, *41*(1), 1–6. https://doi.org/10.1017/CBO9781107415324.004

Yates, M., & Watts, R. (2017). ANCA-associated vasculitis. *Clinical Medicine (London, England)*, *17*(1), 60–64. https://doi.org/10.7861/clinmedicine.17-1-60

APPENDICES

Content

Tables & Figures

APPENDIX I: Systematic Literature Review Supplement

Appendix Table 1 Search criteria – PICO Format

P (population)	**I** (intervention)	**C** (comparison)	**O** (outcome)
Adult patients with systemic and or ANCA associated vasculitis	None	All	healthcare utili*ation healthcare use healthcare expenditure* healthcare spending healthcare consumption healthcare demand healtcare need* cost-effectiveness economic analy*is health service* use cost* cost-effectiveness cost analysis resource consumption resource allocation resource use resource utili*ation hospitali*ation hospital admission* hospital readmission* hospital visit* clinician visit* physician visit* length of stay waiting time* prescription* Antineutrophil Cytoplasmic Antibody ANCA vasculiti*

Appendix Table 2 Summary of findings in the systematic literature review

Author	Title	Year	Journal	Original purpose	Design and Setting	Location	Period	Data source	Type of vasculitis	Population	Sample size	Outcome of interest	Result of interest	Limitations of interest
Cotch, M. F., Hoffman, G.S., Yerg, D.E. et al.	The Epidemiology of Wegener's Granulomatosis. Estimates of the Five-Year Period Prevalence, Annual Mortality, and Geographic Disease Distribution from Population-Based Data Sources	1996	Arthritis & Rheumatism	To estimate - prevalence, - annual mortality and - geographic distribution of WG	- Analysis of national vital statistics data and hospitalization data from a national survey - population-based - hospital-based	State of New York, USA	1986-1990	- national mortality data - hospitalization data from a national sample of hospitals (National Hospital Discharge Survey (NHDS) and Statewide Planning and Research Cooperative System (SPARCS) - hospitalization data from all nonfederal, nonpsychiatric hospitals	GPA	- US-Patients with Wegener's Granulomatosis (n=978) - majority of white descent (76%) - 'rather equal representation of males and females' - mean age 41 (NY) - mean age 56 (nationwide)	978	- LOS -hospitalizations	- LOS: 12 (nationwide) - LOS: 17 (NY) - 1.7 hospitalizations for Wegener's granulomatosis over 5-year period	- diagnosis is code-based (ICD-9-CM), code (ICD 446.4) (7), potentia misclassification bias - potential for underestimation of death records related to WG - hospitalization rate was not measured, but taken from SPARCS, w routinely collects discharge and uniform billing data from all hospitalizations within the state) - focus was on epidemiological measures, not on healthcare utilisat - focus on NY, not nation - unhospitalized patients are not captured, assumption that diagno occurs in hospital OR patients are hospitalized at least once - majority of ethnic groups was white (76%) - no inclusion of people without access to hospital care
Foocharoen, C., Thavornpitak, Y., Mahakkanukrauh, A.	Admission rate and charateristics of hospitalized systemic connective tissue disorders: analysis from a nationwide Thailand healthcare database	2013	International Journal of Rheumatic Diseases	- Clarify admission rate, - disease determination, - hospital mortality rate, - LOS and - hospital charges among hospitalized patients with SCNTD	- Cross-sectional analysis - population-based (nationwide)	Thailand	2010	- 2010 national database of hospitalized patients from the Thai health Coding Center (Ministry of PH), covering four different health care providers - covering four different healthcare providers and self-payment	- Behcet's disease - small vessel vasculitis - medium vessel vasculitis - large vessel vasculitis	- Thais > 18 yrs with SCNTD (n=6861 admissions) - majority (69.5%) was working age (no definition)	6861	- admission rate - LOS - hospital charges	- total admission rate for SCNTDs was 141/100,000 - SVV: 70 admissions in 2010 - mean LOS was 14.5, - mean hospital charge: 2912.83 USD - MVV: 38 admissions in 2010, - mean LOS was 10.4, - mean hospital charge: 2586.47 USD - LVV: 118 admissions in 2010, - mean LOS was 6.7, - mean hospital charge: 1794.9 USD - Behcets: 67 admissions in 2010, - mean LOS was 7.0, - mean hospital charge: 1308.6 USD	- primary diagnosis performed by unknown physician, who could n contacted - diagnosis is code-based (ICD-10), potential for miscoding - only patients >18 with primary diagnosis related to SCNTD were included - also patients who have not been treated in a hospital setting, wer included - data shows number of admissions only, not caring about readmiss of patients - Thai population is not representative for Europeans - "patients are only treated supportively at community hospitals" - "limited knowledge of rare diseases among primary care doctors" - "inadequate reimbursement policy of healthcare schemes"
Janisiewicz, A.M., Klau, M.H., Keschner, D.B. et al.	Higher antineutrophil cytoplasmic antibody (C-ANCA) titers are associated with increased overall healthcare use in patients with sinonasal manifestations of granulomatosis with polyangiitis (GPA)	2015	Am J Rhinol Allergy	Determine impact of C-ANCA levels on radiographic findings and healthcare use in patients ith sinonasal GPA	- Retrospective review - Single center (multidisciplinary rheumatologic/otolaryngologic clinic - cohort study	California, USA	2008-2013	- data collection via retrospective chart review	GPA with sinonasal manifestation	- (n=44) US-patients with GPA with sinonasal manifestation - average age was 52.5 (range 22-86 yrs) 25% men, 75% women	44	- number of CT scans - healthcare use, defined as - rheumatology and otolaryngology office visits - allied health nurse visits - patient emails - patient telephone calls	- patients with C-ANCA titers more than or equal to 1:80 demonstrated a significantly greater overal healthcare use than their counterparts, with a mean of 121 and 69 encounters, (p=0.03) - other measures were not significant	- focus on healthcare interactions within the rheumatology and otolaryngology departments - excluded primary care, nephrology, pumonary and other subspeci visits - strange definition of overall healthcare use (e-mails weigh the san clinic visits?)
Krulichova, I., Gamba, S., Ricci, E. et al.	Direct medical costs of monitoring and treating patients with Takayasu arteritis in Italy	2004	Eur J health Econ	Estimation of resource consumption and direct medical costs for Takayasu arteritis (TA)	- Multicenter, - prospective study - hospital-based, (12 medical departments)	Italy (mostly northern italy)	1998-2000	- hospital discharge database of the Italian Department of Health - the Italian TA registry - Database of the Clinical Research Center for Rare Diseases Aldo and Cele Dacco	Takayasu Arteritis (TA)	- Italian patients with TA (n=67, 45 of which active and 26 of which inactive) - met the American College of Rheumatology 1990 criteria for TA - total mean age was 37.4 - 94% female ("fits epi data") - 38.8% employed, 41.8% unemployed, 19.4% retired	67	- mean resource consumption per patient year - average cost per patient - cost driving factors - hospital admissions	Costs: - Total average cost per patient year was 4,079.3EUR (95%CI 3,131.8-5,333.7; p=0.0093) - Hospital admissions: 44.8% (1,829.3 EUR) of total (95%Ci 1,117.6-2,828.6; p=0.013) - Drugs: 22% (895.7EUR) of total (95%CI 695.9-1,204.4; p=0.0011) - Laboratory tests: 6% (236.6EUR) of total (95%CI 201-290.6; p=0.0001) - GP consultations: 3% (135.4EUR) of total (95%CI 111.9-166.9; p=0.005) - Day-hospital days, instrumental examinations and specialist consultations were not statistically significant Resource consumption: - GP consultations: mean=12.8 (95%CI 10.6-15.7; p=0.005) - Laboratory tests: mean=12.9 (95%CI 11-15.5; p=0.0001) - Hospital admissions mean=0.5 (95%CI 0.3-0.7; p=0.012) - Specialist consultations, Instrumental examinations and Day-hospital days were not significant	- study duration was only one year - small sample - 21% loss to follow-up - comparison of resource consumption in patients with active and inactive TA, not with the general population - participating centers were not randomly selected, majority was lo in northern Italy and is therefore not representative for the whole country - no percentages in the tables, but means - discharge data -> must be coded diagnosis! Not stated in the pape
Michet, C.J., Achenbach, S.J., Crowson, C.S. et al.	Hospitalization rates and utilization among patients with giant cell arteritis: A population-based study from 1987-2012	2015	Seminars in Arthritis and Rheumatism	To understand whether patients with GCA are at greater risk for all-cause hospitalization when compared to the general population	- Retrospective, - population-based - cohort study of GCA patients, comparing GCA and non-GCA	Olmsted County, Minnesota, USA	1987-2012	-Medical records linkage system, containing the complete inpatient and outpatient medical records from all healthcare providers in the area - billing data	GCA	- Patients (n=199, 194 controls) diagnosed with GCA btw 1/1/1950 and 31/12/2009 - met the American College of Rheumatology 1990 criteria for GCA - 79% female - mean age 76.2 - mean follow-up was 8.2	199	- hospitalizations (admission dates, discharge dates, admission and discharge diagnoses) per 100 person-yrs from billing data - LOS	- average LOS was 6 days in both groups (median 4 days; **p=0.64**) - overall hospitalization rate was 49.7/100PY with a rate ratio of 1.13 (**95%CI 1.02-1.25**) compared to non-GCA - similar readmission rates btw the groups with 144 (22% of 65) (GCA) and 147 (25% of 578)(non-GCA) readmissions, **p=0.17**	- 79.9% of patients were females - mean age was quite old with 76.2 9 (controls similar) - follow-up for non-GCA group was slightly longer - most of the results are not statistically significant - diagnosis is code-based, potential for miscoding

Author	Title	Year	Journal	Original purpose	Design and Setting	Location	Period	Data source	Type of vasculitis	Population	Sample size	Outcome of interest	Result of interest	Limitations of interest
Raimundo, K., Farr, A.M., Kim, G. et al.	Clinical and Economic Burden of Antineutrophil Cytoplasmic Antibody-associated Vasculitis in the United States	2015	The Journal of Rheumatology	- Estimate the prevalence of relapse, healthcare use, and costs associated with GPA - develop an algorithm to find MPA patients in admin. claims databases - quantify the clinical and economic burden associated with MPA	- Longitudinal - retrospective - observational - cohort study - population-based	US	2009-2013	- De-identified caims data (Truven Health MarketScan Commercial and Medicare Supplemental Databases) - contain integrated patient-level pharmacy and medical (inpatient and outpatient) claims - variety of fee-for-service, fully capitated, and partially capitated health schemes - ICD-9-CM diagnosis - procedure codes - Current procedure - Coding System codes - National Drug Codes - laboratory tests only for a subset of patients	- GPA - MPA	- US patients with GPA (n=2784, 309 with continuous enrollment) - US patients with MPA (n=612) Total pop: - 79% above 45 - 54% females	2784 (GPA) 612 (MPA)	- Healthcare use (presence and number of inpatient admissions, ER visits, outpatient use) - pharmacy prescriptions - costs (all-cause & GPA-related for GPA; before & after diagnosis for MPA)	- mean total all cause and GPA-related costs were 41,400USD and 24,319 USD during **12-month** follow-up - mean total all cause and GPA-related costs were 76,798USD and 44,740 USD during **24-month** follow-up - relapsed patients incurred higher total all-cause and GPA-related costs than those without relapse (all-cause 88,065 vs. 30,682 USD, p<0.0001; GPA-related 61,636 vs. 15,748 USD, p<0.0001) during **12-month** folow-up - relapsed patients incurred higher total all-cause and GPA-related costs than those without relapse (all-cause 136,007 vs. 51,526 USD, p<0.0001; GPA-related 89,761 vs. 25,531 USD, p<0.0001) during **24-month** folow-up - average all-cause costs after MPA diagnosis were 56,642 USD, compared to 30,166 USD before diagnosis, incremental cost difference of 24,476USD (95%CI 17,678-35,274USD; p<0.0001) - costs related to inpatient admissions being the main driver before and after diagnosis: 15,344 before vs. 34,776 USD after, incremental cost difference of 19,432USD (95%CI 11,586-27,277USD; p<0.0001) - costs related to prescription drugs were 2599 USD before diagnosis and 2909 USD after diagnosis with p<0.0001 - outpatient office visits accounted for 832 USD before and 1213 USD after diagnosis - other outpatient services accounted for 10498 USD before and 16725 USD after diagnosis - ER visits accounted for 893 USD before and 1019 USD after diagnosis - 0.39 inpatient admissions per case before diagnosis - 0.49 inpatient admissions per case after diagnosis	- limited to individuals with commercial coverage or Medcare supplemental cov - potential of mislassification of GPA an diagnosis (no ICD-9 code for MPA) - ICD-9-code 447.6 used, which includes variety of diseases, among which MPA i one very specific one - data was collected for billing purposes than research - no focus on EGPA - unadjusted costs
Reinhold-Keller, E., Herlyn, K., Wagner-Bastmeyer, W. et al.	Effect of Wegener's Granulomatosis on Work Disability, Need for Medical Care, and Quality of Life in Patients Younger Than 40 Years at Diagnosis	2002	Arthritis & Rheumatism	Evaluate effects of Wegener's granulomatosis (WG) on employment status, work disability, and need for medical care of 60 WG patients	- Hospital-based (2 Centers) - cohort - survey via self-administered questionnaires on hospitalization and QoL	Freiburg and Lubeck, Germany	1996	- self-administered questionnaire on hospitalization and QoL - SF-36	GPA	- German patients with WG (n=60) - median age at diagnosis was 31 - median age was 36 - All fulfilled the American College of Rheumatology 1990 criteria for the classification of WG and definition of Chapel Hill	60	- Hospitalization - LOS - Physician visits	- 55% had been hospitalized prior to survey - median LOS 17 days (4-140) - "4 other patients were hospitalized for a median of 19 days (5-37)" - 33.3% visited a physician once or less per quarter - 31,7% visited a physician 1-2 times per month	- age and LOS given as median - self-reported data - small and young study population - short study duration - Focus not on hospitalization/ healthca utilization, but work outcomes
Shut, N., Seyahi, E., Yurdakul, S. et al.	A cost analysis of Behcet's syndrome in Turkey	2007	Rheumatology	Estimation of direct and indirect costs, linked to Behcet's	- Single-center (multi-disciplinary BS out-patient clinic in Istanbul) - hospital-based - survey with standardized questionnaire	Istanbul, Turkey	03.-06.2015	- Standardized questionnaire, addressing: - direct costs (medication, diagnostic tests, hospital visits, hospitalization fees and lodging and transporation expenses), - indirect costs (los workdays and wages)	Behcet's syndrome	- Turkish patients with BS - fulfilling the International Study Group diagnostic criteria for Behc, et's disease criteria (ISG) - 87% had social security coverage - mean age was 35 - 65.55% male - 34.45% female		- cost of illness	- mean annual total cost per patient: 3226-3488USD - direct costs account for 2203-2771USD (68%) of this - of the direct costs, medication costs accounted for 1746-2646USD (79%)	- only 87% had social security coverage - retrospective survey method is subject recall bias - small population - patients well-enough to not need med help were not included, which would ha changed results on costs - out-patient visits were not considered - no lifelong economic impact of BS, cou be achieved by prospective studies

Author	Title	Year	Journal	Original purpose	Design and Setting	Location	Period	Data source	Type of vasculitis	Population	Sample size	Outcome of interest	Result of interest	Limitations of interest
Thorpe, C.T., Thorpe, J.M., Jiang, T. et al.	Healthcare utilization and expenditures for United States Medicare beneficiaries with systemic vasculitis	2017	Science in Arthritis and Rheumatism	Comparison of healtcare utilization and expenditures for medicare beneficiaries with versus without systemic vasculitis	- national, - retrospective - observational - cohort study	US	2010	- claims data (2010) - enrollment data	systemic vasculitis	- n=176,498 (controls n= 46,561) - 100% cohort of Medicare Part A and B beneficiaries - >= 1 claim including a diagnosis for a form of SV - continuously enrolled - LVV: GCA or TA (42.5%) - SVV: GPA, EGPA, Cryoglobulinaemic V, H-S-Purpura or Goodpasture's (18.8%) - Variable vessle vasculitis: Behcet's, Cogan's or other (including MPA): 31.6%	176.498	- costs of illness	- **SV:** mean **per-beneficiary** expenditure was 21,551.60 USD (SD = 35,345.50USD; range = 0-1,187,892.00USD) - **Non-SV**: mean **per-beneficiary** expenditure was 10,518.70USD (SD = 22,661.20USD; range = 0-522,506.60USD) - Average annual **SV**-beneficiary expenditure on **medical services** was 3,329.80USD (SD = 4,545.20USD; range = 0-242,445USD) - Average annual **non-SV**=beneficiary expenditure on **medical services** was 1,785.70USD (SD = 3,167.50; range = 127,286.40) - **Medicare** total annual expenditure for the **cohort** were 3.80 Billion USD (95% CI = 3.77-3.83 Billion USD) - **SV**-beneficiary total annual expenditure for **medical services** were 588 million USD (95% CI = 584-891 Million USD) - mean **incremental** expenditure on **medical services** by **Medicare** (for SV compared to non-SV): 11,004.48USD (95% CI = 10,728.66-11,280.30) -> more than twice as much! - mean **incremental** expenditure **SV**-beneficiaries (compared to non-SV): 1,547.24USD (95% CI = 1,5 - MORE RESTRICTIVE CRITERIA TO CLASSIFY SV **(>=2 claims)**: mean **incremental** expenditure on **medical services** by **Medicare** was 14,035.08USD (slightly more expensive than with the less restrictive classification criteria) - **> 65yrs** was cheaper than **<65yrs** (mean per-beneficiary expenditure by **Medicare** for **medical services** (-65 + SV: 30,100.46USD (SD= 49,546.35USD); -65 -SV: 11,885.00USD (SD= 27,392.08USD); 65+ +SV: 20,185.16USD; 65+ -SV: 9,750.81USD - higher healthcare utilization of SV-beneficiaries compared to non-SV-beneficiaries in (age- and sex matched regression; coefficient [95%CI]): inpatient stays (0.43 [0.42-0.45]), readmissions (0.13 [0.12-0.13]), emergency department visits (0.61 [0.59-0.63]), outpatient visits (22.45 [22.05-22.84]), skilled nursing stays (0.09 [0.080-0.09]), home health visits (2.9 [2.63-3.17]), ambulatory surgery (0.12 [0.11-0.14]), anethisia (0.37 [0.36-0.38]), dialysis (0.40 [0.37-0.42]), imaging (4.66 [4.58-4.75]), tests (16.86 [16.57-17.11]), other procedures (5.28 [5.086-5.48]), durable medical equipment (1.31 [1.225-1.39]), part B medications (1.70 [1.607-1.79]), other Part B events (1.14 [0.989-1.29]). - only no differences or slightly less utilization was seen in hospice stays (-0.000 [-0.003–0.002]) "Although a rare disease, (...) it is a disproportionately costly one for both medicare and patients."	- Medicare is the US federal health insurance program for all individuals aged >=65yrs and younger persons with disabilities, therefore only the older populations with SV are included - diagnosis based on ICD-9-CM - SV seen as a bunch of diagnoses, not accurately splitted up - no specific focus on AAV population - patients that are differently or not at all insured are not covered - single year of claims data (no distinguishing btw incident and prevalent SV cases) - only a partly population of Medicare beneficiaries included, which is the one that i slightly less healthy and with greater healthcare utilization - no generalization possible for the whole nationwide population
Wallace, Z.S., Lu, N., Miloslavsky, E. et al.	Nationwide Trends in hospitalizations and In-Hospital Mortality in Granulomatosis with Polyangiitis (Wegener's)	2016	American College of Rheumatology	Evaluation of nationwide trends in hospitalization and in-hospital moralitz over past 2 decades	- Nationwide - cohort study - observational - longitudinal	US	1993-2011	- National Inpatient Sample (NIS) - largest publicly available all-player inpatient database in the US - includes data and sampling weights from more than 1,000 hospitals in 44 states (97% of US pop) - covers Medicare, Medicaid, private insurance and uninsured - created by Agency for Healthcare Research and Quality	GPA	- All US GPA patients covered by Medicare, Medicaid, private insurance and no insurance - hospitalized btw 1993-2011 - mean age 1993: 55.4 - mean age 2011: 52.8 - 1993 & 2011: 50.3% male, 49.7% female - 1993: 35.6% Medicare, 6.6% Medicaid, 51.9% Privately insured - 2011: 40.7% Medicare,		- Hospitalization rate - LOS	- Hospitalization rate per 100,000, overall - 1993: 5.1 (p<0.0001) - 2011: 6.3 (p<0.0001) - increased by 24% (p<0.0001) - Hospitalization rate per 100,000, <65 - 1993: 3.7 (p<0.0001) - 2011: 4.8 (p<0.0001) - Hospitalization rate per 100,000, >=65 - 1993: 14.5 (p=0.049) - 2011: 16.2 (p=0.049) - Median length of stay in days - 1993: 6.9 (p=0.001) - 2011: 5.5 (p=0.001) - declined by 20% (p=0.001)	- diagnosis based on ICD-9-CM coding, potential for misclassification bias (principal discharge diagnoses limitation should improv the validity of the case definition) - readmissions not captured - They used the ICD-9-CM 2010 diagnosis cod 446.4, which also includes Eosinophilic granulomatosis with polyangiitis, which does not appear in any of the analyses

Author	Title	Year	Journal	Original purpose	Design and Setting	Location	Period	Data source	Type of vasculitis	Population	Sample size	Outcome of interest	Result of interest	Limitations of interest
GREY LITERATURE														
Belk, K.W., Craver, C.W.,	Hospital-Based Resource Utilization among Wegener's Granulomatosis Patients	2014	MedAssets	- Examine hospital-based utilization in WG - identify LOS drivers	- Retrospective - cross-sectional analysis - Hospital based	Charlotte, NC, USA	2009-2013	MedAssets Health System Data	Wegener's Granulomatosis	-Adult patients with WG - 54.6% females - 45.3% males - 56.8% between 40 and 69	7.202	- number of inpatient visits (n=7,202) - outpatient visits (n=24,971) - LOS	- hospital utilization was primarily for outpatient services (77.6%) - mean was 2.5 outpatient visits per patient - average LOS was 8.4 days	- no CI for outcomes of interest - Primary diagnosis was WG for 58.6% of inpatients and 16.6% of outpatients - more women than men - Abstract only
Cotch, M.F., Hoffman, G.S.	The Prevalence, Epidemiology and Cost of hospitalizations for Vasculitis in New York State: 1986 to 1990	1995	Arthritis & Rheumatism, 1995 National Scientific Meeting	Determine prevalence, epidemiology and the cost of hospitalizations	- Analysis of hospitalization data from all non-federal, non-psychiatric facilities in the State of New York	New York, USA	1986-1990	Hospitalization data, based on ICD-9 diagnoses	- Wegener's granulomatosis - GCA - TA	- NY patients with WG (n=571) - 48% male - 83% female - mean age: 56	571	- LOS - Costs	- 12,023USD was the average charge per admission of one WG patient	- Abstract only - gender percentages exceed 100 - no no. of admissions - no possibility to follow calculation - code-based diagnosis
McCormick, N., Marra, C.A., Colley, L. et al.	Growing prescription drug costs, despite more efficient health care delivery, in cases with systemic autoimmune rheumatic diseases in British Columbia, Canada: a population-based study	2012	J Popul Ther Clin Pharmacol	Estimation of health care burden of SARDs	- Cohort of SARDs cases - longitudinal - Population-based	Canada, British Columbia	1996-2007	- paid claims (costs) - case-mix (hospitalizations) - "administrative data, capturing all provincially-funded outpatient services, hospitalizations (1990-2010), dispensed prescriptions ()1995-2010"	Systemic vasculitis	Canadians (n= 18.741, contributing 82,140PY) with SARDs in british Columbia, Canada	18.741	- hospitalizations - outpatient services - prescriptions - related costs	- costs per patient year decreased by 20% over 12 years, from 8901-7123CAD/PY - outpatient encounters and costs decreased by 19% from 34-27/PY and 26% 2205-1641CAD/PY - Mean annual hospital costs decreased by half, from 5579-2776CAD/PY - Admissions decreased by 45% from 0.89-0.48/PY - Dispensed prescriptions increased by 49% (23-34/PY - Prescription costs increased by 50% frin 1117-1670CAD/PY - Comorbidities and complications may be the main contributors to the increase in medication costs, as new therapies only accounted for 4% of the costs (rituximab, etc.)	- no CI for outcomes of interest - Focus on group of diagnoses, not vasculitis only - Abstract only - casemix used, must base on coding - SARDs include Lupus erythematosus, syste sclerosis, Sjoergen's disease, poly/dermatomyositis, and systemic vasculit
McCormick, N., Marra, C., Avina-Zubieta, J.A.	Longitudinal, Incremental Direct Medical Coss of Giant Cell Arteritis for the First Five Years Following Diagnosis: A General Population-Based Cohort Study	2015	ACR poster Session at ACR/ARHP Annual Meeting	Determine the incremental direct medical cost of a general population-based cohort of incident GCA for the first five years after diagnosis	- Population-based cohort of GCA patients, - longitudinal	Canada, British Columbia	1996-2010	- billing data - case-mix - "administrative data, capturing all provincially-funded outpatient services, hospitalizations (1990-2010), dispensed prescriptions (1995-2010)"	Giant Cell Arteritis (GCA)	- Canadian patients (n=797, 7,970 controls) with incident GCA in british columbia - mean age 76 - 72% female - Charlson-Romano comorbidity index of 0 - new diagnosis btw 1/1996-12/2010	797	- absolute and incremental costs for outpatient services and prescriptions as per PY for the first five yrs after diagnosis (summed from billing data; case-mix methodology used for hospitalizations) - absolute and incremental health service utilisation, defined as outpatient encounters, inpatient admissions and dispensed prescriptions	- unadjusted incremental costs of GCA for the first 5 yrs after diagnoses was on average 26.48 CAD/PY with - 78% from hospitalizations, - 15% from outpatient and - 6% from medications - incremental costs decreased over time: - 41.113 CAD/PY for diagnoses over 1996-2002, 34.732 CAD/PY accounting for hospitalizations - 19.033 CAD/PY for diagnoses over 2003-2010, 13.531 CAD/PY accounting for hospitalizations (changed by 61%) - absolute costs for GCA have recently decreased by 54% - Adjusted mean/PY cost ratios between GCA cases and matched controls (95%CI), over the first five years after diagnosis - Outpatient costs: 1.7 (1.6-1.8) - Inpatient hospitalization costs (among hospitalized): 1.3 (1.2-1.4) - Medication costs: 1.2 (1.1-1.4) - Overall costs: 1.9 (1.7-2.0) - Adjusted mean/PY incremental utilization (95%CI) of GCA cases - Outpatient encounters: 55.9 (50.3-61.6) - Inpatient admissions: 1.4 (1.05-1.76) - Dispensed prescriptions: 65.5 (42.9-88.2)	- inclusion criteria were >=40yrs , new GCA diagnosis btw. 1996 and 2010, use of oral glucocorticoids - 598 (from 797) lost to follow-up - majority female (72%) - comparably high mean age of 76 - no self anouncement of limitations - might be partly the same population as see above - Abstract only
Raimundo, K., Farr, A.M., Kim, G. et al.	Clinical and Economical Burden of Granulomatosis with Polyangiitis (GPA) in the US	2015	The Journal of Rheumatology	Describe clinical and economic burden of patients with GPA in the US	- Retrospective, - claims-based - cohort study	**US**	2009-2013	- MarketScan - Medicare Supplemental Databases	GPA	- US patients with GPA		- inpatient admissions - ER visits - mean total annual cost	- inpatient admissions: 1.8 - ER visits: 32% had more than 1 - mean total annual cost: 41,400USD - on average, 58.7% off all-cause costs per GPA patien was associated with GPA - "Reducing the risk of relapse can contribute to decrease the clinical burden and total healthcare costs for this population"	- no CIs - diagnosis based on coding - Abstract only

Newcastle-Ottawa Scale (NOS) quality assessment of sources included in the systematic literature review

Belk, K., & Craver, C. (2014). Hospital-Based Resource Utilization Among Wegener's Granulomatosis Patients. *Value in Health*, *17*(3), A230.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Cotch, M. F., & Hoffmann, G. S. (1995). THE PREVALENCE, EPIDEMIOLOGY AND COSTS OF HOSPITALIZATIONS FOR VASCULITIS IN NEW YORK STATE: 1986 TO 1990. *Arthritis and Rheumatism*, *39*(9 (Supplement)).

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Cotch, M. F., Hoffman, G. S., Yerg, D. E., Kaufman, G. I., Targonski, P., & Kaslow, R. A. (1996). The epidemiology of Wegener's granulomatosis. Estimates of the five-year period prevalence, annual mortality, and geographic disease distribution from population-based data sources. *Arthritis Rheum*, *39*(1), 87–92.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Foocharoen, C., Thavornpitak, Y., Mahakkanukrauh, A., & Suwannaroj, S. (2012). Characteristics of hospitalized patients with systemic connective tissue disorders in Thailand: Analysis from a nationwide hospital database. *International Journal of Rheumatic Diseases*, *15*(SUPPL. 1), 80.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Janisiewicz, A. M., Klau, M. H., Keschner, D. B., Lehmer, R. R., Venkat, K. V., Medhekar, S. S., … Lee, J. T. (2015). Higher antineutrophil cytoplasmic antibody (C-ANCA) titers are associated with increased overall healthcare use in patients with sinonasal manifestations of granulomatosis with polyangiitis (GPA). *American Journal of Rhinology and Allergy*, *29*(3), 202–206.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Krulichova, I., Gamba, S., Ricci, E., & Garattini, L. (2004). Direct medical costs of monitoring and treating patients with Takayasu arteritis in Italy. *European Journal of Health Economics*, *5*(4), 330–334.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

McCormick, N., & Marra, C. (2015). Longitudinal, incremental direct medical costs of giant cell arteritis for the first five years following diagnosis: A general population-based cohort study. *Arthritis and Rheumatology*, *67*(SUPPL. 10), 2–5.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

McCormick, N., Marra, C., Colley, L., Grubisic, M., Sayre, E., & Avina-Zubieta, J. (2012). Growing prescription drug costs, despite more efficient health care delivery, in cases with systemic autoimmune rheumatic diseases in British Columbia, Canada: a population-based study, *19*(June 2014), e133.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Michet, C. J., Achenbach, S. J., Crowson, C. S., & Matteson, E. L. (2015). Hospitalization rates and utilization among patients with giant cell arteritis: A population-based study from 1987 to 2012. *Seminars in Arthritis and Rheumatism*, *45*(1), 70–74.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Raimundo, K., Farr, A. M., Kim, G., & Duna, G. (2015). Clinical and economic burden of antineutrophil cytoplasmic antibody-associated vasculitis in the United States. *Journal of Rheumatology*, *42*(12), 2383–2391.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Raimundo, K., Farr, A., Kim, G., & Duna, G. (2015). Clinical and economic burden of granulomatosis with polyangiitis (GPA) in the US (Scientific Abstract). *Annals of the Rheumatic Diseases*, 521

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Reinhold-Keller, E., Herlyn, K., Wagner-Bastmeyer, R., Gutfleisch, J., Peter, H. H., Raspe, H. H., & Gross, W. L. (2002). Effect of Wegener's granulomatosis on work disability, need for medical care, and quality of life in patients younger than 40 years at diagnosis. *Arthritis and Rheumatism*, *47*(3), 320–325.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Sut, N., Seyahi, E., Yurdakul, S., Senocak, M., & Yazici, H. (2007). A cost analysis of Behçet's syndrome in Turkey. *Rheumatology*, *46*(4), 678–682.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Thorpe, C. T., Thorpe, J. M., Jiang, T., Atkinson, D., Kang, Y., Schleiden, L. J., … Hogan, S. L. (2018). Healthcare utilization and expenditures for United States Medicare beneficiaries with systemic vasculitis. *Seminars in Arthritis and Rheumatism*, *47*(4), 507–519.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Wallace, Z. S., Lu, N., Miloslavsky, E., Unizony, S., Stone, J. H., & Choi, H. K. (2017). Nationwide Trends in Hospitalizations and In-Hospital Mortality in Granulomatosis With Polyangiitis (Wegener's). *Arthritis Care and Research*, *69*(6), 915–921.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

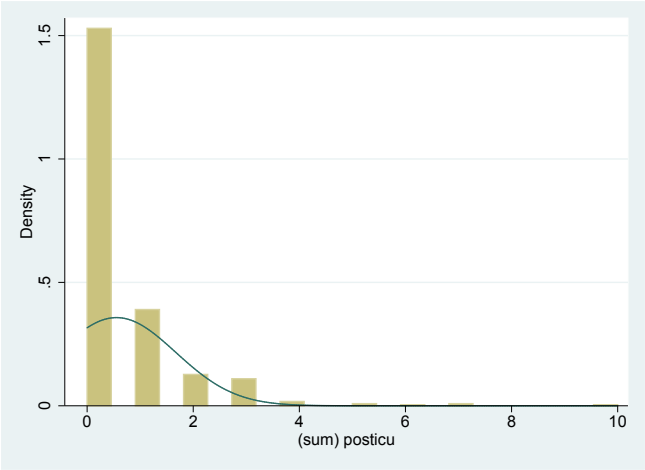
Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

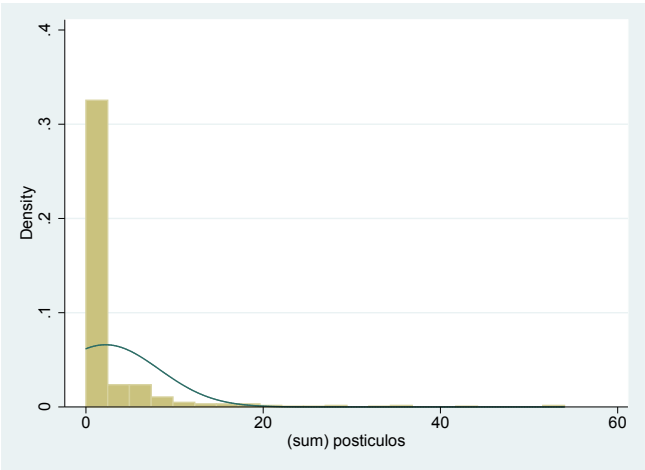
2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

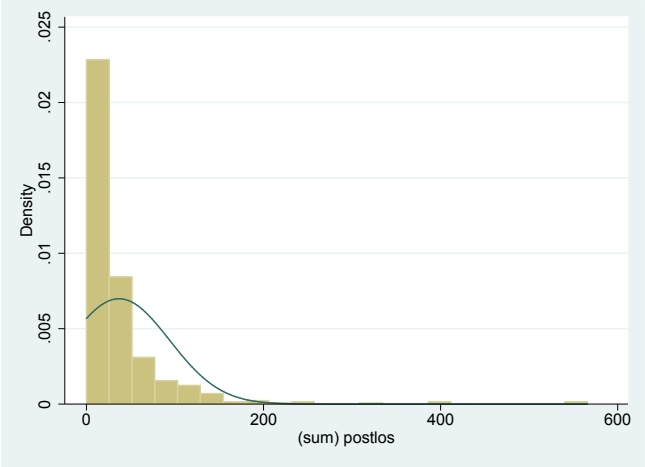
APPENDIX II: Poisson Regression Supplement



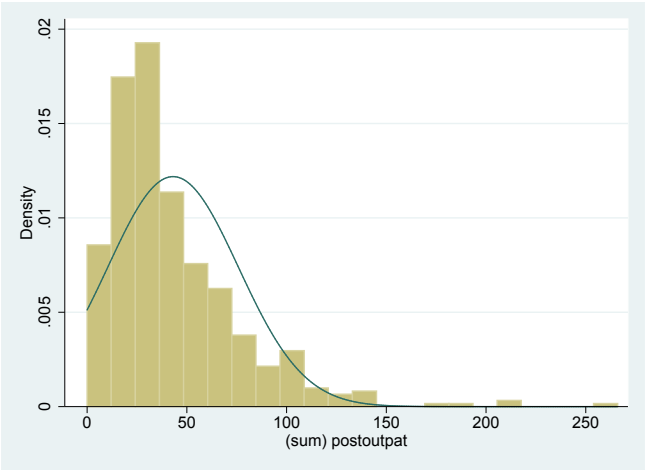
Appendix Figure 1 Histogram of ICU episodes in AAV patients, post diagnosis



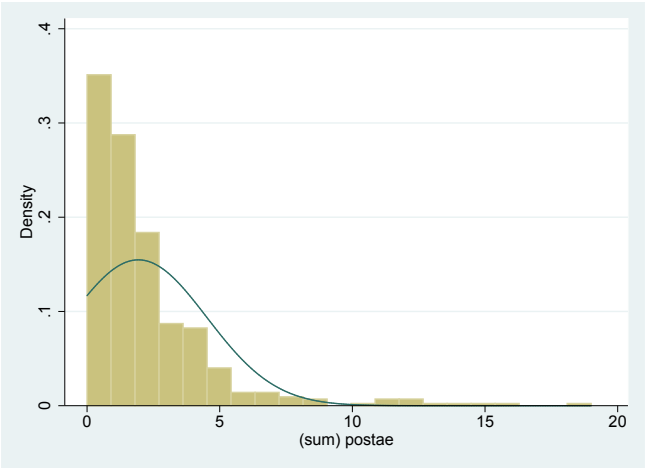
Appendix Figure 2 Histogram of length of stay on ICU in AAV patients, post diagnosis



Appendix Figure 3 Histogram of length of stay (general ward) in AAV patients, post diagnosis



Appendix Figure 4 Histogram of outpatient encounters in AAV patients, post diagnosis



Appendix Figure 5 Histogram of A&E in AAV patients, post diagnosis

Translation of codes

Sex (1=male, 2=female)

AAV type (1=GPA, 2=MPA, 3=EGPA)

ANCA status (1=PR3, 2=MPO, 3=ANCA negative)

Socio-economic status (Q1=most deprived, Q5=most affluent)

NHS Health board (1=Grampian, 2=Lothian, 3=Glasgow, 4=Tayside, 5=Highland, 6=Fife)

Univariable Poisson regression analyses

12: Inpatient hospital admission (general ward) and age

13: Inpatient hospital admission (general ward) and sex

14: Inpatient hospital admission (general ward) and AAV type

15: Inpatient hospital admission (general ward) and ANCA status

16: Inpatient hospital admission (general ward) and socio-economic status (SIMD-Quintiles)

17: Inpatient hospital admission (general ward) and NHS health board

19: Length of stay (general ward) and age

20: Length of stay (general ward) and sex

21: Length of stay (general ward) and AAV type

22: Length of stay (general ward) and ANCA status

23: Length of stay (general ward) and socio-economic status (SIMD-Quintiles)

24: Length of stay (general ward) and NHS health board

```
12. ///uni
  > poisson postadmis ageatindex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood =  -2231.361
Iteration 1:   log pseudolikelihood = -2231.3568
Iteration 2:   log pseudolikelihood = -2231.3568

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       9.82
                                                Prob > chi2       =     0.0017
Log pseudolikelihood = -2231.3568               Pseudo R2         =     0.0156
```

postadmis	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ageatindex	1.010438	.0033479	3.13	0.002	1.003897	1.017021
_cons	.6763721	.1420799	-1.86	0.063	.448106	1.020917
ln(fup)	1	(exposure)				

```
13. poisson postadmis i.sex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood =  -2263.797
Iteration 1:   log pseudolikelihood =  -2263.797

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       0.71
                                                Prob > chi2       =     0.3990
Log pseudolikelihood =  -2263.797               Pseudo R2         =     0.0013
```

postadmis	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.sex	1.086639	.1070449	0.84	0.399	.8958464	1.318067
_cons	1.168661	.0703415	2.59	0.010	1.038616	1.314989
ln(fup)	1	(exposure)				

```
14. poisson postadmis i.aavtype, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2215.0723
Iteration 1:   log pseudolikelihood = -2215.0647
Iteration 2:   log pseudolikelihood = -2215.0647

Poisson regression                              Number of obs     =        500
                                                Wald chi2(2)      =      13.17
                                                Prob > chi2       =     0.0014
Log pseudolikelihood = -2215.0647               Pseudo R2         =     0.0198
```

postadmis	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
aavtype						
2	1.112324	.1128365	1.05	0.294	.9117669	1.356998
3	.6337015	.0949252	-3.05	0.002	.4724753	.849944
_cons	1.254873	.0816005	3.49	0.000	1.104711	1.425446
ln(fup)	1	(exposure)				

```
15. poisson postadmis i._ancastatus, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2253.2819
Iteration 1:   log pseudolikelihood = -2253.2722
Iteration 2:   log pseudolikelihood = -2253.2722

Poisson regression                              Number of obs     =        499
                                                Wald chi2(2)      =       0.39
                                                Prob > chi2       =     0.8223
Log pseudolikelihood = -2253.2722               Pseudo R2         =     0.0005

------------------------------------------------------------------------------
             |               Robust
  postadmis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _ancastatus |
           2 |   .9440207   .0887064    -0.61   0.540     .7852303    1.134922
           3 |   1.005861   .2240485     0.03   0.979     .6500382    1.556457
             |
       _cons |   1.233086   .0706373     3.66   0.000     1.102129    1.379604
    ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------

16. poisson postadmis i.simdquin00, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2220.9985
Iteration 1:   log pseudolikelihood = -2220.9931
Iteration 2:   log pseudolikelihood = -2220.9931

Poisson regression                              Number of obs     =        497
                                                Wald chi2(4)      =      10.86
                                                Prob > chi2       =     0.0282
Log pseudolikelihood = -2220.9931               Pseudo R2         =     0.0137

------------------------------------------------------------------------------
             |               Robust
  postadmis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  simdquin00 |
           2 |   .8106477   .1280778    -1.33   0.184     .5947673    1.104886
           3 |   .7520366   .1078273    -1.99   0.047     .5677974    .9960578
           4 |   .6345483   .0895999    -3.22   0.001     .4811415    .8368672
           5 |   .7987099   .1353239    -1.33   0.185     .5730225    1.113285
             |
       _cons |    1.59552   .1783587     4.18   0.000      1.28159     1.98635
    ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------
```

```
17. poisson postadmis i.healthboard, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood =  -2211.216
Iteration 1:   log pseudolikelihood = -2211.0719
Iteration 2:   log pseudolikelihood = -2211.0719

Poisson regression                              Number of obs     =        502
                                                Wald chi2(5)      =      17.17
                                                Prob > chi2       =     0.0042
Log pseudolikelihood = -2211.0719               Pseudo R2         =     0.0246

------------------------------------------------------------------------------
             |               Robust
   postadmis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  healthboard |
 NHS Lothian |   1.206073   .1436704     1.57   0.116      .954941    1.523249
NHS Greater Glasgow and Clyde |   1.451359   .2555649     2.12   0.034     1.027755    2.049557
 NHS Tayside |   1.034902   .1461999     0.24   0.808     .7846043    1.365048
 NHS Highland |   .9441748   .1309688    -0.41   0.679     .7194162    1.239152
    NHS Fife |   1.752984   .2911452     3.38   0.001     1.265919    2.427449
             |
       _cons |   1.053443   .0765396     0.72   0.474     .9136202    1.214665
      ln(fup) |          1  (exposure)
------------------------------------------------------------------------------

19. poisson postlos ageatindex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -14166.275
Iteration 1:   log pseudolikelihood = -14166.275

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =      53.36
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -14166.275               Pseudo R2         =     0.1224

------------------------------------------------------------------------------
             |               Robust
     postlos |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ageatindex |   1.035712   .0049749     7.31   0.000     1.026007    1.045508
       _cons |    .839761   .2459009    -0.60   0.551     .4730475    1.490756
      ln(fup) |          1  (exposure)
------------------------------------------------------------------------------

20. poisson postlos i.sex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -16003.786
Iteration 1:   log pseudolikelihood = -16003.786

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       2.91
                                                Prob > chi2       =     0.0881
Log pseudolikelihood = -16003.786               Pseudo R2         =     0.0085

------------------------------------------------------------------------------
             |               Robust
     postlos |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |   1.276419   .1826532     1.71   0.088     .9642458    1.689657
       _cons |   5.837558   .5663968    18.18   0.000      4.82661    7.060251
      ln(fup) |          1  (exposure)
------------------------------------------------------------------------------
```

```
21. poisson postlos i.aavtype, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -15628.903
Iteration 1:   log pseudolikelihood = -15628.874
Iteration 2:   log pseudolikelihood = -15628.874

Poisson regression                              Number of obs     =        500
                                                Wald chi2(2)      =      14.63
                                                Prob > chi2       =     0.0007
Log pseudolikelihood = -15628.874               Pseudo R2         =     0.0285
```

postlos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
aavtype						
2	1.417399	.2080755	2.38	0.017	1.063004	1.889945
3	.6690677	.1367483	-1.97	0.049	.448225	.9987206
_cons	6.311086	.6493831	17.90	0.000	5.158449	7.721274
ln(fup)	1	(exposure)				

```
22. poisson postlos i._ancastatus, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -15713.442
Iteration 1:   log pseudolikelihood = -15713.422
Iteration 2:   log pseudolikelihood = -15713.422

Poisson regression                              Number of obs     =        499
                                                Wald chi2(2)      =       4.79
                                                Prob > chi2       =     0.0910
Log pseudolikelihood = -15713.422               Pseudo R2         =     0.0078
```

postlos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_ancastatus						
2	1.146155	.1713161	0.91	0.361	.8550939	1.536288
3	.7625763	.1423227	-1.45	0.146	.5289567	1.099377
_cons	6.462148	.6856854	17.59	0.000	5.248771	7.956026
ln(fup)	1	(exposure)				

```
23. poisson postlos i.simdquin00, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -15958.699
Iteration 1:   log pseudolikelihood = -15958.684
Iteration 2:   log pseudolikelihood = -15958.684

Poisson regression                              Number of obs     =        497
                                                Wald chi2(4)      =       2.91
                                                Prob > chi2       =     0.5736
Log pseudolikelihood = -15958.684               Pseudo R2         =     0.0080
```

postlos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
simdquin00						
2	1.045582	.2441257	0.19	0.849	.6616324	1.652339
3	1.101226	.2494099	0.43	0.670	.7064687	1.716563
4	.8513744	.1874151	-0.73	0.465	.5530234	1.310683
5	.8327326	.1644179	-0.93	0.354	.565511	1.226225
_cons	6.916431	1.085354	12.32	0.000	5.085201	9.407104
ln(fup)	1	(exposure)				

```
24. poisson postlos i.healthboard, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -15718.576
Iteration 1:   log pseudolikelihood = -15718.198
Iteration 2:   log pseudolikelihood = -15718.198

Poisson regression                              Number of obs     =        502
                                                Wald chi2(5)      =      14.37
                                                Prob > chi2       =     0.0134
Log pseudolikelihood = -15718.198               Pseudo R2         =     0.0262
```

postlos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
healthboard						
NHS Lothian	.676575	.1077388	-2.45	0.014	.4951871	.9244055
NHS Greater Glasgow and Clyde	.8192371	.2036629	-0.80	0.423	.5032692	1.33358
NHS Tayside	.4838394	.1009158	-3.48	0.000	.3214871	.7281801
NHS Highland	.7024784	.1681176	-1.48	0.140	.4394636	1.122905
NHS Fife	.9355861	.2497899	-0.25	0.803	.5544004	1.578861
_cons	8.008169	1.008976	16.51	0.000	6.255867	10.2513
ln(fup)	1	(exposure)				

```
26. poisson posticu ageatindex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -634.03907
Iteration 1:   log pseudolikelihood =  -634.0388
Iteration 2:   log pseudolikelihood =  -634.0388

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       3.72
                                                Prob > chi2       =     0.0539
Log pseudolikelihood =  -634.0388               Pseudo R2         =     0.0062
```

posticu	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ageatindex	1.012315	.006428	1.93	0.054	.9997941	1.024992
_cons	.0493886	.0193925	-7.66	0.000	.0228771	.1066234
ln(fup)	1	(exposure)				

```
27. poisson posticu i.sex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -637.98324
Iteration 1:   log pseudolikelihood = -637.98324

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       0.01
                                                Prob > chi2       =     0.9027
Log pseudolikelihood = -637.98324               Pseudo R2         =     0.0000
```

posticu	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.sex	.9768519	.1871317	-0.12	0.903	.6710693	1.421969
_cons	.0996236	.0120013	-19.15	0.000	.0786721	.1261547
ln(fup)	1	(exposure)				

```
28. poisson posticu i.aavtype, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -609.87027
Iteration 1:   log pseudolikelihood = -609.85898
Iteration 2:   log pseudolikelihood = -609.85898

Poisson regression                              Number of obs     =        500
                                                Wald chi2(2)      =      17.21
                                                Prob > chi2       =     0.0002
Log pseudolikelihood = -609.85898               Pseudo R2         =     0.0418

------------------------------------------------------------------------------
             |               Robust
    posticu |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    aavtype |
          2 |   1.744833   .3503634     2.77   0.006     1.177151    2.586281
          3 |   .2983329   .1356861    -2.66   0.008     .1223383    .7275116
            |
      _cons |   .0935273   .0106701   -20.77   0.000     .0747875    .1169629
   ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------

29. poisson posticu i._ancastatus, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -628.72103
Iteration 1:   log pseudolikelihood = -628.72077
Iteration 2:   log pseudolikelihood = -628.72077

Poisson regression                              Number of obs     =        499
                                                Wald chi2(2)      =       5.98
                                                Prob > chi2       =     0.0504
Log pseudolikelihood = -628.72077               Pseudo R2         =     0.0119

------------------------------------------------------------------------------
             |               Robust
    posticu |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _ancastatus |
          2 |   1.369428   .2829199     1.52   0.128     .9134473    2.053029
          3 |   .5901596   .1980782    -1.57   0.116     .3056866    1.139364
            |
      _cons |   .0942447   .0110863   -20.08   0.000      .074839    .1186824
   ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------

30. poisson posticu i.simdquin00, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -632.12937
Iteration 1:   log pseudolikelihood = -632.12891
Iteration 2:   log pseudolikelihood = -632.12891

Poisson regression                              Number of obs     =        497
                                                Wald chi2(4)      =       2.78
                                                Prob > chi2       =     0.5960
Log pseudolikelihood = -632.12891               Pseudo R2         =     0.0056

------------------------------------------------------------------------------
             |               Robust
    posticu |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  simdquin00 |
          2 |   .8396938   .3180018    -0.46   0.645     .3997247    1.763928
          3 |   .6326657   .2239236    -1.29   0.196     .3161571    1.266035
          4 |   .6313077   .2135724    -1.36   0.174     .3252956    1.225191
          5 |   .7705023   .2611897    -0.77   0.442     .3964864    1.497337
            |
      _cons |   .1354687   .0391343    -6.92   0.000      .076903    .2386352
   ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------
```

```
31. poisson posticu i.healthboard, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -628.69063
Iteration 1:   log pseudolikelihood =  -628.6872
Iteration 2:   log pseudolikelihood =  -628.6872

Poisson regression                              Number of obs     =        502
                                                Wald chi2(5)      =       7.83
                                                Prob > chi2       =     0.1657
Log pseudolikelihood =  -628.6872               Pseudo R2         =     0.0146

------------------------------------------------------------------------------
             |               Robust
     posticu |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  healthboard |
NHS Lothian  |   1.626432   .3434956     2.30   0.021     1.075142    2.460402
NHS Greater Glasgow and Clyde |   1.017547   .4850456     0.04   0.971     .3997628    2.590039
NHS Tayside  |   .6747842   .2965229    -0.90   0.371     .2851795    1.596657
NHS Highland |   1.402764   .4598902     1.03   0.302     .7377713     2.66715
NHS Fife     |   1.205028   .5085033     0.44   0.659     .5269909    2.755441
             |
       _cons |   .0800544   .0116868   -17.30   0.000     .0601343    .1065733
      ln(fup) |          1  (exposure)
------------------------------------------------------------------------------

33. poisson posticulos ageatindex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2564.7615
Iteration 1:   log pseudolikelihood = -2564.7594
Iteration 2:   log pseudolikelihood = -2564.7594

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       0.54
                                                Prob > chi2       =     0.4644
Log pseudolikelihood = -2564.7594               Pseudo R2         =     0.0013

------------------------------------------------------------------------------
             |               Robust
  posticulos |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ageatindex |   1.005529   .0075776     0.73   0.464     .9907865    1.020491
       _cons |   .2893611   .1339017    -2.68   0.007      .116828    .7166932
      ln(fup) |          1  (exposure)
------------------------------------------------------------------------------

34. poisson posticulos i.sex, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2563.9684
Iteration 1:   log pseudolikelihood = -2563.9684

Poisson regression                              Number of obs     =        502
                                                Wald chi2(1)      =       0.46
                                                Prob > chi2       =     0.4954
Log pseudolikelihood = -2563.9684               Pseudo R2         =     0.0016

------------------------------------------------------------------------------
             |               Robust
  posticulos |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |   1.187623   .2995552     0.68   0.495     .7244011    1.947053
       _cons |   .3633706   .0630608    -5.83   0.000     .2586001    .5105882
      ln(fup) |          1  (exposure)
------------------------------------------------------------------------------
```

```
35. poisson posticulos i.aavtype, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2434.4736
Iteration 1:   log pseudolikelihood = -2433.6859
Iteration 2:   log pseudolikelihood = -2433.6858

Poisson regression                              Number of obs     =        500
                                                Wald chi2(2)      =      22.62
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -2433.6858               Pseudo R2         =     0.0501
```

posticulos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
aavtype						
2	1.512577	.3879021	1.61	0.107	.9150103	2.500398
3	.1445205	.0697333	-4.01	0.000	.0561322	.372089
_cons	.4022249	.0664048	-5.52	0.000	.2910321	.5559003
ln(fup)	1	(exposure)				

```
36. poisson posticulos i._ancastatus, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2519.8578
Iteration 1:   log pseudolikelihood = -2519.8415
Iteration 2:   log pseudolikelihood = -2519.8415

Poisson regression                              Number of obs     =        499
                                                Wald chi2(2)      =       8.84
                                                Prob > chi2       =     0.0120
Log pseudolikelihood = -2519.8415               Pseudo R2         =     0.0162
```

posticulos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_ancastatus						
2	1.116136	.2960912	0.41	0.679	.6636029	1.877266
3	.3605574	.1359804	-2.70	0.007	.1721681	.7550855
_cons	.4165009	.0702539	-5.19	0.000	.2992528	.5796872
ln(fup)	1	(exposure)				

```
37. poisson posticulos i.simdquin00, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2511.6949
Iteration 1:   log pseudolikelihood = -2511.6756
Iteration 2:   log pseudolikelihood = -2511.6756

Poisson regression                              Number of obs     =        497
                                                Wald chi2(4)      =       5.30
                                                Prob > chi2       =     0.2578
Log pseudolikelihood = -2511.6756               Pseudo R2         =     0.0184

------------------------------------------------------------------------------
             |               Robust
  posticulos |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  simdquin00 |
           2 |   .6203143   .3291073    -0.90   0.368     .2192842    1.754754
           3 |   .5087136    .238586    -1.44   0.150     .2028901    1.275515
           4 |   .3866132   .1779778    -2.06   0.039     .1568286    .9530773
           5 |   .6422614   .2891167    -0.98   0.325     .2657907    1.551972
             |
       _cons |   .6961587   .2763732    -0.91   0.362      .31973    1.515769
    ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------

38. poisson posticulos i.healthboard, exp(fup) irr vce(robust)

Iteration 0:   log pseudolikelihood = -2560.7927
Iteration 1:   log pseudolikelihood = -2560.7531
Iteration 2:   log pseudolikelihood = -2560.7531

Poisson regression                              Number of obs     =        502
                                                Wald chi2(5)      =       0.85
                                                Prob > chi2       =     0.9740
Log pseudolikelihood = -2560.7531               Pseudo R2         =     0.0028

------------------------------------------------------------------------------
             |               Robust
  posticulos |        IRR   Std. Err.      z    P>|z|     [95% Conf.
> Interval]
-------------+----------------------------------------------------------------
             |
  healthboard |
  NHS Lothian |   1.168706    .300738     0.61   0.545     .7057803
>    1.935269
NHS Greater Glasgow and Clyde |   1.156633   .9746474     0.17   0.863     .2217802
>    6.032098
  NHS Tayside |    .906477   .7292663    -0.12   0.903     .1873095
>    4.386859
  NHS Highland |   .9428771   .3735746    -0.15   0.882     .4337136
>     2.04978
     NHS Fife |   1.379476   .6915531     0.64   0.521     .5164083
>    3.684978
             |
       _cons |   .3629739   .0658827    -5.58   0.000      .254317
>    .5180545
    ln(fup)  |          1  (exposure)
------------------------------------------------------------------------------
```

Multivariable Poisson regression on inpatient hospital admissions and goodness of fit test

```
94. xi: stepwise, pr(.05): poisson postadmis ageatindex (i.sex) (i._ancastatus) (i.simdq
  > uin00) (i.healthboard), exp(fup) irr vce(robust)
  i.sex             _Isex_0-1           (naturally coded; _Isex_0 omitted)
  i._ancastatus     _I_ancastat_1-3     (naturally coded; _I_ancastat_1 omitted)
  i.simdquin00      _Isimdquin0_1-5     (naturally coded; _Isimdquin0_1 omitted)
  i.healthboard     _Ihealthboa_1-6     (naturally coded; _Ihealthboa_1 omitted)
                        begin with full model
  p = 0.7079 >= 0.0500  removing _Isex_1
  p = 0.4917 >= 0.0500  removing _I_ancastat_2 _I_ancastat_3
  p = 0.1853 >= 0.0500  removing _Isimdquin0_2 _Isimdquin0_3 _Isimdquin0_4
                                 _Isimdquin0_5

  Poisson regression                              Number of obs     =        494
                                                  Wald chi2(6)      =      40.90
                                                  Prob > chi2       =     0.0000
  Log pseudolikelihood = -2135.6919               Pseudo R2         =     0.0464
```

postadmis	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ageatindex	1.011628	.0034601	3.38	0.001	1.004869	1.018433
_Ihealthboa_2	1.212297	.1446757	1.61	0.107	.9594595	1.531763
_Ihealthboa_3	1.637817	.2915476	2.77	0.006	1.155428	2.321601
_Ihealthboa_4	1.024662	.1414572	0.18	0.860	.7817542	1.343047
_Ihealthboa_5	.9815622	.1262695	-0.14	0.885	.762813	1.263041
_Ihealthboa_6	1.828131	.3165019	3.48	0.000	1.302087	2.566698
_cons	.54435	.1143599	-2.89	0.004	.3606229	.8216808
ln(fup)	1	(exposure)				

```
95. estat gof

          Deviance goodness-of-fit =  2624.643
          Prob > chi2(487)         =    0.0000

          Pearson goodness-of-fit  =  4015.756
          Prob > chi2(487)         =    0.0000
```

Multivariable Poisson regression on length of stay (general ward) and goodness of fit test

```
96. xi: stepwise, pr(.05): poisson postlos ageatindex (i.sex) (i._ancastatus) (i.simdqui
  > n00) (i.healthboard), exp(fup) irr vce(robust)
  i.sex             _Isex_0-1           (naturally coded; _Isex_0 omitted)
  i._ancastatus     _I_ancastat_1-3     (naturally coded; _I_ancastat_1 omitted)
  i.simdquin00      _Isimdquin0_1-5     (naturally coded; _Isimdquin0_1 omitted)
  i.healthboard     _Ihealthboa_1-6     (naturally coded; _Ihealthboa_1 omitted)
                    begin with full model
  p = 0.2032 >= 0.0500  removing _I_ancastat_2 _I_ancastat_3
  p = 0.1583 >= 0.0500  removing _Isimdquin0_2 _Isimdquin0_3 _Isimdquin0_4
                                 _Isimdquin0_5
  p = 0.1205 >= 0.0500  removing _Isex_1

  Poisson regression                              Number of obs     =        494
                                                  Wald chi2(6)      =      68.19
                                                  Prob > chi2       =     0.0000
  Log pseudolikelihood = -13304.208               Pseudo R2         =     0.1571

------------------------------------------------------------------------------
             |               Robust
    postlos |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ageatindex |    1.03744   .0051444     7.41   0.000     1.027406    1.047572
_Ihealthboa_2|   .6681043   .1020061    -2.64   0.008     .4953166    .9011679
_Ihealthboa_3|   1.195228   .2821864     0.76   0.450      .752466    1.898517
_Ihealthboa_4|   .4816977   .0915412    -3.84   0.000     .3319047    .6990944
_Ihealthboa_5|   .7627721   .1787654    -1.16   0.248     .4818422    1.207493
_Ihealthboa_6|   .9729676   .2609652    -0.10   0.919     .5751654    1.645902
       _cons |   .9089013   .2811329    -0.31   0.757     .4957138    1.666489
      ln(fup)|          1  (exposure)
------------------------------------------------------------------------------

97. estat gof

          Deviance goodness-of-fit =  24317.79
          Prob > chi2(487)         =    0.0000

          Pearson goodness-of-fit  =   50616.7
          Prob > chi2(487)         =    0.0000
```

Multivariable Poisson regression on ICU episodes and goodness of fit test

```
98. xi: stepwise, pr(.05): poisson posticu ageatindex (i.sex) (i._ancastatus) (i.simdqui
  > n00) (i.healthboard), exp(fup) irr vce(robust)
  i.sex             _Isex_0-1          (naturally coded; _Isex_0 omitted)
  i._ancastatus     _I_ancastat_1-3    (naturally coded; _I_ancastat_1 omitted)
  i.simdquin00      _Isimdquin0_1-5    (naturally coded; _Isimdquin0_1 omitted)
  i.healthboard     _Ihealthboa_1-6    (naturally coded; _Ihealthboa_1 omitted)
                        begin with full model
  p = 0.5261 >= 0.0500  removing _Isimdquin0_2 _Isimdquin0_3 _Isimdquin0_4
                                 _Isimdquin0_5
  p = 0.4939 >= 0.0500  removing _Isex_1
  p = 0.1663 >= 0.0500  removing _Ihealthboa_2 _Ihealthboa_3 _Ihealthboa_4
                                 _Ihealthboa_5 _Ihealthboa_6
  p = 0.1248 >= 0.0500  removing ageatindex

  Poisson regression                              Number of obs     =        494
                                                  Wald chi2(2)      =       6.09
                                                  Prob > chi2       =     0.0477
  Log pseudolikelihood = -626.22645               Pseudo R2         =     0.0122
```

posticu	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_I_ancastat_2	1.375729	.2839803	1.55	0.122	.9179658	2.061766
_I_ancastat_3	.5891732	.197788	-1.58	0.115	.3051343	1.137614
_cons	.094944	.0111688	-20.01	0.000	.0753937	.1195637
ln(fup)	1	(exposure)				

```
99. estat gof

          Deviance goodness-of-fit =  888.4054
          Prob > chi2(491)         =    0.0000

          Pearson goodness-of-fit  =  2419.985
          Prob > chi2(491)         =    0.0000
```

Multivariable Poisson regression on length of stay on ICU and goodness of fit test

```
100 xi: stepwise, pr(.05): poisson posticulos ageatindex (i.sex) (i._ancastatus) (i.simd
  > quin00) (i.healthboard), exp(fup) irr vce(robust)
  i.sex             _Isex_0-1          (naturally coded; _Isex_0 omitted)
  i._ancastatus     _I_ancastat_1-3    (naturally coded; _I_ancastat_1 omitted)
  i.simdquin00      _Isimdquin0_1-5    (naturally coded; _Isimdquin0_1 omitted)
  i.healthboard     _Ihealthboa_1-6    (naturally coded; _Ihealthboa_1 omitted)
                        begin with full model
  p = 0.9968 >= 0.0500  removing _Ihealthboa_2 _Ihealthboa_3 _Ihealthboa_4
                                 _Ihealthboa_5 _Ihealthboa_6
  p = 0.6240 >= 0.0500  removing _Isex_1
  p = 0.4120 >= 0.0500  removing ageatindex
  p = 0.1902 >= 0.0500  removing _Isimdquin0_2 _Isimdquin0_3 _Isimdquin0_4
                                 _Isimdquin0_5

  Poisson regression                              Number of obs     =        494
                                                  Wald chi2(2)      =       8.91
                                                  Prob > chi2       =     0.0116
  Log pseudolikelihood = -2510.0237               Pseudo R2         =     0.0164
```

posticulos	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_I_ancastat_2	1.121271	.2973182	0.43	0.666	.6668141	1.885458
_I_ancastat_3	.3599547	.1357784	-2.71	0.007	.1718567	.7539272
_cons	.419591	.070783	-5.15	0.000	.301462	.5840094
ln(fup)	1	(exposure)				

```
101 estat gof

          Deviance goodness-of-fit =  4509.365
          Prob > chi2(491)         =    0.0000

          Pearson goodness-of-fit  =  28469.57
          Prob > chi2(491)         =    0.0000
```

Multivariable Poisson regression on outpatient encounters and goodness of fit test

```
> xi: stepwise, pr(.05): poisson postoutpat ageatindex (i.sex) (i._ancastatus) (i.simd
> quin00) (i.healthboard), exp(fup) irr vce(robust)
i.sex             _Isex_0-1           (naturally coded; _Isex_0 omitted)
i._ancastatus     _I_ancastat_1-3     (naturally coded; _I_ancastat_1 omitted)
i.simdquin00      _Isimdquin0_1-5     (naturally coded; _Isimdquin0_1 omitted)
i.healthboard     _Ihealthboa_1-6     (naturally coded; _Ihealthboa_1 omitted)
                      begin with full model
p = 0.9016 >= 0.0500  removing _I_ancastat_2 _I_ancastat_3
p = 0.3256 >= 0.0500  removing ageatindex
p = 0.2019 >= 0.0500  removing _Isex_1
p = 0.0991 >= 0.0500  removing _Isimdquin0_2 _Isimdquin0_3 _Isimdquin0_4
                               _Isimdquin0_5

Poisson regression                              Number of obs     =        494
                                                Wald chi2(5)      =      50.47
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -4350.0426               Pseudo R2         =     0.0786
```

postoutpat	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Ihealthboa_2	.9898282	.0596275	-0.17	0.865	.8795959	1.113875
_Ihealthboa_3	1.271738	.1784694	1.71	0.087	.9659272	1.674369
_Ihealthboa_4	.4577994	.0588903	-6.07	0.000	.3557776	.5890768
_Ihealthboa_5	1.167347	.122672	1.47	0.141	.9500585	1.434331
_Ihealthboa_6	1.168278	.107326	1.69	0.090	.9757737	1.398761
_cons	7.691327	.3160256	49.65	0.000	7.096213	8.33635
ln(fup)	1	(exposure)				

```
122 estat gof

          Deviance goodness-of-fit =  6068.366
          Prob > chi2(488)         =    0.0000

          Pearson goodness-of-fit  =  7071.193
          Prob > chi2(488)         =    0.0000
```

Multivariable Poisson regression on A&E and goodness of fit test

```
> xi: stepwise, pr(.05): poisson postae ageatindex (i.sex) (i._ancastatus) (i.simdquin
> 00) (i.healthboard), exp(fup) irr vce(robust)
i.sex             _Isex_0-1           (naturally coded; _Isex_0 omitted)
i._ancastatus     _I_ancastat_1-3     (naturally coded; _I_ancastat_1 omitted)
i.simdquin00      _Isimdquin0_1-5     (naturally coded; _Isimdquin0_1 omitted)
i.healthboard     _Ihealthboa_1-6     (naturally coded; _Ihealthboa_1 omitted)
                      begin with full model
p = 0.6604 >= 0.0500  removing _Isex_1
p = 0.4241 >= 0.0500  removing _I_ancastat_2 _I_ancastat_3
p = 0.2862 >= 0.0500  removing _Isimdquin0_2 _Isimdquin0_3 _Isimdquin0_4
                               _Isimdquin0_5

Poisson regression                              Number of obs     =        461
                                                Wald chi2(6)      =      76.89
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -1068.1428               Pseudo R2         =     0.1042
```

postae	IRR	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ageatindex	1.014839	.0049441	3.02	0.002	1.005194	1.024575
_Ihealthboa_2	2.988598	.4725987	6.92	0.000	2.192117	4.074472
_Ihealthboa_3	4.496155	1.101492	6.14	0.000	2.781694	7.267301
_Ihealthboa_4	1.443151	.3264542	1.62	0.105	.9263222	2.248337
_Ihealthboa_5	2.348437	.543351	3.69	0.000	1.492241	3.695889
_Ihealthboa_6	2.456004	.5053046	4.37	0.000	1.64097	3.675847
_cons	.0691714	.0216213	-8.55	0.000	.0374854	.1276414
ln(fup)	1	(exposure)				

```
estat gof

       Deviance goodness-of-fit =   1288.63
       Prob > chi2(454)         =    0.0000

       Pearson goodness-of-fit  =  2208.964
       Prob > chi2(454)         =    0.0000
```

APPENDIX III: Linear Regression Supplement

Translation of codes

Sex (1=male, 2=female)

AAV type (1=GPA, 2=MPA, 3=EGPA)

ANCA status (1=PR3, 2=MPO, 3=ANCA negative)

Socio-economic status (Q1=most deprived, Q5=most affluent)

NHS Health board (1=Grampian, 2=Lothian, 3=Glasgow, 4=Tayside, 5=Highland, 6=Fife)

Univariable linear regression analyses on (log) costs post diagnosis

14: Age

15: Sex

16: AAV type

17: ANCA status

18: Socio-economic status

19: NHS Health board

14. regress log_post_costs ageatindex

Source	SS	df	MS	Number of obs	=	502
				F(1, 500)	=	3.50
Model	4.24646364	1	4.24646364	Prob > F	=	0.0619
Residual	606.172692	500	1.21234538	R-squared	=	0.0070
				Adj R-squared	=	0.0050
Total	610.419155	501	1.21840151	Root MSE	=	1.1011

log_post_c~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ageatindex	.0063705	.0034039	1.87	0.062	-.0003172	.0130582
_cons	9.730329	.2048797	47.49	0.000	9.327798	10.13286

15. regress log_post_costs i.sex

Source	SS	df	MS	Number of obs	=	502
				F(1, 500)	=	0.49
Model	.597724123	1	.597724123	Prob > F	=	0.4842
Residual	609.821431	500	1.21964286	R-squared	=	0.0010
				Adj R-squared	=	-0.0010
Total	610.419155	501	1.21840151	Root MSE	=	1.1044

log_post_c~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.sex	.0691715	.0988082	0.70	0.484	-.1249589	.2633018
_cons	10.07033	.0674604	149.28	0.000	9.937793	10.20287

16. regress log_post_costs i.aavtype

Source	SS	df	MS	Number of obs	=	500
				F(2, 497)	=	5.52
Model	13.2425049	2	6.62125247	Prob > F	=	0.0042
Residual	595.856935	497	1.19890731	R-squared	=	0.0217
				Adj R-squared	=	0.0178
Total	609.09944	499	1.22064016	Root MSE	=	1.0949

log_post_c~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
aavtype						
2	.0437242	.1098066	0.40	0.691	-.1720183	.2594667
3	-.4715686	.1524853	-3.09	0.002	-.771164	-.1719733
_cons	10.15083	.0649731	156.23	0.000	10.02317	10.27848

17. regress log_post_costs i._ancastatus

Source	SS	df	MS		
Model	5.40675365	2	2.70337683	Number of obs	= 499
Residual	602.967751	496	1.21566079	F(2, 496)	= 2.22
				Prob > F	= 0.1093
				R-squared	= 0.0089
				Adj R-squared	= 0.0049
Total	608.374505	498	1.22163555	Root MSE	= 1.1026

log_post_c~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_ancastatus						
2	-.1469086	.1069174	-1.37	0.170	-.3569754	.0631582
3	-.3060119	.1599987	-1.91	0.056	-.6203707	.008347
_cons	10.19135	.068117	149.62	0.000	10.05752	10.32518

18. regress log_post_costs ib(5).simdquin00

Source	SS	df	MS		
Model	3.30673688	4	.82668422	Number of obs	= 497
Residual	602.593449	492	1.22478343	F(4, 492)	= 0.67
				Prob > F	= 0.6096
				R-squared	= 0.0055
				Adj R-squared	= -0.0026
Total	605.900186	496	1.22157296	Root MSE	= 1.1067

log_post_c~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
simdquin00						
1	.0682371	.1857124	0.37	0.713	-.2966501	.4331243
2	-.1329595	.1566661	-0.85	0.396	-.4407766	.1748577
3	-.0948177	.1482216	-0.64	0.523	-.3860431	.1964078
4	-.1745663	.1421768	-1.23	0.220	-.4539148	.1047822
_cons	10.19201	.1045733	97.46	0.000	9.986545	10.39748

19. regress log_post_costs i.healthboard

Source	SS	df	MS		
Model	14.6476175	5	2.9295235	Number of obs	= 502
Residual	595.771538	496	1.20115229	F(5, 496)	= 2.44
				Prob > F	= 0.0337
				R-squared	= 0.0240
				Adj R-squared	= 0.0142
Total	610.419155	501	1.21840151	Root MSE	= 1.096

log_post_costs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
healthboard						
NHS Lothian	-.1645203	.117344	-1.40	0.162	-.3950729	.0660323
NHS Greater Glasgow and Clyde	-.5496123	.2032291	-2.70	0.007	-.9489084	-.1503161
NHS Tayside	-.5134502	.2200057	-2.33	0.020	-.9457081	-.0811922
NHS Highland	-.2682015	.1868615	-1.44	0.152	-.6353392	.0989362
NHS Fife	-.303119	.1819215	-1.67	0.096	-.6605509	.0543128
_cons	10.27932	.083567	123.01	0.000	10.11513	10.4435

Multivariable linear regression model on costs post diagnosis

```
21. regress log_post_costs i._ancastatus ib(5).simdquin00 i.healthboard, vce(robust)

Linear regression                               Number of obs     =        494
                                                F(11, 482)        =       2.14
                                                Prob > F          =     0.0163
                                                R-squared         =     0.0450
                                                Root MSE          =     1.0938

------------------------------------------------------------------------------
             |               Robust
log_post_costs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_ancastatus |
          2  |  -.1846129   .1074492    -1.72   0.086    -.3957397    .0265138
          3  |  -.3642599   .1624068    -2.24   0.025    -.6833728   -.0451471
             |
simdquin00 |
          1  |   .2136821   .1960049     1.09   0.276    -.1714475    .5988117
          2  |  -.0928685   .1572376    -0.59   0.555    -.4018243    .2160873
          3  |  -.1043377   .1519157    -0.69   0.493    -.4028365    .1941611
          4  |  -.1647405   .1412597    -1.17   0.244    -.4423013    .1128203
             |
healthboard |
NHS Lothian  |  -.1932517   .1189172    -1.63   0.105     -.426912    .0404085
NHS Greater Glasgow and Clyde  |  -.5926338   .2186798    -2.71   0.007    -1.022317   -.1629504
NHS Tayside  |  -.5838687   .2338494    -2.50   0.013    -1.043359   -.1243786
NHS Highland  |  -.2644738   .1597384    -1.66   0.098    -.5783435    .049396
NHS Fife  |  -.4052869   .2222047    -1.82   0.069    -.8418965    .0313226
             |
_cons  |   10.48133   .1293621    81.02   0.000     10.22715    10.73552
------------------------------------------------------------------------------
```

APPENDIX IV: Excerpt of the Stata code (SMR01 Costs)

```
set more off

adopath + "S:\ado_lib/"


cd "\\Farr-FS1\Study Data\1516-0194\Research\LauraBrunoDatasets\OrigData"

use SMR01, clear

merge m:1 patientid using StudyCohortsFinalORIG

drop if _merge==1
```

```
bys patientid (dadmis): gen m = _n
gen _aav = (case == 0)
label var _aav "Is this an AAV patient?"

gen _ancastatus=1 if ancastatus==1 | ancastatus==12 | ancastatus==13
replace _ancastatus=2 if ancastatus==2
replace _ancastatus=3 if ancastatus==3

gen neversmr01 = (_merge == 2)
label var neversmr01 "patients with zero SMR01 admissions"

keep if _aav ==1

drop _merge
order patientid _casenumber _aav

todate dadmis, gen(_dadmis) p(yyyymmdd) f(%td)

clonevar entrydate = _indexdateanalysis
label var entrydate "entry date - diagnosis date"

bys patientid contstay (_dadmis): gen admission = _n
replace admission = .a if admission != 1
replace admission = .a if neversmr01==1
label var admission "is this row an admission?"
```

```
sort patientid _dadmis

todate ddisch, gen(_ddisch) p(yyyymmdd) f(%td)


clonevar _los = los

replace _los=1 if _los==0


replace admission = .a if mi(admission)

replace entrydate = _indexdateanalysis if mi(entrydate)


gen yadmis= year(_dadmis)


bys patientid: gen markicu= 1 if sigfac=="13" | sigfac=="1H"

replace markicu=.a if markicu!=1

label var markicu "marker for icu episode"


gen iculos=_ddisch-_dadmis if markicu==1

replace admission=.a if markicu==1

replace iculos=1 if iculos==0 & markicu==1

replace _los=. if !mi(iculos)


replace admission=1 if contstay==contstay[_n-1] & markicu[_n-1]==1 & markicu!=1


bys patientid yadmis: gen yearmarker=_n


gen fup=(exitdate-entrydate)/365.25

label var fup "follow-up time in years"
```

```
gen pre_fup=1

forvalues j=1995/2017{
bys patientid yadmis (_dadmis): egen sumpreadmis`j'=sum(admission) if yadmis<year(entrydate) & yadmis==`j' & admission==1 & markicu!=1
replace sumpreadmis`j'=0 if mi(admission)
bys patientid yadmis (admission _dadmis): gen seqpreadmis`j'=_n
bys patientid yadmis: gen preadmis`j'=sumpreadmis`j' if seqpreadmis`j'==1 & sumpreadmis`j'!=0
}

forvalues j=1995/2017{
bys patientid yadmis (_dadmis): egen sumpostadmis`j'=sum(admission) if yadmis>=year(entrydate) & yadmis==`j'& admission==1 & markicu!=1
replace sumpostadmis`j'=0 if mi(admission)
bys patientid yadmis (admission _dadmis): gen seqpostadmis`j'=_n
bys patientid yadmis: gen postadmis`j'=sumpostadmis`j' if seqpostadmis`j'==1 & sumpostadmis`j'!=0
}

forvalues j=1995/2017{
bys patientid yadmis (_dadmis): egen sumprelos`j'=sum(_los) if yadmis<year(entrydate) & yadmis==`j' & !mi(_los) & markicu!=1
replace sumprelos`j'=0 if mi(_los)
bys patientid yadmis (admission _dadmis): gen seqprelos`j'=_n
bys patientid yadmis: gen prelos`j'=sumprelos`j' if seqprelos`j'==1 & sumprelos`j'!=0
}
```

```
forvalues j=1995/2017{
bys patientid yadmis (_dadmis): egen sumpostlos`j'=sum(_los) if yadmis>=year(entrydate) & yadmis==`j'& !mi(_los) & markicu!=1
replace sumpostlos`j'=0 if mi(_los)
bys patientid yadmis (admission _dadmis): gen seqpostlos`j'=_n
bys patientid yadmis: gen postlos`j'=sumpostlos`j' if seqpostlos`j'==1 & sumpostlos`j'!=0
}

forvalues j=1995/2017{
bys patientid yadmis (_dadmis): egen sumpreicu`j'=sum(markicu) if yadmis<year(entrydate) & yadmis==`j' & markicu==1
replace sumpreicu`j'=0 if mi(markicu)
bys patientid yadmis markicu (_dadmis): gen seqpreicu`j'=_n
bys patientid yadmis: gen preicu`j'=sumpreicu`j' if seqpreicu`j'==1 & sumpreicu`j'!=0
}

forvalues j=1995/2017{
bys patientid yadmis (_dadmis): egen sumposticu`j'=sum(markicu) if yadmis>=year(entrydate) & yadmis==`j'& markicu==1
replace sumposticu`j'=0 if mi(markicu)
bys patientid yadmis markicu (_dadmis): gen seqposticu`j'=_n
bys patientid yadmis: gen posticu`j'=sumposticu`j' if seqposticu`j'==1 & sumposticu`j'!=0
}

forvalues j=1995/2017{
```

```
bys patientid yadmis (_dadmis): egen sumpreiculos`j'=sum(iculos) if yadmis<year(entrydate) & yadmis==`j' & markicu==1

replace sumpreiculos`j'=0 if mi(iculos)

bys patientid yadmis markicu (_dadmis): gen seqpreiculos`j'=_n

bys patientid yadmis: gen preiculos`j'=sumpreiculos`j' if seqpreiculos`j'==1 & sumpreiculos`j'!=0

}

forvalues j=1995/2017{

bys patientid yadmis (_dadmis): egen sumposticulos`j'=sum(iculos) if yadmis>=year(entrydate) & yadmis==`j'& markicu==1

replace sumposticulos`j'=0 if mi(iculos)

bys patientid yadmis markicu (_dadmis): gen seqposticulos`j'=_n

bys patientid yadmis: gen posticulos`j'=sumposticulos`j' if seqposticulos`j'==1 & sumposticulos`j'!=0

}

drop seqpre* seqpost* sumpre* sumpost*

drop no matchedno m yearmarker

drop if entrydate<mdy(1,1,2000)

forvalues i=1995/2017 {

replace preadmis`i'=. if `i'-year(entrydate)<-1

replace postadmis`i'=. if `i'-year(entrydate)<-1

replace prelos`i'=. if `i'-year(entrydate)<-1

replace postlos`i'=. if `i'-year(entrydate)<-1
```

```
replace preicu`i'=. if `i'-year(entrydate)<-1
replace posticu`i'=. if `i'-year(entrydate)<-1
replace preiculos`i'=. if `i'-year(entrydate)<-1
replace posticulos`i'=. if `i'-year(entrydate)<-1
}

egen preadmis=rowtotal(preadmis1995-preadmis2017)
egen postadmis=rowtotal(postadmis1995-postadmis2017)
egen prelos=rowtotal(prelos1995-prelos2017)
egen postlos=rowtotal(postlos1995-postlos2017)
egen preicu=rowtotal(preicu1995-preicu2017)
egen posticu=rowtotal(posticu1995-posticu2017)
egen preiculos=rowtotal(preiculos1995-preiculos2017)
egen posticulos=rowtotal(posticulos1995-posticulos2017)

clonevar healthboard=_indexhealthboard

label var preadmis "total number of admissions prior to diagnosis"
label var postadmis "total number of admissions after diagnosis"
label var prelos "total length of stay on general ward prior to diagnosis"
label var postlos "total length of stay on general ward after diagnosis"
label var preicu "total number of icu episodes prior to diagnosis"
label var posticu "total number of icu episodes after diagnosis"
label var preiculos "total number length of stay on icu prior to diagnosis"
label var posticulos "total number length of stay on icu after diagnosis"
```

```
bys patientid: gen patmark=_n
replace pre_fup=1

forvalues i=1995/2017 {
replace preadmis`i'=. if `i'-year(entrydate)<-1
replace postadmis`i'=. if `i'-year(entrydate)<-1
replace prelos`i'=. if `i'-year(entrydate)<-1
replace postlos`i'=. if `i'-year(entrydate)<-1
replace preicu`i'=. if `i'-year(entrydate)<-1
replace posticu`i'=. if `i'-year(entrydate)<-1
replace preiculos`i'=. if `i'-year(entrydate)<-1
replace posticulos`i'=. if `i'-year(entrydate)<-1
}

cd "\\Farr-FS1\Study Data\1516-0194\Research\LauraBrunoDatasets\HCU Master"
save smr01_long_data, replace

collapse (sum) pre* post*, by(patientid _aav entrydate healthboard ageatindex sex aavtype _ancastatus fup inpatday)
order preadmis prelos preicu preiculos postadmis postlos posticu posticulos, last

cd "\\Farr-FS1\Study Data\1516-0194\Research\LauraBrunoDatasets\HCU Master"
merge m:1 patientid using simd00
keep if _merge==3
drop _merge
```

```
lab def sex 0 "male" 1"female"

lab def aavtype 1"GPA" 2"MPA" 3"EGPA"

lab def _ancastatus 1"PR3" 2"MPO" 3"Negative"

lab def simdquin00 1"Q1" 2"Q2" 3"Q3" 4"Q4" 5"Q5"

lab var simdquin "Scottish Index of Multiple Deprivation, 1 is most deprived"


replace pre_fup=1


forvalues i=1995/2017 {

replace preadmis`i'=. if `i'-year(entrydate)<-1

replace postadmis`i'=. if `i'-year(entrydate)<-1

replace prelos`i'=. if `i'-year(entrydate)<-1

replace postlos`i'=. if `i'-year(entrydate)<-1

replace preicu`i'=. if `i'-year(entrydate)<-1

replace posticu`i'=. if `i'-year(entrydate)<-1

replace preiculos`i'=. if `i'-year(entrydate)<-1

replace posticulos`i'=. if `i'-year(entrydate)<-1

}


save smr01_wide_cost_data, replace


import excel Final_Tariffs, sheet("Final Tariffs") firstrow clear

save tariffs, replace


use smr01_wide_cost_data, clear
```

```
merge m:1 patientid using tariffs
drop _merge

foreach var of varlist preadmis1995-posticulos2017 {
gen cost_`var'=.
}

foreach var of varlist cost_i_admis1995-cost_ae2017 {
gsort - patientid
replace `var'=`var'[1] if `var'==.
}

order preadmis prelos preicu preiculos postadmis postlos posticu posticulos toadmis tolos toicu toiculos sending_centre simdquin00 simddec00 _simddate c_age, last
order cost_i_admis1995-cost_ae2017, seq
order cost_ae1995-cost_nurseclinic2017, after(posticulos2017)
order preadmis1995-cost_nurseclinic2017, last
order cost_preadmis1995-cost_posticulos2017, last

forvalues j=1995/2017{
replace cost_prelos`j'=prelos`j'*cost_i_admis`j' if prelos!=0 & inpatday=="I"
replace cost_prelos`j'=prelos`j'*cost_d_admis`j' if prelos!=0 & inpatday=="D"
replace cost_preiculos`j'=preiculos`j'*cost_icu`j' if preiculos!=0
replace cost_postlos`j'=postlos`j'*cost_i_admis`j' if postlos!=0 & inpatday=="I"
replace cost_postlos`j'=postlos`j'*cost_d_admis`j' if postlos!=0 & inpatday=="D"
replace cost_posticulos`j'=posticulos`j'*cost_icu`j' if posticulos!=0
```

}

egen cost_prelos=rowtotal(cost_prelos1995-cost_prelos2017)

egen cost_preiculos=rowtotal(cost_preiculos1995-cost_preiculos2017)

egen cost_postlos=rowtotal(cost_postlos1995-cost_postlos2017)

egen cost_posticulos=rowtotal(cost_posticulos1995-cost_posticulos2017)

collapse(sum)cost_prelos cost_postlos cost_preiculos cost_posticulos, by(patientid fup pre_fup _aav sex ageatindex aavtype _ancastatus healthboard)

save smr01_cost, replace

STATURORY DECLARATION

I herewith declare that I have authored the present Master thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources

_____ _____

Date Signature

DECLARATION OF CONSENT

I herewith declare my consent regarding the publication of the present Master thesis in the library of the department Life Sciences of the University of Applied Sciences Hamburg. The rights of third parties will be preserved.

_____ _____

Date Signature