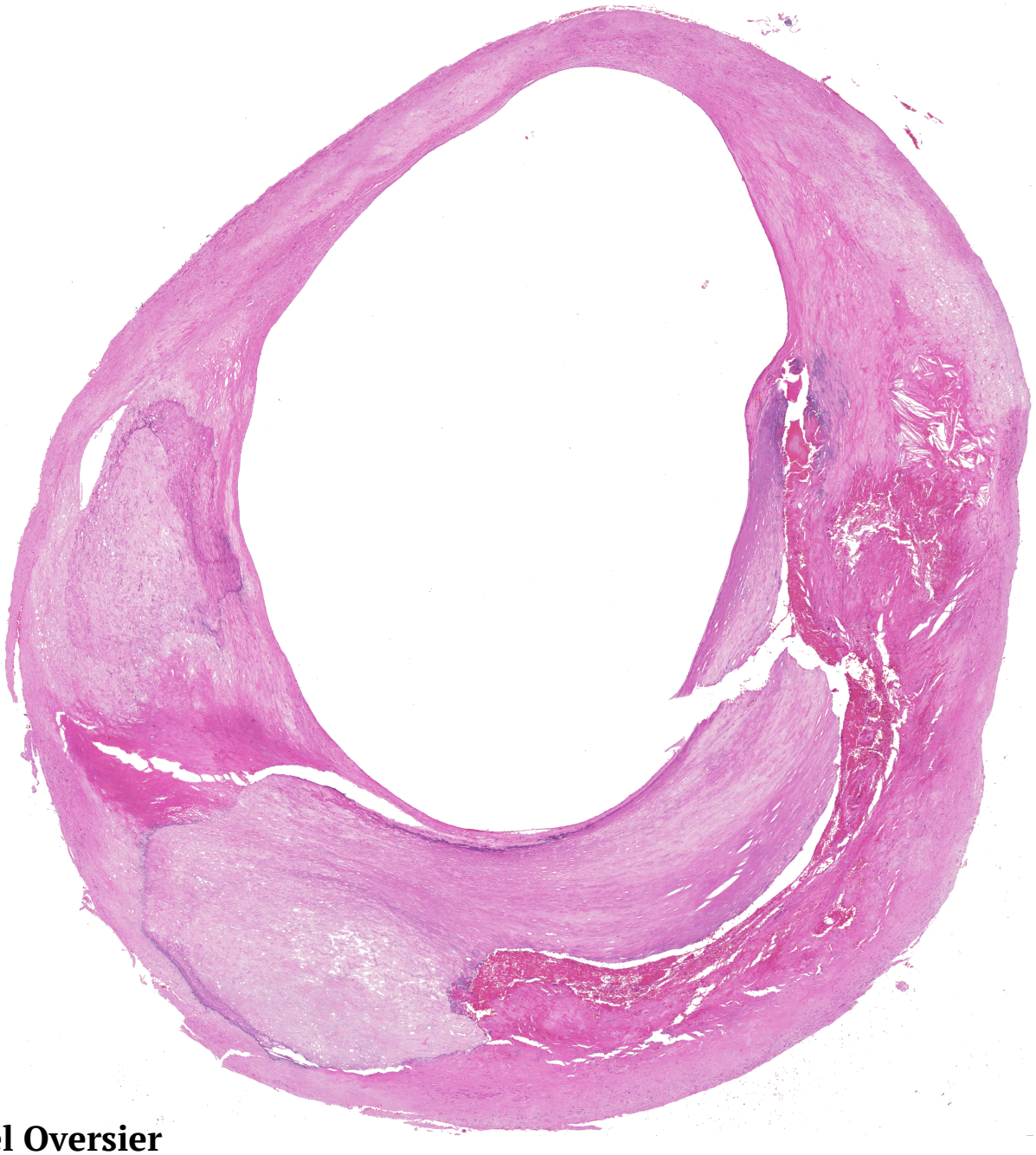
Local atherosclerotic plaque composition and plaque rupture

An explorative study

MSc Thesis in Biomedical Engineering, Technical University Delft



Rachel Oversier
4722515
Technical University Delft
MSc Biomedical Engineering

TU Delft

Dimitra Dodou
Technical University Delft
Associate Professor

Ali Akyildiz
Erasmus Medical Centre
Assistant Professor

Su Güvenir
Erasmus Medical Centre
PhD student

Margreet de Vries
Leiden University Medical Centre
Assistent professor

Graduation report

Local atherosclerotic plaque composition and plaque rupture

An explorative study

By

Rachel Marianne Oversier

in partial fulfilment of the requirements for the degree of

Master of Science
in Biomedical Engineering

at the Delft University of Technology,
to be defended publicly on Friday, December 18, 2020 at 15:30

Student number:	4722515	
Supervisors:	Dr.ir. A.C. Akyildiz	Erasmus Medical Centre
	ir. S. Guvenir	Erasmus Medical Centre
	Dr. D. Dodou	Technical University Delft
Co-supervisor:	Dr. M.R. de Vries	Leiden University Medical Centre
Thesis committee:	Dr. D. Dodou	Technical University Delft
	Dr.ir. A.C. Akyildiz	Erasmus Medical Centre
	Dr.ir. S. Iskander-Rizk	Technical University Delft
	Dr.ir. Mirzaali Mazandarani	Technical University Delft

An electronic version of this thesis is available at https://repository.tudelft.nl/.

TU Delft

Erasmus MC
Universitair Medisch Centrum Rotterdam

LUMC Leids Universitair Medisch Centrum

Acknowledgements

Atherosclerosis is the main cause of cardiovascular diseases, whereas cardiovascular diseases lead to one-third of mortality worldwide. Arteriosclerosis strikes people from all countries, independent of their level of income. The subject of this thesis spoke to me, because of the enormous impact of atherosclerosis on global health. I am a Biomedical Engineering student at the TU Delft, as well as a medical student at the Leiden University. To look at this disease from both a medical and a technical perspective was very insightful.

During the graduation process I gained and improved skills, including the interpretation of H&E stained atherosclerotic plaque sections, research design (translating visual data to quantitative data and statistical methods for the analysis of clustered data), project management and scripting. Next to practical skills, this thesis has given me the opportunity for personal growth in perseverance, stress management and self-esteem in general.

I am very grateful that the Biomedical Engineering department offered me a position as graduation intern. There are some people I'd like to thank specifically.

First of all, I would like to thank my daily supervisors Dr. ir. A.C. Akyildiz and S. Güvenir of the Erasmus Medical Centre. Thank you both for overseeing the thesis process, the many detailed reviews, your positivity and flexibility. Ali, thank you for the creative out-of-the box ideas and for the support when I needed it. Su, thank you for your tips on making a story more fluent.

Furthermore, many thanks to my supervisor Dr. D. Dodou of the Technical University Delft. I really appreciated your flexibility and help with the adjustment of the thesis topic due to the COVID-19 crisis. You have also been a big help in the translation of theoretical ideas to practical implementation. Thank you for always being available, your clear advices and the feedback on this project.

I would like to thank my co-supervisor Dr. M.R. de Vries from the Leiden University Medical Centre for thinking along in this project, reviewing the report and for making sure the thesis fits the requirements of the research internship of my medical study.

Additionally, I would like to express my gratitude towards my friend MD X.M. Matthijssen of the Leiden University Medical Centre. Thank you for your giving me confidence in this thesis and for your input on the statistical part. I am grateful for your always quick and clear answers on my many statistics-related questions.

Moreover, thanks to K. Kuijt-van Gaalen and Dr.ir. S.A. Korteland of the Erasmus Medical Centre for the help with the MeVisLab module. Kim, I really appreciated the many hours of supervision on the segmentation of the histology sections and your emotional support. Suze-Anne, thank you for responding quickly to my programming-related questions and for the original Python code for segmentation, which I could adjust to fit the purpose of this thesis.

Thanks to my friend Caroline: with our similar personality traits, similar medical/BME master situation and similar Duo deadline, there was no one that understood my troubles as well as you did. I hope I can be as much of an added value to you as you are to me in this graduation process.

I would like to thank my parents and my grandparents for their endless support and for always having my best interest at heart. Thank you for believing in me.

Adriana, I could not have done this thesis without you. I cannot express how much you've helped me on the technical level, as well as on the personal level. Thank you for always listening, for all of your advice and for distracting me when necessary.

Abstract

Introduction: Carotid atherosclerosis is characterized by the build-up of the so-called plaque tissue in the wall of the carotid artery. The rupture of carotid plaques can lead to acute manifestations, including ischemic stroke and transient ischaemic attack (TIA). The mechanism responsible for rupture is not yet unravelled, complicating the medical management of the disease. There are indications that plaque morphology plays a role in plaque rupture risk. Previous histopathological studies found that ruptured plaques and non-ruptured plaques tend to differ in their structural composition. These studies did not perform a detailed analysis of the local plaque features at the rupture sites, even though this might increase the insight in the rupture event. The aim of this thesis was to investigate the relation between local plaque composition and rupture in carotid plaques.

Methods: Digitized histological sections of 21 ruptured carotid endarterectomy samples (obtained from fourteen different patients) were segmented for the following nine tissue structures: *cholesterol crystals, plaque, calcium*, *necrotic core*, *fresh haemorrhage*, *foam cells*, *neovessels*, *macrophages* and *old haemorrhage*. The rupture site on the sections was identified as a luminal disruption in the presence of red blood cells. The digitized sections were then divided into 8 bins (pie slices), resulting in a total of 168 bins (21 sections x eight bins). Of these observations, 23 contained rupture, whereas the other 145 were rupture-free. The plaque area in each bin was measured and the other tissue structure sizes were calculated relative to the plaque area in the bin. Spatial heatmaps demonstrated the location of the tissue structures relative to the rupture site. They were created for the three variables *cholesterol crystals*, *calcium* and *macrophages,* as these tissue structures were often recognized close to the rupture sites. Kendall Tau-b correlation coefficients were calculated to determine the relation between tissue structures in close proximity. The two-sided clustered Wilcoxon rank sum test (with clusters at the patient level) was used to investigate differences between the composition of ruptured and non-ruptured bins. Additionally, the association of tissue structures with rupture was tested by a generalized estimating equations (GEE) logistic regression model (working correlation matrix: independence). Clusters were assigned at the patient level. A sensitivity analysis was carried out (using 2-, 4- and 16-bins) to determine the effect of the bin size on the outcome of the statistical analyses.

Results: Visual analysis demonstrated that ruptures tended to extend to (or start at) areas of calcification or cholesterol crystals and that the rupture sites could be markedly inflamed. The spatial heatmaps showed that areas of calcification were closer to the rupture sites than areas with cholesterol crystals. Moreover, the prevalence of macrophages near the rupture sites was lower when compared to the prevalence of calcium and cholesterol crystals. In ruptured bins, the strongest tissue structure correlation was found between *calcium* and *cholesterol crystals* (τ_b= -0.44, $p \leq 0.05$). This correlation was less evident when ruptured and non-ruptured bins were analysed together (τ_b= -0.07, $p > 0.05$). The clustered Wilcoxon rank sum test showed that *calcium* (p=0.05) and *fresh haemorrhage* ($p=2.17 \times 10^{-11}$) tended to take on higher values in the ruptured bins, when compared to the non-ruptured bins. Lastly, with the GEE logistic regression model an association was found between rupture and the predictor variables *calcium* (p= 0.00094) and *plaque* (p=0.049). The odds ratio of rupture was 1.05 [1.02-1.07] for a plaque size increase of 0.34 mm^2, if the other plaque characteristics in the bin were held fixed. The odds ratios of rupture belonging to plaque-bins that contained 10% *calcium* and 50% *calcium* were 1.96 [1.60-2.40] and 3.01 [2.16-4.20], respectively, relative to plaque-bins with 0% *calcium* content and equal other plaque characteristics in the bin. The sensitivity analysis showed that the choice for an 8-bin approach was reasonable.

Conclusion. The rupture sites of the carotid sections showed an enhanced calcium content. Moreover, the calcium content and the plaque size showed to be significantly associated with plaque rupture. The developmental progress of non-invasive imaging techniques, using magnetic resonance imaging and ultrasound, is promising for the characterization of carotid plaques *in vivo*. Understanding the relation between plaque composition and plaque rupture at a local level can contribute to rupture risk assessment and eventually to the medical management of atherosclerosis.

Table of contents

Introduction

Atherosclerosis is the main cause of cardiovascular diseases and leads to a major burden on health worldwide.[1] In atherosclerosis the arteries are locally thickened by the build-up of so-called plaque tissue in the presence of chronic inflammation (Appendix A). The disease commonly occurs in the carotid area, where the proximal internal carotid artery and the common carotid bifurcation are most often involved.[2] The stenosis of an artery can lead to a reduced blood supply to the downstream tissues and organs and thereby cause a deteriorating health problem.[3] Acute manifestations of the disease can arise if the atherosclerotic plaque ruptures. The rupture of a plaque causes the highly thrombogenic plaque content to come in contact with the blood stream. By the activation of the coagulation cascade, a thrombus forms. This possibly leads to an ischaemic stroke or a transient ischaemic attack (TIA), if the thrombus originates from the carotid area.[4]

Currently medical management planning of carotid atherosclerotic disease is based on the stenosis degree of the artery and the presence of symptoms (ischaemic events in medical history).[2,5] However, research has indicated that the rupture risk of the plaque is not necessarily associated to stenosis degree or symptomatology.[6] The disease treatment is established by either intensive medical treatment or invasive revascularization procedures, such as the placement of a carotid stent and endarterectomy operations.[2,5] These procedures relief stenosis-related symptoms and can reduce the risk of stroke in the long term, but there is a risk of peri- and post-operative complications.[7,8] A need exists for better plaque rupture risk stratification in order to prevent under- and over-treatment.[6]

Even though plaque rupture is the result of a complex interplay between biochemical, systemic and mechanical factors,[9,10] the event itself can be seen as a mechanical failure.[4] The mechanical loading of the artery (e.g. by the blood pressure) induces stresses and strains in the arterial wall and the plaque. The magnitude and distribution of these stresses and strains are determined by the loading conditions, geometry and material properties.[11,12] The knowledge about the failure mechanism responsible for plaque rupture is still limited. No consensus has been reached about the underlying mechanical failure criterion. Proposed criteria include that rupture occurs if the local stress in the plaque exceeds the local strength[13–15], or by the excess of a defined failure strain[16]. Previous research has provided some evidence for the potential impact of plaque composition on plaque biomechanics, and as such on rupture risk.[10]

Arteriosclerotic plaques are heterogenous in nature and consist of different tissue structures.[17] Plaques can, for example, contain a large necrotic core with cholesterol crystals that is covered by a fibrous cap. They can also have a necrotic core that is calcified, and many other variations exist (Appendix A).[18] It is known that the morphological components of the plaque play an important role in the stress and strain magnitude and distribution within the plaque.[13] Recent studies show that the material properties of the plaque components are not

the same. A mismatch between these material properties can lead to local stress concentrations.[17,19]

Finite element analysis (FEA) of plaque biomechanics indicates that the peak (cap) stress experienced in atherosclerotic plaques is influenced by the fibrous cap thickness[12,20–22], necrotic core size and thickness [22–26] and the presence of calcifications[27,28]. The peak (cap) stress drastically increased if the fibrous cap thickness was reduced by decreasing the stenosis severity without the adjustment of other parameters.[12,21] The necrotic core size and thickness were positively related to the peak (cap) stress.[24–26] The necrotic core thickness was more important in the evaluation of plaque stability.[26] Inconclusive results were found on the effect of calcification on the maximum principal stress experienced in the plaque. A study by Huang et al. (2001) showed this stress was not significantly different if parts of fibrous tissue were replaced with calcification.[25] Contrarily, Teng et al. (2014) found that there was an increase in the maximum principal stress with increasing calcification. However, the stress decreased again after calcification was more extensive.[27] The location of calcification appeared to be an important factor in plaque stability.[28,29]

Mechanical experiments underline the importance of plaque morphology in plaque biomechanics. Uniaxial tensile tests showed that the failure stress is significantly related to the calcification pattern[11], the calcium-to-lipid ratio[30–32] and the macrophage density.[33]

Histology is the gold standard for the characterisation of atherosclerotic plaque composition.[34,35] Histological studies compared sections of ruptured and intact plaques from varying arterial territories. They found that the plaque composition differed between the two groups. The histology findings are in line with the FEA results. Multiple studies reported that ruptured plaques tend to have a thinner fibrous cap[10,36–38] and a larger necrotic core.[37–40] Again, the role of calcification was more uncertain. Ruptured plaques showed a significantly increased or comparable calcification when compared to intact plaques.[37,39] Mauriello et al. (2011) found that a higher calcification area was related to the presence of minute plaque disruption, while a smaller calcification area was associated with a full cap rupture.[41]

Moreover, based on histological studies it was concluded that ruptured plaques tend to have an increased cholesterol crystal content,[39] an increased overall plaque inflammation (presence of macrophages and/or lymphocytes)[36,42], as well as locally in the fibrous cap[36,43], and an increased overall neovessel density when compared to non-ruptured plaques.[44,45] Different hypothesis exist about the underlying mechanisms by which these factors influence plaque stability. Abela & Aziz (2006) reported that *in-vitro* cholesterol crystallization leads to volume expansion. They suggested that this event could have implications for plaque stability: the force of crystal growth can damage a biological membrane and a crystal tip can puncture a membrane.[46] Cholesterol crystals that perforated the intima and plaque surfaces were found to surround the location of plaque rupture in a study performed by Abela et al (2009).[47] Inflammatory processes trigger vascular smooth muscle cell (VSMC) apoptosis, whereas plaque rupture occurs in regions with VMSC depletion.[48,49] Macrophages produce matrix

metalloproteinases (MMPs). MMPs can influence plaque stability by stimulating the breakdown of matrix proteins (including collagen) and promoting neovascularization. In general pro-inflammatory conditions stimulate MMP expression.[50] The neovessel density is shown to be positively correlated to the atherosclerotic stage[51], the presence of an inflammatory infiltrate[52], the deposition of free cholesterol and the enlargement of the necrotic core. Therefore, neovessels might influence plaque stability.[53]

The contribution of the histology-based studies to the understanding of plaque rupture at the local level is not evident, since all studies compared plaque composition at the section level. This thesis investigates the association between the local plaque composition and rupture in carotid plaques. The analysed dataset included 21 digitized histological sections of ruptured carotid endarterectomy samples, obtained from fourteen patients. Plaque rupture was defined as a luminal surface disruption, accompanied by the presence red blood cells. Manual segmentation of the slides was carried out in order to determine the plaque composition. This was followed by the division of the sections into eight bins (pie slices). Statistical methods, suitable for clustered data, were used to compare the composition of the plaque locally around the plaque rupture site to the composition further away from the rupture and to find out whether the presence of certain tissue structures was associated with plaque rupture. The clinical translation of this study becomes clear with the development of (non-invasive) visualisation techniques that can determine the plaque composition *in vivo*.

Methods

The Seattle dataset

The histological images of carotid plaques used in this study represented a subset of a dataset belonging to the Washington University in Seattle. The Seattle dataset contains high resolution images of histological sections from 244 carotid plaques obtained from 244 patients.[54] The patients underwent a carotid endarterectomy operation. They were either symptomatic with a stenosis degree of more than 50% (i.e. had a TIA or stroke ipsilateral to the index artery within 6 months of the surgery) or they were asymptomatic with a stenosis degree of more than 80%. During surgery, the carotid plaques were approached by a longitudinal cut through the adventitia and the outer media. An endarterectomy plane around the plaque was created and the plaque was removed as a whole, without assaulting the plaque. In this way, the lumen surface remained undisturbed. After removal, the plaques were flushed with Ringer's lactate and bathed in a 10% neutral buffer formalin for at least 24 hours. Then, the specimen were decalcified in a 10% formic acid, followed by further processing and paraffin embedding. The blocks were sectioned at a thickness of 10μm and the sections were stained either with haematoxylin and eosin (H&E) or Mallory's Trichrome. The histological slides enveloping the plaque sections were microscopically photographed and all images were digitized.[54]

The available subset

A subset of the Seattle dataset was randomly made available to students. The subset contained the digitized images of 58 histological sections from carotid plaques of 22 patients, which each contributed one up to six sections. The subset was valuable, as the images showed high-quality sections with grossly intact and undistorted plaque tissue. Additionally, the rarity of plaque rupture was visible on some of these sections.[54] The images had a high resolution and each pixel corresponded to an area of 3.40 μm^2.

All sections were stained with H&E, a staining commonly applied in histology and used for the interpretation of a variety of tissue structures.[55] Eosin binds to positively charged complexes, colouring them in differing shades of red and pink. Hence, eosin can be used to differentiate between cell types (by staining its cytoplasm) and connective tissue matrices. Haematoxylin provides the cell nuclei, as well as calcium deposits with a black-blue colour.[56]

Manual segmentation of the digitized sections

Segmentation refers to the process of partitioning the image into multiple segments that share the same characteristics.[57] For this thesis, the histological images of all 58 sections were manually segmented for the following tissue structures: *cholesterol crystals, plaque, calcium, necrotic core, fresh haemorrhage, foam cells, neovessels, macrophages* and *old haemorrhage*. These structures were chosen, because they are all reasonably recognizable in a H&E stained section. Figure 1 shows an example of all different tissue structures. The area between the outer wall of the endarterectomy section and the lumen was segmented as *plaque*. *Cholesterol*

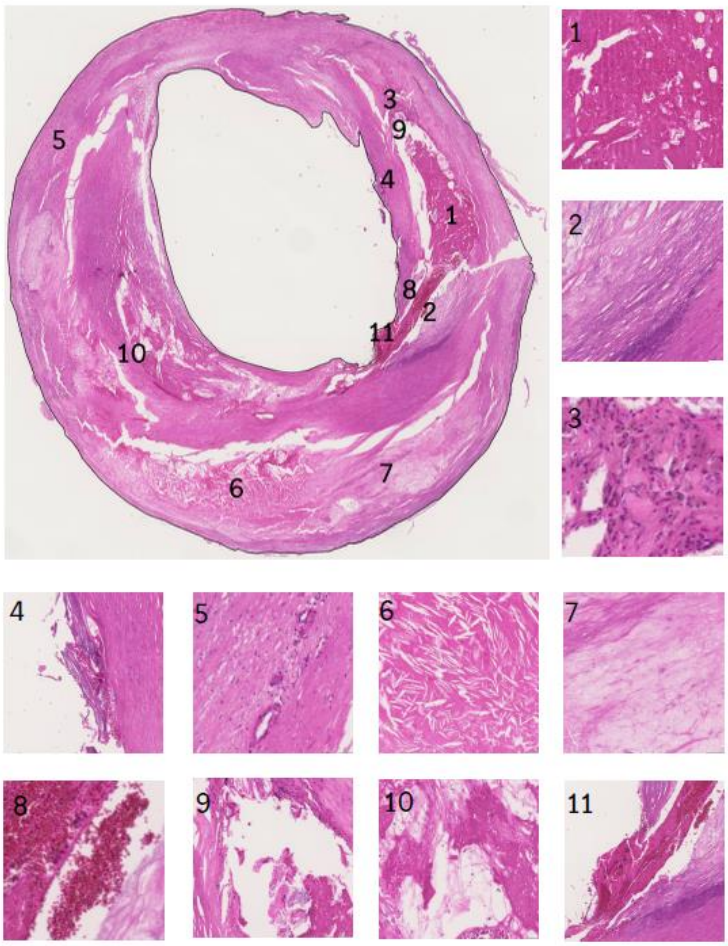


Figure 1: Histological slide showing plaque rupture. Black lines show the borders of the plaque. The different tissue structures that can be characterized include: 1. Old hemorrhage, 2. Calcium, 3. Macrophages, 4. Rupture (in calcified tissue), 5. Neovessels, 6. Cholesterol crystals, 7. Necrotic core (extracellular lipid), 8. Fresh hemorrhage (individual red blood cells are visible), 9. Artefact (tissue has disappeared), 10. Foam cells, 11. Plaque rupture (in calcified tissue).

crystals dissolve during tissue processing with lipid solvents and leave characteristic clefts. These needle-like structures can be observed with routine H&E staining.[46,47] After decalcification part of the *calcium* disappears. However, traces of mineral deposits are still observable by their black-blue colour when stained with H&E.[58,59] *Necrotic cores* are acellular necrotic regions that appear pale on the H&E stained section compared to the healthy, fibre-rich tissue.[60] By this definition, the areas that contained accumulated extracellular lipid were also segmented as necrotic core. These areas were not necessarily covered with a fibrous cap and might better fit the definition of "lipid pool" in Appendix A. *Fresh haemorrhage* can be recognized by the presence of individual red blood cells (RBC), which stain intensely red by the eosin dye.[61] *Foam cells* can be distinguished from other cells on the H&E stained section by their large size and by the presence of vacuolized cytoplasm.[62,63] If a *neovessel* is sliced transversely, its structure can be recognized by an approximately circular or ellipse-formed hole (sometimes filled with RBC). The hole is surrounded by flat endothelial cells of which the nuclei are coloured blue by the haematoxylin dye.[64] The recognition of the different types of

immune cells was difficult due to the low magnification of the digitized histological sections in the database, as well as to the absence of specific immune cell markers. Inflammatory infiltrates of carotid plaque tissue can include T-lymphocytes, B-lymphocytes, mast cells, dendritic cells, but mainly consist of macrophages.[65–67] Therefore, the term *macrophages* was used to indicate areas with marked inflammation. These areas showed accumulations of blue-coloured nuclei. *Old haemorrhage* is characterised by an amorphous region that stains intensely pink on a H&E stained sections.[68] Individual RBC are not recognisable anymore. Regions were segmented as *artefact* when tissue had disappeared and when it was unclear which tissue structure was originally present. *Ruptures* were segmented merely at the site of luminal discontinuation. For the segmentation an in-house developed software tool was used (MeVisLab 2.7.1, MeVis Medical Solutions, Bremen, Germany).

The segmentation of the first five histological images was confirmed by the vascular histopathology expert. The images belonged to the sections of five different patients. Additionally, complex cases where discussed with the expert and consensus was reached for these cases.

Selection of the ruptured sections

The digitized histological sections were divided into two categories: sections with plaque rupture and sections without plaque rupture. Two definitions of plaque rupture are:

(1) There is a discontinuation/disruption of the tissue surrounding the lumen. Additionally there is fresh haemorrhage visible somewhere along the rupture path.
(2) The tissue surrounding the lumen is still intact by one or two fibres, but the cavity beneath these fibres is filled with red blood cells. In these cases, it is plausible that the "definition 1" rupture was "just missed" due to the section thickness (10µm).

In total, 31 out of the 58 sections showed plaque rupture. In three cases the sections exhibited multiple plaque ruptures. Plaque rupture is a 3D process and therefore the same ruptures could be visible in multiple adjacent sections. In order to prevent "double observations" in the data, for each rupture the most representative section was chosen. This was the section that showed the rupture path connected to the lumen most clearly. For example, if the same rupture was categorized by definition 1 on one section and by definition 2 on an adjacent section, the section with definition 1 was preferred for further analysis.

The categorization of the digitized sections, as well as the choice for the most representative section was confirmed by the vascular histopathology expert. It was agreed upon that the dataset exhibited 23 "original ruptures" that were seen on 21 sections, which were obtained from fourteen patients. Each patient contributed one up to four ruptured sections. These 21 digitized ruptured sections were selected for further analysis. It should be noted that two of these sections also showed ruptures that were already counted as an "original rupture" on another included section. These "repeated ruptures" were ignored. A flow diagram of the section selection can be found in Figure 2.

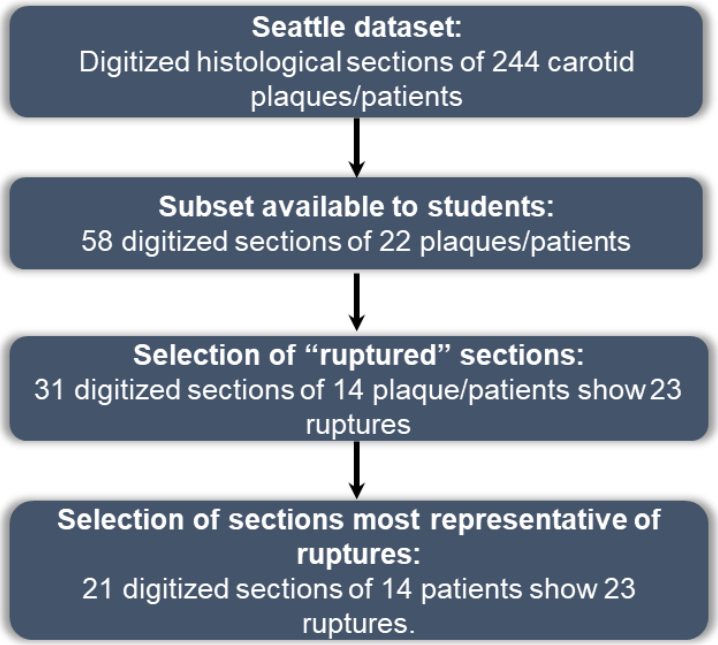


Figure 2: Flow diagram of the selection of the digitized ruptured sections. The lower block presents the set that was used for further analysis in this thesis.

Extraction of the tissue structures and binning procedure

The output of the software tool used for segmentation was a NIFTII file. This file contained the same number of pixels as the original histological image. By the segmentation of the different tissue structures, a value was assigned to each pixel. The 21 NIFTII files of the "representative" ruptured sections were loaded into Python (version 3.7, Python Software Foundation, https://www.python.org/). The scripts used for data processing can be found in Appendix D1 and were based on a script that was already developed in the department. The binary masks of the different tissue structures were extracted from the pixel values. For every tissue structure, one binary mask was created per section. The pixels in these masks could take on either a value of 1, indicating the presence, or 0, indicating the absence of the tissue structure. Areas (consisting out of pixels) could contain multiple tissue structures, as the other tissue structures were always segmented as *plaque* too.

The next step included the division of the segmentation into 32 bins (pie slices), visualised in Figure 3. Binning was performed by the assignment of a bin group number to each pixel. The formation of the bins originated at the centre point of the lumen. The number of degrees per bin equalled 360/32 and the first bin was aligned along the negative x-axis. Consequently, the number of tissue structure pixels was determined per bin. The output of the Python script was an excel sheet with 32 observations (bins) that noted the area of the tissue structure *plaque* in the bin, expressed as the number of pixels. The other tissue structures, including *cholesterol crystals*, *calcium*, *necrotic core*, *fresh haemorrhage*, *foam cells*, *neovessels*, *macrophages* and *old haemorrhage*, were expressed as the percentage relative to the *plaque* area in the corresponding bin. For the statistical analysis, the 32 bins were combined to form eight bins. Four adjacent bins formed one new bin and the rupture was located in the centre of one of the

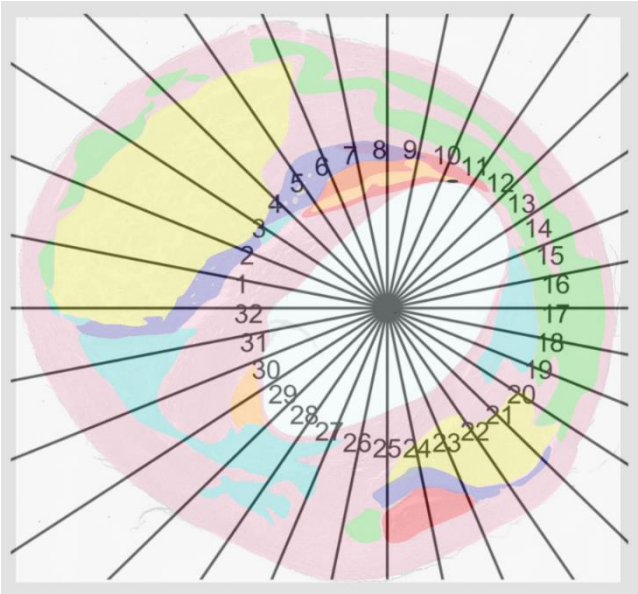


Figure 3. The 32 bins of a segmented histological slide. The first bin is located along the negative x-axis. In this slide, the rupture is visible as a black spot in bin 11. For the 16-bin analysis, bin 10 and 11 were combined to represent the ruptured bin. For the 8-, 4- and 2-bin analysis, bin 9-12, bin 7-14 and bin 3-18 were combined, respectively, to represent the ruptured bin.

new bins (Figure 3). The *plaque* areas in the four bins were added up. The scores of the other tissues structures were recalculated and therefore still represented the percentage of the *plaque* area occupied by the concerning tissue structure. In the end, 168 observations/bins (21 slides x 8 bins per section) were obtained. Of these bins, 23 showed rupture and 145 did not show rupture.

Spatial analysis

Spatial heat maps of plaque rupture were generated in MATLAB (Version 9.5 update 2, The MathWorks, Inc., Natick, Massachusetts, United States), in order to enhance the insight in the tissue structures surrounding the rupture. The heatmaps were created for the tissue structures *plaque*, *cholesterol crystals*, *calcium, macrophages* and *necrotic core. Plaque* was chosen to familiarize the reader with the concept of a spatial heat map. The tissue structures *cholesterol crystals* and *calcium* were selected, because visual inspection of the digitized sections revealed that the majority of the ruptures extended to either an area of cholesterol crystals or calcium (Histology Atlas). Rupture sites sometimes showed a marked invasion of *macrophages*, or broader, inflammatory cells (Histology Atlas). These observations were confirmed with the vascular histopathology expert. The heatmap for *necrotic core* can be found in Appendix C1.

A code available via a GitHub repository (https://github.com/kwikteam/npy-matlab) was able to load the Python arrays, i.e. the 21 binary masks of the tissue structures, and convert them into MATLAB matrices.[69] In these masks, as stated before, a 1 indicated the presence of a tissue structure at the corresponding location, while a 0 indicated absence.

A circle with the rupture (at the luminal site) as the centre point and a fixed radius of 1300 pixels (i.e. 2.40 mm) was cut out from the binary masks of the tissue structures (Appendix D2). There were 23 ruptures in the 21 histological sections, so 23 cropped masks were obtained per tissue structure. The circular masks were rotated so that a vertical line could be drawn between the rupture and the centre of the lumen. The lumen was on the lower side of the circle. The masks were similarly oriented due to this rotation, increasing the interpretability of the heatmap. Figure 4 visualizes the described procedure.

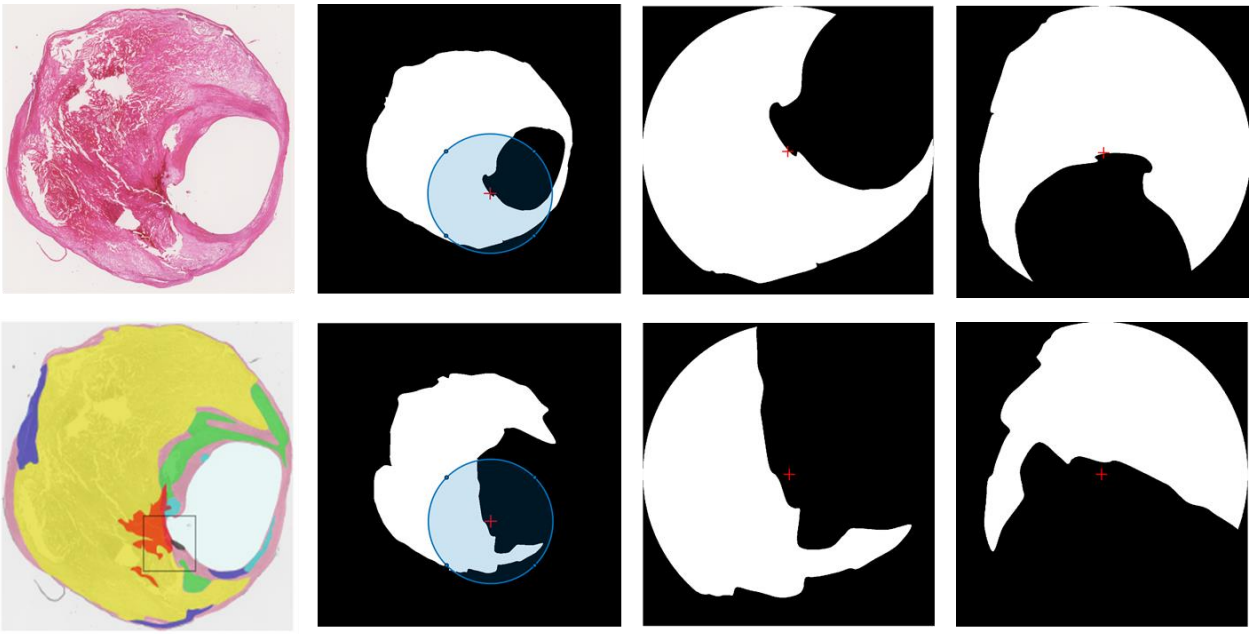


Figure 4. Selection of the region of interest in the binary *plaque* mask (upper row) and the *cholesterol crystal* mask (lower row). The left upper image shows a histology section and the image below shows its segmentation (the area with cholesterol crystals colored yellow). The red arrows in the different images indicate the rupture site. The last images in the upper and lower row show that the cropped masks are rotated.

The heatmaps of the tissue structures *plaque*, *cholesterol crystals*, *calcium, macrophages* and *necrotic core* were created by the summation of the 23 rotated circular masks. The pixel values of the heat map represented the prevalence of the concerning tissue structure at the corresponding location. The maximum possible pixel value equalled 23.

Statistical analysis

Statistical techniques were used to increase the understanding of plaque rupture at the local level. The general relation between tissue structures in close proximity was examined by investigating correlations at the bin-level. By zooming in on the correlations found in the ruptured bins only, divergencies could be revealed. Additionally, the local composition of the plaque at the rupture site and further away from the rupture site was compared. This was obtained by the detection of compositional differences between the ruptured bins and the non-ruptured bins. Furthermore, it was tested whether certain tissue structures were associated with plaque rupture.

Appendix B describes the statistical procedures and the mathematical equations behind the statistical analyses that were used. All statistical analyses were carried out in RStudio (Version 1.3.1056, R Foundation for Statistical Computing, Vienna, Austria) with the following packages: "psych", "stats", "ggplot2", "geepack" and "car" (Appendix D3). A p-value ≤ 0.05 was considered statistically significant.

RELATIONS BETWEEN TISSUE STRUCTURES IN CLOSE PROXIMITY

In order to investigate the relation between two tissue structures in close proximity at the bin level and more specifically in ruptured bins, the Kendall Tau-b correlation coefficient was used (Appendix B1). This non-parametric measure calculates the strength of the association. It depends on the number of concordances and disconcordances in the data and can effectively deal with ties.[70] The outcome is a value between -1 and 1. A positive relation shows that the ranks of both tissue structures increase together. A negative implies that the rank of one tissue structure increases with the decrease of the rank of the other tissue structure.[71] The Kendall Tau-b correlation coefficients were obtained by the use of the cor-function.[72]

COMPOSITIONAL DIFFERENCES BETWEEN RUPTURED BINS AND NON-RUPTURED BINS

The distributions of the tissue structures in the ruptured bins (group 1) and the non-ruptured bins (group 2) were compared. A clustered Wilcoxon rank sum test was used for this purpose (Appendix B2.2). This test is a variation on the Wilcoxon rank sum test that assumes independency between the groups (Appendix B2.1).[73,74] The independency assumption is not validated in the dataset used in this thesis. The digitized sections obtained from the same patient, as well as the 8 bins of a section, are not independent. The clustered Wilcoxon rank sum test was proposed by Rosner et al. (2006) and takes into account the dependency between the two groups.[75] The clusters were assigned at the patient level and the bins formed the subunits. The cluster size, i.e. the number of bins per patient, is allowed to vary. The test assumes an exchangeable relation between the ranks of the tissue structures in the subunits of each cluster. Thus, it assumes a similar correlation between the ranks of the tissue structures in the bins obtained from the same patient.[76]

The two-sided clustered Wilcoxon rank sum test was carried out for all tissue structures (i.e. *cholesterol crystals, plaque, calcium*, *necrotic core*, *neovessels*, *macrophages*, *fresh haemorrhage, foam cells* and *old haemorrhage*). The tests were executed by the clusWilcox.test-function.[77] The hypotheses that were tested were:[75,76]

H_0: The cumulative distribution functions of the tissue structure in the ruptured bin group and the non-ruptured bin group are the same
H_1: The cumulative distribution functions of the tissue structure in the ruptured bin group and non-ruptured bin group are not the same.

ASSOCIATION OF TISSUE STRUCTURES WITH PLAQUE RUPTURE

In order to find out whether the tissue structures were significantly associated with plaque rupture, the GEE method was applied. The GEE method can be used as an extension to generalized linear models (GLM) and is suitable for the analysis of clustered data.[78,79] There is

a dependency between the bins of the same section and the sections of the same patient in the dataset. In this thesis, the GEE method was used in combination with a logistic regression model (Appendix B4.2). The tissue structures were the independent (predictor) variables and rupture was the dependent variable. The clusters were assigned at the patient level and the bins formed the subunits.

The GEE logistic regression model gives an estimation of the regression parameters under weak assumptions of a working correlation matrix.[78] The working correlation matrix defines the relationship between the subunits of a cluster. The matrices that can be assigned include: independence, exchangeable, autoregressive, unstructured and user-defined (Appendix B4.2). The exchangeable working correlation matrix was chosen, since it assumes a similar correlation between all subunits within a cluster.[79] Hereby, the model took into account differences in systemic risk factors that vary between individuals. It is, however, also known that atherosclerotic plaques are highly heterogenous.[80] Therefore, the independence working correlation was tried as well. The independence working correlation matrix treats all observations as independent and in this case reduces to a standard logistic regression (Appendix B4.1). The fit of the models was compared with a measure called the quasi-likelihood under the independence model criterion (QIC). The QIC is based on the Akaike Information Criterion (AIC), but uses the quasi-likelihood instead of likelihood and the penalty term is modified (Appendix B4.2).[81] A lower QIC indicates a better fit of the model.[82]

The predictor variables (except *plaque*) were transformed by a $log(x + 1)$ function for the analysis (Appendix C2.1). In this way, the GEE logistic regression model better fitted the linearity assumption (Appendix C2.3). Additionally, a decreased number of predictor variables was entered into the model. There is a chance of model overfitting if the number of events is small compared to the number of explanatory variables.[83] A principal component analysis was carried out to explore the possibility of variable reduction by the use of principal components, but the results were not satisfactory (Appendix B3&C2.2). The parameters *cholesterol crystals*, *calcium*, *necrotic core*, *neovessels* and *macrophages* were included in the model. Available literature (histology-based studies and FEA research) indicated that these parameters might play a role in plaque stability.[12,20,36–45,21,47,84,22–28] *Plaque* was also included, because it stands to reason that the amount of pathological tissue could be associated to plaque rupture. The author was not aware of hypotheses concerning *old haemorrhage* and *foam cells* and their influence on plaque rupture in the literature. Therefore, these parameters were left out of the model. Additionally, *fresh haemorrhag*e was not entered, because it is a consequence of plaque rupture.

In a logistic regression model, the odds ratio of the outcome belonging to a one-unit increase in the value of the predictor variable (while fixing the other covariates) is calculated by the exponential of the corresponding regression coefficient (Appendix B4.1).[85] The population-averaged interpretation of the GEE logistic regression coefficients is similar as in a standard model.[86] *Plaque* was divided by 10^5, because the odds ratio of rupture belonging to a one pixel increase in *plaque* is not informative. A one-unit increase in *plaque* thus corresponded to a 10^5

increase in pixels in the bin, i.e. a 0.34 mm^2 area increase. The *log(x+1)* transformation of the other variables complicated the interpretation of the one-unit term. Therefore, the effect of these parameters was obtained by filling in the fitted GEE logistic regression equation. The odds ratio of rupture belonging to plaque-bins with a 10%, 30% and 50% and 70% calcium content were calculated, relative to plaque-bins with 0% calcium content. An example of this calculation can be found in Appendix B4.1). In these calculations the other plaque characteristics in the bin are held fixed. The GEE logistic regression model was executed by the geeglm-function.[87]

SENSITIVITY ANALYSIS

Initially, the author chose an 8-bin approach. In this case, the bins were large enough to actually represent something (and harbour multiple tissue structures), but small enough to still represent a "local composition". The clustered Wilcoxon rank sum test and the GEE logistic regression were repeated for a 2-, 4- and 16-bin approach in order to evaluate the choice for an 8-bin approach. The bins were formed by different combinations of the 32 bins that resulted from the Python scripts (Appendix D1), Figure 3. In the 2-bin approach, sixteen adjacent bins were combined to form the ruptured halve of the plaque and the other sixteen bins to form the non-ruptured halve. The two histological sections that harboured two plaque ruptures were excluded in this case, because it was not possible divide the section in two representative halves. In the 4-bin approach, eight adjacent bins were combined to form one ruptured bin and three non-ruptured bins. Likewise, for the 16-bin analysis eight times two neighbouring bins were combined. It was repeatedly assured that the plaque rupture was found in the centre of the "newly formed" ruptured bin. The GEE logistic regression model was run with the independence and the exchangeability working correlation matrix. The results of the best fitting model (evaluated by the QIC value), as well as the results of the clustered Wilcoxon rank sum test were reported in Appendix C3.

Results

Representative example of section segmentation

The segmentation of one of the digitized carotid endarterectomy sections is shown in Figure 5. The tissue structures are indicated with distinctive colours. Two ruptures are visible.

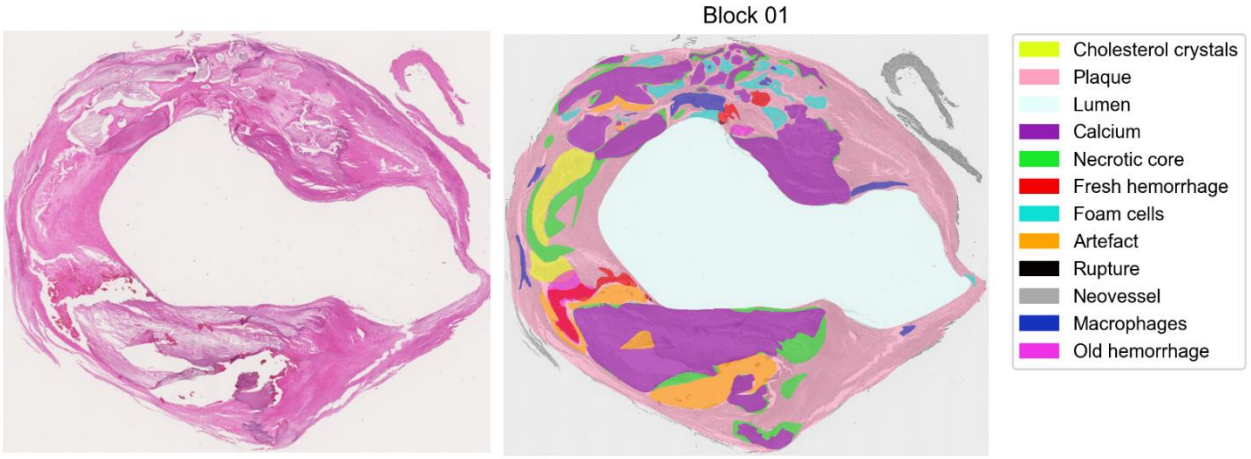


Figure 5. Representative example of one of the analyzed histological carotid endarterectomy sections (patient 4, slide 1). On the left the histology section is visible, the image on the right shows the segmentation. The section shows two ruptures. All the different tissue structures are present in the section, as indicated by the colour bar. The Histology Atlas shows the segmentation of the complete set analyzed histological sections.

The segmentation of the complete set of analysed sections (n=21) can be found in the Histology Atlas that is attached to this thesis. There, the images are shown in a higher magnification.

Spatial analysis

Ruptures were found to extend to (or start in) areas with *calcium* or *cholesterol crystals* in twenty out of 23 cases (Histology Atlas). The rupture sites were sometimes accompanied by marked inflammation (8/23 cases, Histology Atlas). Figures 6-9 show the rupture heatmaps obtained by the stacking of 23 circular binary masks from the tissue structures *plaque, cholesterol crystals, calcium* and *macrophages,* respectively. The x- and y-axis of the images represent the diameter of the clipped-out circles from the full-size masks, which always equalled 1300 pixels (i.e. 2.40 mm). The red arrow shows the centre of the image, which coincides with the luminal rupture location. The colour bar on the right of each image indicates the colour that corresponds to the adopted pixel value. The pixel value represents the prevalence of the respective tissue structure at that location.

The maximum value in the *plaque* heat map is 23 in the centre, which confirms that all 23 masks of the structure *plaque* had a pixel value of 1 at the location of rupture (Figure 6). If an imaginary vertical line is drawn trough the red arrow, it can be seen that beneath the rupture location there is no plaque, this area corresponds to the lumen. The upper part of the vertical line crosses decreasing values as it is followed towards the periphery of the image. This is the logical consequence of the fact that the plaque thickness varies.

Rupture heatmap: Plaque

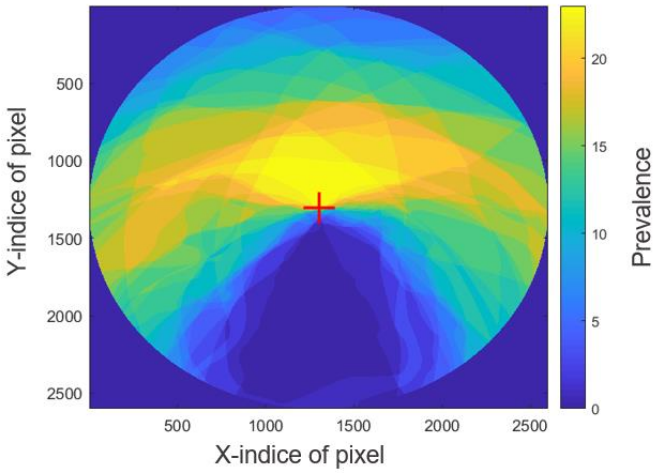


Figure 6 Rupture heatmap of the variable *plaque* (n=23)

Rupture heatmap: Cholesterol Crystals

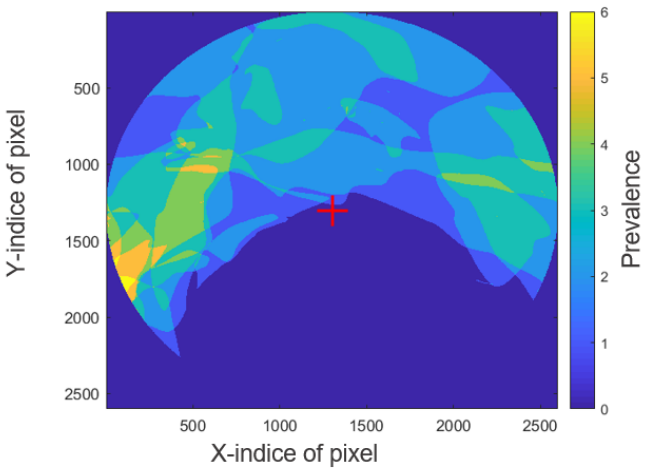


Figure 7 Rupture heatmap of the variable *cholesterol crystals* (n=23)

Rupture heatmap: Calcium

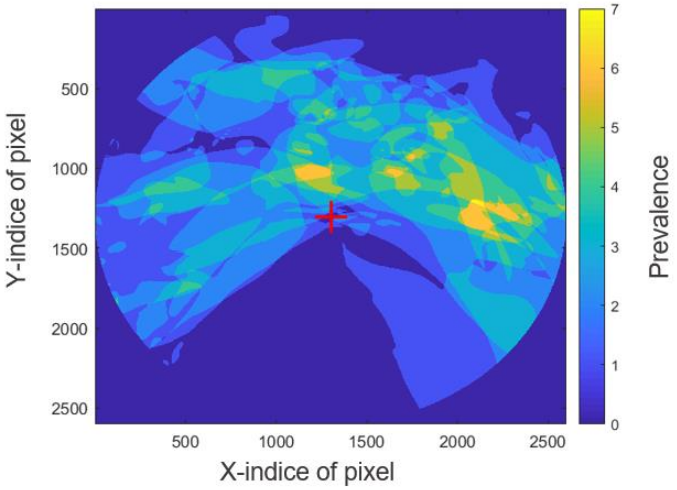


Figure 8 Rupture heatmap of the variable *calcium* (n=23)

Rupture heatmap: Macrophages

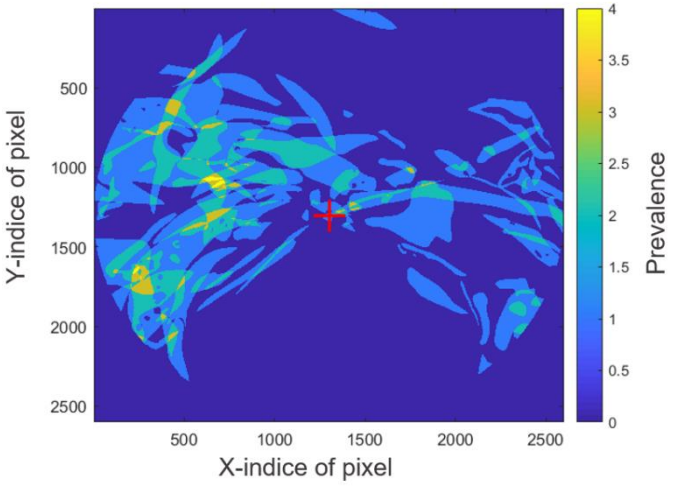


Figure 9 Rupture heatmap of the variable *macrophages* (n=23)

The maximum value in the *cholesterol crystals* heat map is 6 (Figure 7). It should be noted that the maximum score does not indicate the number of sections that show the tissue structure. Rather, it represents the number of sections that shows the tissue structure at the same location with respect to the rupture site. Based on the colour distribution it can be observed that the areas with cholesterol crystals are quite large. Therefore, it is assumed that these cholesterol crystal areas are obtained from a relatively small number of sections. Moreover, the heatmap shows that the cholesterol crystals are not necessarily positioned in the plaque tissue right behind the rupture site. They are also found more to the sides of the image.

Figure 8 shows that the *calcium* heatmap has a maximum value of 7. The calcium areas with the highest prevalence are found relatively close to the rupture site. By the colour distribution in the heatmap, small and large calcification chunks are recognized. Theoretically, all chunks can be obtained from different sections. Inflammatory hotspots sometimes surround the

location of rupture, as indicated by the heatmap of *macrophages* (Figure 9). The maximum value observed is 4. The vast majority of this heatmap shows a value of 0 or 1.

Statistical analysis

DESCRIPTION OF THE DATA

The data frame used for statistical analysis contained 168 observations/bins (21 sections x 8 bins) and eleven variables. The bins were obtained from 14 patients, which each contributed 1-4 sections. The variable *plaque* was expressed as the number of pixels in the bin and in mm^2. The variables *cholesterol crystals*, *calcium*, *necrotic core*, *fresh haemorrhage*, *foam cells*, *neovessels*, *macrophages* and *old haemorrhage* were expressed as the percentage relative to the *plaque* area in the corresponding bin. *Rupture* was indicated with a 0 (not present) or a 1 (present). Two of the sections harboured two ruptures and therefore the total amount of ruptured bins equalled 23. The variable *filename* indicated the patient number and was used to define the clusters. The descriptive statistics of the bins can be found in Table 1.

Table 1 Descriptive statistics of all bins (n=168), non-ruptured bins (n=145) and ruptured bins (n=23). All variables are expressed as percentages (relative to the plaque area in the bin), except the statistics for *plaque* which is expressed as the number of pixels in the bin and in mm^2.

	All bins (n=168)	Non-ruptured bins (n=145)	Ruptured bins (n=23)
Tissue classes	**Median [Q1-Q3]**	**Median [Q1-Q3]**	**Median [Q1-Q3]**
Cholesterol Crystals	0.00% [0.00-8.55]	00.00% [0.00-9.04]	0.87% [0.00-5.29]
Plaque	$1.08x10^6$ pixels [$6.54x10^5$-$1.61x10^6$] 3.68 mm^2 [2.22-5.49]	$1.06 x10^6$ pixels [$6.20x10^5$-$1.57x10^6$] 3.60 mm^2 [2.11-5.39]	$1.23x10^6$ pixels [$8.88 x10^6$-$1.94 x10^6$] 4.18 mm^2 [3.02-6.60]
Calcium	0.00% [0.00-13.23]	0.00% [0.00-9.90]	4.46% [0.00-15.85]
Necrotic core	9.32% [3.00-20.16]	9.47% [3.01-20.09]	8.69% [3.70-21.02]
Fresh hemorrhage	0.00% [0.00-3.56]	0.00% [0.00-0.00]	7.47% [2.81-17.18]
Foam Cells	0.00% [0.00-1.10]	0.00% [0.00-1.08]	0.21% [0.00-1.63]
Neovessels	0.00% [0.00-0.14]	0.00% [0.00-0.12]	0.00% [0.00-0.21]
Macrophages	0.70% [0.00-4.48]	0.63% [0.00-4.25]	2.65% [0.00-6.25]
Old hemorrhage	0.00% [0.00-0.04]	0.00% [0.00-0.00]	0.00% [0.00-1.70]

RELATIONS BETWEEN TISSUE STRUCTURES IN CLOSE PROXIMITY

The matrices in Figure 10 show the Kendall's Tau-b correlation coefficients between two tissue structures. The left image presents the correlation coefficients when all observations are taken into account. The right image shows the correlation coefficients in the ruptured bins only. The squares outlined in black indicate the significant coefficients. If ruptured and non-ruptured bins are considered (left image), all the significant coefficients are positive, except for the correlation between the *necrotic core* and the *neovessels*. Only two coefficients take on a value that is higher than 0.30: The correlation between *neovessels* and *macrophages* is 0.33 and between *cholesterol crystals* and *plaque* is 0.37. If only the ruptured bins are considered (right image), the three significant coefficients have values above 0.30. The correlation between *neovessels* and *macrophages* is 0.35 and between *neovessels* and *old haemorrhage* is 0.38. A negative correlation of -0.44 is found between *cholesterol crystals* and *calcium*.

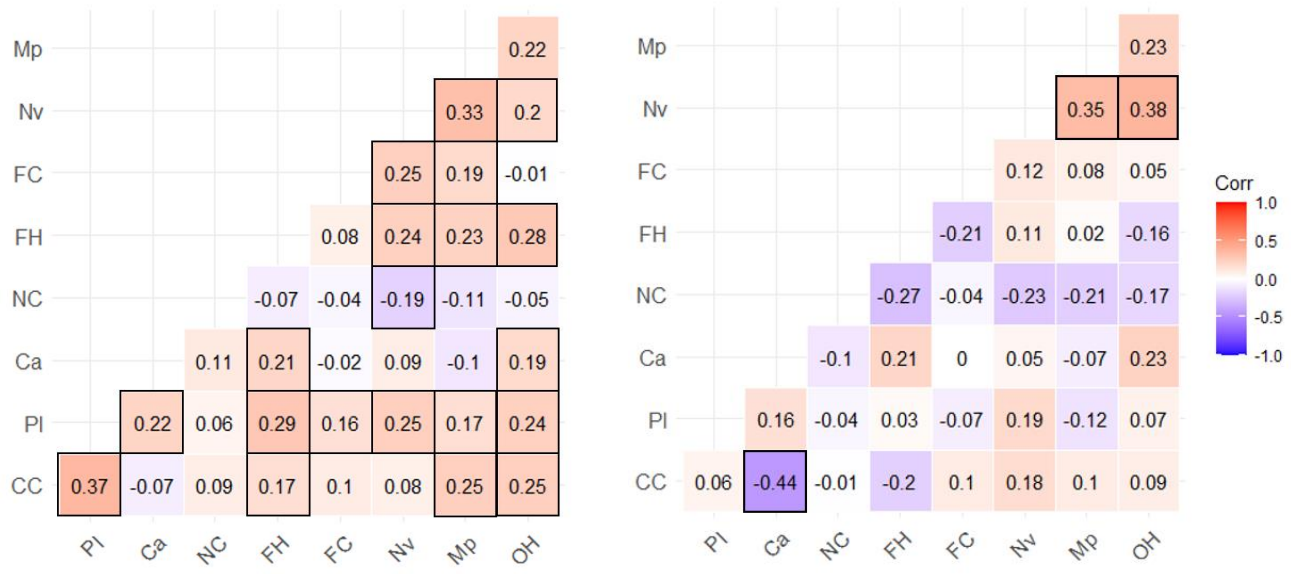


Figure 10 Kendall Tau-b correlation coefficients of all observations (left, n = 168) and in ruptured bins (right, n=23). Mp = macrophages, Nv = neovessels, FC = foam cells, NC = necrotic core, Ca = calcium, Pl = plaque, CC = cholesterol crystals, OH = old hemorrhage

COMPOSITIONAL DIFFERENCES BETWEEN RUPTURED BINS AND NON-RUPTURED BINS

The outcomes of the two-sided clustered Wilcoxon rank sum tests can be found in Table 2. The groups were created based on the presence or absence of rupture. This led to 145 observations in the ruptured bin group and 23 observations in the non-ruptured bin group.

Table 2 Outcome of clustered Wilcoxon Rank Sum test (rgl method). Group 1: ruptured bins (n=23), group 2: non-ruptured bins (n=145). Clusters are defined at the patient level. The Z-value is the test statistic. Significant p-values (≤0.05) are indicated with *.

Variables	Z-value	P-value
Cholesterol Crystals	1.57	0.12
Plaque	1.75	0.080
Calcium	1.96	0.050*
Necrotic core	0.92	0.36
Fresh hemorrhage	6.69	2.17×10^{-11}*
Foam Cells	0.61	0.54
Neovessels	0.74	0.46
Macrophages	1.33	0.18
Old hemorrhage	1.16	0.25

In total, fourteen clusters (patients) with a maximum size of 32 observations (a patient maximally contributed 4 sections x 8 bins) were present in the data. There was a (borderline) significant difference in the distributions of the variables *calcium* (p=0.050) and *fresh haemorrhage* ($p=2.17 \times 10^{-11}$) between the ruptured bin group and the non-ruptured bin group. The values of *calcium* and *fresh haemorrhage* tended to be higher in the ruptured bin group. The comparison of the distributions of the other tissue structures did not reveal significant differences between the two groups.

Table 4 shows the outcome of the GEE logistic regression model with the exchangeable working correlation matrix. Fourteen clusters (patients) were identified, with a maximum cluster size of 32 (bins). The estimated intracluster correlation parameter alpha was -0.061 and the QIC of the model was 136.42. The fitted model is represented in the following equation:

$$Log\left[\frac{P_1}{P_0}\right] = -3.39 - 0.045\log(CC+1) + 0.041\log(Pl+1) + 0.22\log(Ca+1) + 0.20\log(NC+1) - 0.86\log(Nv+1) + 0.45\log(Mp+1),$$

where *CC* = cholesterol crystals (%), *Pl* = plaque (amount of pixels/10^5) *Ca* = calcium (%), *NC* = necrotic core (%), *Nv* = neovessels (%) and *Mp* = macrophages (%).

Table 4 Outcome of GEE logistic regression model, working correlation matrix: exchangeable. Dependent variable: rupture. "Estimate" indicates the parameter estimates, "std.error" indicates the standard error of the parameter, estimated with the robust (sandwich) estimator. "Wald" indicates the Wald test statistic. "Pr(>|W|)" indicates the corresponding p-value. A p-value ≤0.05 is considered significant, indicated with *.

Variables	Estimate	Std. error	Wald	Pr(>|W|)
Intercept	-3.39	0.63	28.58	9.00×10^{-8} *
Log(Cholesterol crystals+1)	-0.045	0.13	0.11	0.74
Plaque/10^5	0.041	0.017	5.89	0.015 *
Log(Calcium+1)	0.22	0.061	12.33	0.00044 *
Log(Necrotic core+1)	0.20	0.21	0.91	0.34
Log(Neovessels +1)	-0.86	0.66	1.72	0.19
Log(Macrophages +1)	0.45	0.25	3.06	0.080

Plaque and *calcium* showed a significant association with plaque rupture, as is indicated by the Wald test (Table 4). The odds ratio of rupture belonging to a 0.34 mm^2 increase (or a 10^5 pixel increase) in *plaque* equalled 1.04 [1.02-1.06]. Table 5 shows the odds ratios of rupture belonging to plaque-bins with a 10%, 30%, 50% and 70% calcium content, relative to plaque-bins with 0% calcium content.

Table 5 Odds ratios of rupture belonging to plaque-bins with a calcium content of 10%, 30%, 50% and 70%, relative to plaque-bins with 0% calcium content. The other parameter values are assumed to remain unchanged.

Calcium content (%)	Odds ratio [CI]
10	1.68 [1.45-1.94]
30	2.09 [1.70-2.58]
50	2.33 [1.83-2.97]
70	2.50 [1.93-3.25]

Table 6 shows the outcome of a GEE logistic regression model with the independence working correlation matrix. The QIC of this model was 134.37. The fitted model is represented in the following equation:

$$Log\left[\frac{P_1}{P_0}\right] = -3.06 - 0.12\log(CC+1) + 0.046\log(Pl+1) + 0.28\log(Ca+1) + 0.048\log(NC+1) - 1.33\log(Nv+1) + 0.40\log(Mp+1)$$

Table 6 Outcome of GEE logistic regression model, working correlation matrix: independence. Dependent variable: rupture. "Estimate" indicates the parameter estimates, "std.error" indicates the standard error of the parameter, estimated with the robust (sandwich) estimate. "Wald" indicates the Wald test statistic. "Pr(>|W|)" indicates the corresponding p-value. A p-value ≤0.05 is considered significant, indicated with *

Variables	Estimate	Std. error	Wald	Pr(>|W|)
Intercept	-3.06	0.58	27.73	$1.40\text{x}10^{-7}$ *
Log(Cholesterol crystals+1)	-0.12	0.18	0.48	0.49
Plaque/10^5	0.046	0.023	3.89	0.049 *
Log(Calcium+1)	0.28	0.085	10.93	0.00094 *
Log(Necrotic core+1)	0.048	0.23	0.040	0.84
Log(Neovessels +1)	-1.33	0.98	1.84	0.17
Log(Macrophages +1)	0.40	0.28	2.05	0.15

Again, *plaque* and *calcium* were significantly associated with plaque rupture. The odds ratio of rupture belonging to a 0.34 mm^2 increase (or a 10^5 pixel increase) in *plaque* equalled 1.05 [1.02-1.07]. Table 7 shows the odds ratios of rupture belonging to plaque-bins with a 10%, 30%, 50% and 70% calcium content, relative plaque-bins with 0% calcium content.

Table 7 Odds ratios of rupture belonging to plaque-bins with a calcium content of 10, 30, 50 and 70%, relative to plaque-bins with 0% calcium content. The other parameter values are assumed to remain unchanged.

Calcium content (%)	Odds ratio [CI]
10	1.96 [1.60-2.40]
30	2.62 [1.96-3.50]
50	3.01 [2.16-4.20]
70	3.30 [2.30-4.74]

SENSITIVITY ANALYSIS

The clustered Wilcoxon rank sum test (Wilcoxon test) and the GEE logistic regression models (GEE models) were repeated for the 2-bin, 4-bin and 16-bin approach (Appendix C3). The fit of the GEE model with an exchangeable and an independence working correlation matrix was compared.

In the 2-bin analysis, the Wilcoxon test showed that the parameters *plaque* (p=0.012)*, macrophages* (p=0.044) and *fresh haemorrhage* (p=0.00018) tended to take on higher values in the ruptured halves when compared to the non-ruptured halves. *Plaque* (p=0.049) was

significantly associated with rupture in the preferred GEE model (exchangeable, alpha = -0.32). In the 4-bin analysis, the Wilcoxon test found that the parameters *plaque* (p=0.42) and *fresh haemorrhage* ($p=1.36 \times 10^{-8}$) tended to take on higher values in the ruptured bin group when compared to the non-ruptured bin group. *Plaque* (p=0.040) and *calcium* ($p=5.80 \times 10^{-5}$) were significantly associated with rupture in the preferred GEE model (independence). In the 16-bin analysis, the Wilcoxon test revealed that the parameters *calcium* (p=0.022) and *fresh haemorrhage* ($p=1.75 \times 10^{-13}$) tended to take on higher values in the ruptured bin group when compared to non-ruptured bin group. *Calcium (*p=0.0016) was significantly associated with rupture in the GEE model (independence).

Discussion

The aim of this thesis was to investigate the relation between local plaque composition and rupture in carotid plaques. Spatial and statistical analyses were carried out on 21 digitized ruptured carotid endarterectomy sections that were segmented for different tissue structures. The sections were divided into eight bins for statistical analysis. A sensitivity analysis was carried to determine the influence of the bin size on the outcomes.

Spatial analysis

Heat maps were used as a visualization method to give a spatial overview of the (local) tissue structures surrounding carotid plaque rupture. From the maps it could be observed that calcification is more often present in the nearby rupture environment than cholesterol crystals. The maximum value in the *calcium* heatmap (7) was higher than in the *cholesterol crystals* heatmap (6). Furthermore, the *calcium* chunks had a large variety in size, whereas the *cholesterol crystals* often occupied larger areas. Therefore, the areas with cholesterol crystals were assumed to be obtained from a smaller amount of sections. The observation that rupture often extended to (or started in) areas with *calcium* was also made by Daemen et al. (2016), which described the Seattle dataset.[54] The *macrophages* heatmap indicates that marked inflammation was visible at some of the rupture sites. It is known that inflammatory cells weaken tissue by the secretion of MMPs and the promotion of vascular smooth muscle cell apoptosis.[48–50] Macrophages are, however, also known to recruit to sites of tissue damage.[88] The maximum prevalence (4) and the overall prevalence of *macrophages* was low compared to the values found in the other heatmaps.

Statistical analysis

RELATIONS BETWEEN TISSUE STRUCTURES IN CLOSE PROXIMITY

In order to investigate the relation between the tissue structures in close proximity, the Kendall Tau-b correlation coefficients were calculated. When all 168 observations (the ruptured and non-ruptured bins) were taken into account, most of the significant coefficients took on a value between 0.2 and 0.3. This can be interpreted as a poor to fair relation between the variables.[89] A fair correlation[89] ($\tau_b = 0.33$) was found between *neovessels* and *macrophages*. A similar observation was made by Kumamoto et al. (1995), but it was not clear whether the inflammatory infiltrate was a consequence of the neovessels or vice versa.[52] The other significant fair[89] correlation was found between *plaque* and *cholesterol crystals* ($\tau_b = 0.37$). Thus, the cholesterol crystals content increased with the increase in plaque area. This relationship makes sense, considering the fact the plaque size increases by the "growth" of pathological tissue. Indeed, the correlation coefficients of all variables with *plaque* took on a positive value. The relation between *plaque* and *cholesterol crystals* was, however, strongest and might be influenced by the volume expansion due to crystallization as proposed by Abela & Aziz (2006).[46]

If only the ruptured bins (n=23) were analysed, three significant correlation coefficients were found. Again, a fair correlation[89] between *neovessels* and *macrophages* ($\tau_b = 0.35$) was

observed. *Neovessel* and *old haemorrhage* showed a fair correlation[89] ($\tau_b = 0.38$) as well. Milei et al. (2003) and Teng et al.(2014)[90], who did a comparable observation, suggested that the bleeding was of local origin as a result of neovessel rupture and leakage.[91] Moreover, a notable and stronger negative correlation was found between *cholesterol crystals* and *calcium* ($\tau_b = -0.44$). Apparently, the increase in the calcium content was associated with a decrease in the cholesterol crystals content in the ruptured bins. This relation was more pronounced than in the case where ruptured and non-ruptured bins were combined (-0.07, p>0.05).

COMPOSITIONAL DIFFERENCES BETWEEN RUPTURED BINS AND NON-RUPTURED BINS

Compositional differences between the ruptured bins and non-ruptured bins were tested variable-wise by the clustered Wilcoxon rank sum test. The clusters were assigned at the patient-level. The approach based on Rosner et al. (2006) (Appendix B2.2) assumes an exchangeable intra-cluster correlation.[75,76] This indicates that the correlation between the bins was assumed to be similar for every patient. The article, however, mentions that the test seems to be robust for other correlation structures as well.[76]

Significant differences in the distributions of *calcium* (p=0.050, borderline significant) and *fresh haemorrhage* ($p=2.17 \times 10^{-11}$) were found between the ruptured bin group and the non-ruptured bin group. These variables tended to have higher values in the ruptured group. Intuitively, the last observation is reasonable, since plaque rupture was only confirmed in the presence of red blood cells. The significance of the *calcium* parameter is in line with the results of the studies by Virmani et al. (2006) and Mauriello et al. (2011), which both concluded that calcification is associated with plaque rupture.[39,41]

The observations in other histological studies, i.e. plaque rupture is related to a larger necrotic core,[37–40] an increased amount of cholesterol crystals,[39,47] an increased inflammation[36,42] and an increased neovessels density,[44,45] were not supported by results of the clustered Wilcoxon rank sum test. This could be due to the fact that these studies compared non-ruptured and ruptured plaques at the section level, and not at the local level as was done in this thesis. Furthermore, other studies investigated coronary[37,39,47], aortic[38,40,44] or carotid plaques[36,42,45,47], whereas this study is based on the segmentation of carotid plaque sections only. Due to the exploratory character of this study and to the unavailability of histological studies examining local plaque composition, it was decided to look into plaque rupture studies based on plaques from different arterial territories. Alternative factors that could have influenced the outcome include limitations in the thesis dataset (small and clustered) and the segmentation. Moreover, all extracellular lipid accumulations were segmented as *necrotic core* in this thesis, while other studies reserved this term (or the term "lipid pool") for the large lipid-rich lesions covered by a fibrous cap.[37–40] They also included the cholesterol crystals in the necrotic core parameter.[37–40] Certainly, these choices could lead to different results.

ASSOCIATION OF TISSUE STRUCTURES WITH PLAQUE RUPTURE

The association of the tissue structures with plaque rupture was examined with a GEE logistic regression model (from now on called GEE model) that had clusters defined at the patient level. The GEE method models the population mean response to given changes in the

variables, and cannot be used for subject-specific predictions. The term "population" corresponds to all observations/bins within the dataset.[92]

The first GEE model was executed with an exchangeability working correlation matrix, in which the intracluster correlation is similar for all observations (Appendix B4.2). Because the strength of the associations can still vary across the clusters, the model uses the average correlation, indicated by the correlation coefficient alpha.[79] This implies that correlation between all bins of a patient was assumed to be similar and that value of this correlation was assumed to be similar for every patient. The parameter alpha was estimated as -0.061, indicating that the intracluster correlation is negative. Taking into account the heterogeneity of atherosclerotic plaques[80] and the low value of the estimated parameter alpha, the independence working correlation matrix was tried in the second GEE model (Appendix B4.2). In this case, the observations in a cluster are assumed to be uncorrelated (i.e. all bins are considered to be independent observations) and the model reduces to a standard logistic regression.[79,93]

In both models, two of the included parameters (*cholesterol crystals, plaque, calcium*, *necrotic core, neovessels* and *macrophages*) were found to be significantly associated with plaque rupture, namely *plaque* and *calcium*. Based on the QIC score, a preference was found for the GEE model with the independence working correlation matrix. In this model, a 0.34 mm^2 increase in *plaque* was associated with an odds ratio of rupture (OR) of 1.05 [1.02-1.07]. The OR represents the odds that rupture occurs in bins with a one-unit increase in plaque size compared to the odds that rupture occurs in bins without a one-unit increase in plaque size (Appendix B4.1). The OR of rupture belonging to plaque-bins with a 10% *calcium* content and 50% *calcium* were 1.96 [1.60-2.40] and 3.01 [2.16-4.20], respectively, relative to plaque-bins with 0% calcium content. In the calculation of odds ratios, the other plaque parameters in the bin were held fixed.

The outcome of the GEE model suggests that *calcium*, and *plaque* size are associated to plaque instability. It is known that a mismatch between the material properties of tissue structures can lead to stress concentrations. One of the proposed failure mechanisms is that rupture occurs when local plaque stress is higher than the local plaque material strength.[13–15] A 2D-finite element study performed by Teng et al. (2004) found a positive relation between the degree of calcification and the maximum principal stress experienced in coronary plaques. More extensive calcification, however, led to a plateau in the peak stress or even a decrease.[27] Other 2D-FEA studies showed that the location of the calcification is important in carotid plaques. Li. et al (2007) found that calcification in the fibrous cap led to high stress concentrations in the plaque. Additionally, they showed that addition of calcium in the necrotic core or further away for the lumen did not have a great effect on the experienced maximum von Mises stress.[29] Contrarily, Wong et al. (2012) found that an increase in the calcification gap (i.e. the width of necrotic core between the fibrous cap and the calcification) increased the peak principal stress within the plaque.[28] Moreover, uniaxial failure tests carried

out by Mulvihill et al. (2013) indicated that clustering calcium nodes created voids that lead to carotid plaque rupture.[94]

In a histological study by Virmani et al. (2006), an increased degree of calcification was found in ruptured coronary plaques compared to non-ruptured plaques. They did not speculate about the underlying mechanism.[39] Additionally, Mauriello et al. (2011) reported that in carotid plaques large calcifications were associated with asymptomatic minute ruptures, while symptomatic and wide cap ruptures had little calcifications. The author does not contemplate on the physiology behind the observations, but mentions the possibility of different pathogenetic mechanisms.[41] Daemen et al.(2016) proposed that minor luminal surface disruptions in carotid plaques could occur due to delamination by "shear-like forces" at the border of loose matrix and calcification.[54] In this thesis, no distinction was made between wide cap ruptures and minute plaque ruptures.

As stated before, it stands to reason that the plaque size is positively associated to rupture, as the increase in plaque size is caused by the growth of pathological tissue.[95] The results are in line with Narula et al. (2013) and Moreno et. al (2002) which both found an increased plaque area in mm^2 in ruptured sections compared to non-ruptured sections of coronary and aortic plaques, respectively.

It was remarkable that the intracluster parameter alpha was negative in all GEE models (i.e. in the 8-, 2-, 4- and 16-bin approach, Appendix C3). Intuitively, it would be expected that bins obtained from the same patient are more alike, rather than being negatively correlated. Especially in the 2-bin analysis the parameter was high (-0.30). Even though the reason of this negative correlation is not entirely clear, one important factor could have contributed. Staining variations, for example in colour intensity, affect the appearance of histological images and make it challenging to use other sections as a control.[96] Often, the segmentation results in a sort of "relative segmentation". For example, *necrotic core* was defined as "relatively pale", and as a control the other plaque tissue in the same section was used, rather than the plaque tissue of a different section.[60]

SENSITIVITY ANALYSIS

The discussed results were obtained by an 8-bin approach. A 2-, 4- and 16-bin analysis were carried out to determine the effect of the bin size on the outcome of the statistical models. In the 2- and 4-bin analysis, the clustered Wilcoxon rank sum test showed that *plaque* and *macrophages* tended to take on higher values in the ruptured bins, when compared to the non-ruptured bins. Contrarily, the 8- and the 16-bin analysis found that *calcium* had a tendency to show higher values in the ruptured bins. Apparently, when considering the plaque rupture site at an increasingly local level, *plaque* and *macrophages* lost significance and *calcium* gained significance. In both the 2- and 16-bin GEE logistic regression models, only one parameter showed significance: the *plaque* and *calcium* parameter, respectively. The 4- and 8-bin approach both showed a significant association between the parameters *plaque* and *calcium* and plaque rupture. The choice for using 8 bins was reasonable, taking into account the

corresponding bin size (i.e. the degree of "localness" of the observations) and the interesting results of this approach.

Clinical translation

Estimating the plaque rupture risk based on plaque composition generates an added clinical value if there are *in-vivo* visualization techniques that can determine the plaque morphology. Two intravascular imaging techniques known for their ability to characterize coronary plaques are Optical Coherence Tomography (OCT) and Intravascular Ultrasound (IVUS).[97] OCT is a high resolution imaging technique that uses near infrared light. It can distinguish between fibrous plaques, fibro-calcific plaques and lipid-rich plaques.[98,99] Additionally, neovessels, cholesterol crystals and macrophage accumulations can be identified.[99] Contrary to OCT, IVUS uses sound waves to visualize plaque morphology.[97] IVUS has a limited resolution and is less convenient for the assessment of microstructural components. However, IVUS is superior to OCT in terms of penetration depth and the imaging techniques can complement each other under certain circumstances.[100] A disadvantage of both visualisation techniques is that they are invasive.

Currently non-invasive imaging techniques based on magnetic resonance imaging (MRI), computed tomography (CT) and ultrasound (US) are being studied for their applicability in carotid plaque characterization.[101] Especially MRI shows a high specificity and sensitivity in the determination of the plaque composition.[102] The implementation of MRI for screening purposes is, however, impractical, due to the long duration of the examination, limited availability and high costs.[103] CT angiography of the carotid arteries has the ability to discriminate between different plaque components, but requires iodinated contrast and radiation.[101] US is useful in the determination of the stenosis severity and can make some distinction in plaque type (lipid- or calcium-rich), but a detailed study of plaque composition is not yet possible.[101,104] Techniques based on US are especially interesting from a screening perspective, since US examination is fast, relatively cheap, radiation-free and there are almost no contraindications.[104]

With the improvement of these imaging techniques, *in-vivo* segmentation of the plaque tissue structures is becoming realistic. It is known that carotid endarterectomy can reduce the risk of stroke in carotids with a high stenosis degree (>70%).[105] The benefit of this procedure is less established in patients that show a lower carotid stenosis degree, but a significant number of strokes still occurs in these patients.[105,106] A surgical intervention has peri-and postoperative risks and may not be preferred in certain patients due to their physical condition or other restrictions.[5] Besides, atherosclerosis is a systemic disease that does not disappear by surgical intervention and neither does the risk of stroke. The Athero-express study found that the reoccurrence of local restenosis after carotid endarterectomy is related to the plaque composition.[107] Personalized rupture (and restenosis) risk assessment that takes into account the plaque composition could therefore contribute to medical decision making.

Limitations and future research

There are some limitations to this study. Firstly, the dataset was small and comprised 21 ruptured sections (obtained from fourteen patients). The dataset is, however, valuable, since high quality sections that show plaque rupture are hard to acquire. With an increased sample size, the chance of the detecting small differences grows. Also, it is expected that with a greater sample size the behaviour of the "whole group" of carotid plaques is better represented.

Secondly, the segmentation of the majority of the sections was solely carried out by the author. The inter-observer variation was not tested. The author looked multiple times through the sections and sometimes the segmentation was slightly adjusted. This indicates that intra-observer variation is present. Although not quantified, inter- and intra-observer variation are not expected to change the conclusions of this study. Moreover, future studies with automated segmentation tools could take away the inter- and intra-observer variation that is now present.

Even though the segmented tissue structures can be recognized on H&E staining, the use of staining combinations (e.g. H&E and Masson Trichrome[38,47,52,108]) or immunohistochemistry (e.g. CD68 to detect macrophages [37,43,109], CD34 to mark endothelial cells[108,110]) is usually preferred. The identification of small structures *(neovessels* and *foam cells)* was challenging, and so was finding the exact borders of e.g. areas with *necrotic core* and *macrophages* (inflammatory hotspots). The distinction between *macrophages* and *foam cells* could be difficult. For these reasons, the results of the segmentation more likely show a tendency than absolute values. However, H&E staining is good enough for general staining and the conclusions of this study are expected to remain valid.

No distinction could be made between intraplaque haemorrhage or haemorrhage due to plaque rupture. The parameter *fresh haemorrhage* was, therefore, left out of the GEE logistic regression model. There are studies that found a link between intraplaque haemorrhage and the presence of clinical symptoms.[111] It is not expected that tissue manipulation by the surgeon led to plaque rupture and/or haemorrhage. The surgeon approached the plaque from the adventitia and removed it as a whole, thereby the luminal surface of the plaque was kept intact. [54]

This thesis investigated local plaque composition by the division of the histological sections into bins. In future studies, it would be interesting to add a measure of luminal distance to the tissue structures. This could enhance our insight in the importance of the depth of different tissue structures in the process of plaque rupture.

The robust sandwich estimator was used to determine the standard errors of the GEE logistic regression parameters (Appendix B4.2). There are indications that in certain situations (e.g. small sample size and low event rate) the sandwich estimator might lead to underestimation of the parameter variance. Proposed alternative estimators also had performance issues and literature has not yet agreed upon a solution.[112]

Lastly, it is possible that linearity (assumed by the GEE logistic regression model) is not the most effective way to describe the relation of some predictor variables to the outcome of plaque rupture (Appendix C2.3). The implementation of machine learning techniques in future studies can determine the fit of non-linear models.

Conclusion

To the author's knowledge, this is the first study that investigated the relation between local plaque composition and rupture. The study was based on the segmentation of digitized H&E stained sections of ruptured carotid plaques. A statistical analysis revealed a greater calcium content locally around rupture sites. Additionally, the calcium content and the plaque size were found to be positively associated with plaque rupture. An increased understanding of the plaque rupture process can contribute to rupture risk stratification and eventually to more informed medical decision making. Furthermore, this thesis used statistical methods that were suitable for clustered data. It thereby provides a method that can be used to examine clustered datasets that often appear in the medical field.

Appendix A: Introduction

Anatomy of the healthy arterial wall

An artery consists of three layers: the *tunica intima*, the *tunica media* and *the tunica adventitia* (Figure 1).[113,114] The *tunica intima* is composed of endothelial cells, a subendothelial layer and the internal elastic lamina (IEL). The endothelial cells form a physical barrier that can protect the artery from external influences.[55] The IEL marks the transition to the *tunica media*. The *tunica media* contains concentric smooth muscle cells (SMCs) alternated by sheets of elastin and collagen fibres. Extracellular matrix (ECM), including elastin and collagen fibres, is produced by the SMCs.[55,115] The separation between the *tunica media* and the *tunica adventitia* is less well defined by the external elastic lamina.[115] The *tunica adventitia* consists of a collagenous ECM and fibroblasts. Additionally, perivascular (autonomic) nerves, lymphatic vessels, resident populations of immune cells, and the vasa vasorum can be found in this layer. The latter supplies the *tunica adventitia* and *the tunica media*.[116]

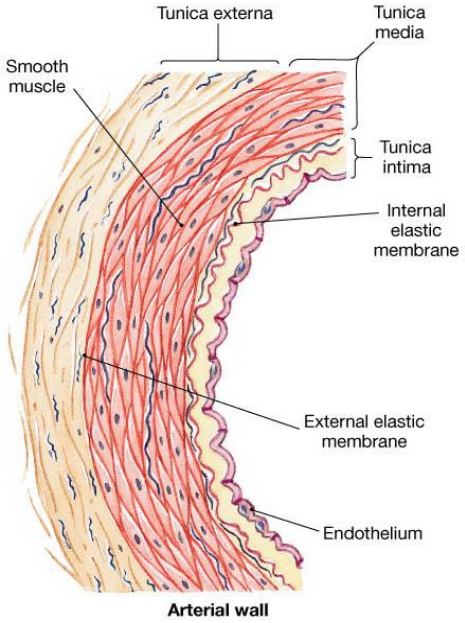


Figure 1 The composition the arterial wall, showing the three arterial layers, endothelial cells, smooth muscle cells and the internal elastic membrane. [114]

Pathophysiology of atherosclerosis

Atherosclerotic disease disorganizes the arterial wall. There are uncertainties about the pathophysiology, as many different factors play a role in this disease.[18] Endothelial dysfunction can be seen as an initial step.[117] Low-density lipo-protein cholesterol (LDL) and monocytes that are present in the blood can hereby enter the intimal layer.[118] Inside the subendothelial space, the monocytes differentiate into pro-inflammatory macrophages and the LDL becomes oxidized, acting as a chemoattractant.[119] The oxidized LDL accumulates in the macrophages, which transform into foam cells.[118,119] The apoptosis of foam cells releases cytokines and growth factors that further increase the pro-inflammatory state.[117] The SMCs respond by migrating from the tunica media into the intimal layer. They produce ECM that

can help in plaque stabilization. SMCs are also capable of taking up lipids. Additionally, SMCs produce cytokines that increase monocyte accumulation and promote endothelial cell dysfunction.[120] A proposed sequence of plaque formation in atherosclerosis is visualised in Figure 2.

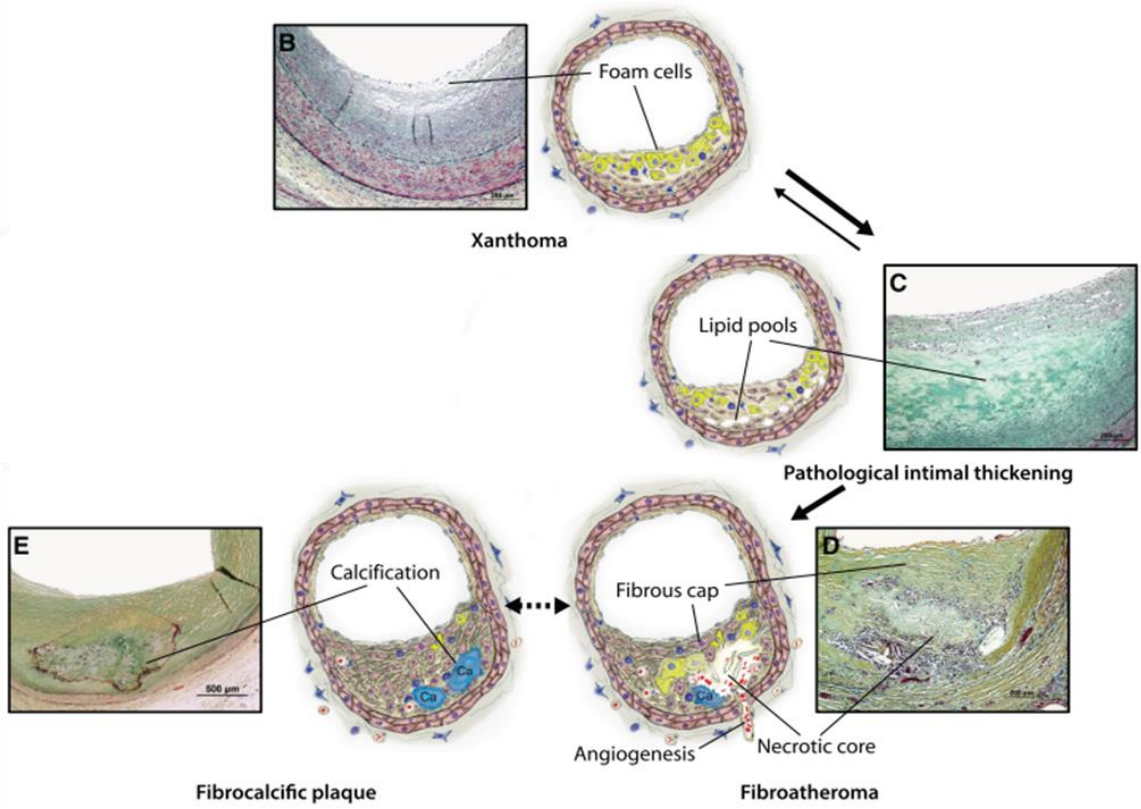


Figure 2 Proposed sequence for the progression of atherosclerotic lesions. Foam cells are macrophages that took up oxidized LDL. The *xanthoma* is composed of multiple layers of foam cells (2B). Lipid pools can form beneath these foam cell layers (2C). This process is called pathological intimal thickening and is irreversible. The coalescence of lipid pools leads to the formation of a necrotic core, which is covered by a fibrous cap (2D: fibroatheroma). Angiogenesis denotes the growth of neovessels growing in the advanced plaque. In time, the necrotic core can calcify or show fibrosis (2E: *fibrocalcific plaque).*[18]

Xanthomas represent initial lesions and consist of multiple foam cells layers. These lesions are reversible. Extracellular lipid pools can form beneath the xanthomas, leading to irreversible *pathological intimal thickening*. Lipid pools can coalescence to form a necrotic core, a process stimulated by macrophage infiltration and presumably by the apoptosis of foam cells and SMCs.[18] A fibrous cap, consisting of ECM, SMCs and sometimes macrophages, covers the necrotic core.[117,121] A lesion with a necrotic core is called a *fibroatheroma*. Neovessels (primarily originating from the vasa vasorum) can grow into the advanced plaque, a process known as angiogenesis.[116] In time, the ECM production of the SMCs can lead to the replacement of loose fibro-cellular tissue by collagen-rich fibrous tissue.[18] It is assumed that calcification could start with the apoptosis of SMCs and macrophages that contain matrix vesicles.[122] The necrotic core, as well as the ECM, can show partial or complete calcification. The *fibrocalcific plaque* lost its necrotic core due to fibrosis or calcification.[18] The *fibroatheroma* and the *fibrocalcific plaque* represent two possible end stages of atherosclerotic lesions, but many more exist.[18]

Appendix B: Methods

1. Kendall Tau-b correlation coefficient

The Kendall Tau-b correlation coefficient (τ_b) can be used to calculate the strength of the association between two ordinal or continuous variables. The calculation is based on the number of concordances and disconcordances in the data.[70,71] Suppose that the data consists of *n* observations of variables *X* and *Y*.

The two observations (X_i, Y_i) and (X_j, Y_j) are concordant if:
(1) $X_i < X_j$ and $Y_i < Y_j$
or
(2) $X_i > X_j$ and $Y_i > Y_j$
And they are disconcordant if:
(1) $X_i < X_j$ and $Y_i > Y_j$
or
(2) $X_i > X_j$ and $Y_i < Y_j$
If $X_i = X_j$ and/or $Y_i = Y_j$ the observations are tied.[70]

The total number of possible comparisons of observations (X_i, Y_i) and (X_j, Y_j) (i.e. the number of pairs that can be assembled) *N* can be calculated by:[70,71]

$$N = \frac{n(n-1)}{2}$$

N equals the sum of the number of concordant pairs (C), disconcordant pairs (D), the number of pairs tied on the X-variable (X_0), the number of pairs tied on the Y-variable (Y_0) and the number of pairs tied on the X and the Y variable $(XY)_0$.[70]

The Kendall Tau-b correlation efficient is found by the application of the following equation[70,71]:

$$\tau_b = \frac{(C-D)}{\sqrt{(C+D+X_0)(C+D+Y_0)}}$$

The τ_b always takes on a value between -1 and 1. Here, -1 and 1 imply a perfect relationship between the two variables and 0 indicates that the parameters are completely independent. A positive relation shows that the ranks of both variables increase together. A negative implies that rank of one variable increases with the decrease of the rank of the other variable.

2.1. The Wilcoxon rank sum test

The original Wilcoxon rank sum test is not used in this thesis. However, a description of this test is given to provide the reader with some background knowledge. This knowledge might help the reader to grasp the idea behind the related clustered Wilcoxon rank sum test.

The Wilcoxon rank sum test can be used to determine whether the distribution of a variable is the same in two independent groups. A null hypothesis (H_0) and an alternative hypothesis (H_1) are defined. For two-sided version of the Wilcoxon rank sum test, these are:[73]
H_0: The distribution of the variable in both groups is the same
H_1: The distribution of the variable is not the same in both groups

The one sided upper-tail version tests the following hypotheses:[123]
H_0: The distribution of the variable in both groups is the same
H_1: The distribution of one group is shifted to the right compared to the distribution of the other group

Consider $X_1, \ldots, X_m$ to be m observations from a population with distribution F and $Y_{1,\ldots,} Y_n$ to be n observations from a population with distribution G. Suppose that $n < m$. The test assumes that the two groups, as well as the observations within the different groups, are independent. Additionally, it is assumed that both distributions F and G are continuous. The X's and Y's are combined and ranked. The combined sample has size $N = m + n$. If observations are tied, they receive the mean of the ranks that would have been assigned if there were no tied observations. The sum of the ranks is calculated in the group with the smallest number of observations and is known as the W-statistic. If S_j denotes the rank of Y_j $(j = 1, \ldots, n)$, then:[74,124,125]

$$W = \sum_{j=1}^{n} S_j$$

For a large sample approximation of W, the W*-statistic asymptomatically follows a standard normal distribution and is calculated by:[73,74,125]

$$W^* = \frac{W - E_0(W)}{\sqrt{var_0(W)}}$$

In this equation E_0 (W) = $n(m + n + 1)/2$ and represents the expected value of W under the null hypothesis. $var_0(W)$ is the null variance of W. If there are ties in the data, the following equation should be used to obtain this parameter:

$$var_0(W) = \frac{mn}{12}\left[m + n + 1 - \frac{\sum_{j=1}^{g}(t_j - 1)t_j(t_j + 1)}{(m+n)(m+n-1)}\right]$$

Here g represents the number of tied groups and t_j denotes the size of tied group j. If no observations are tied, t_j = 1, g = N and the fraction between the square brackets repeatedly reduces to zero.[125]

If $|W^*| \geq w_{1-\alpha/2}$, the H_0 of the two sided-version of test can be rejected and if $W^* \geq w_\alpha$, the H_0 of the one-sided upper tail version of the test can be rejected. $w_{1-\alpha/2}$ denotes the (1- $\alpha/2$) quantile, whereas w_α corresponds to the α quantile of the standard normal distribution w. α is the type 1 error probability (significance level).[73,125]

2.2. The clustered Wilcoxon rank sum test

A somewhat different approach can be used when the observations are clustered. Rosner et al. (2006) developed an extension to the Wilcoxon rank test that can be applied if data is clustered and when the members of the different clusters belong to different groups.[75,76]

In short, the test statistic is based on estimator $\hat{\theta}$ of θ, θ equalling the probability that an observation in group 1 is greater than the observation in group 2, which is ½ under H_0.[75,76] The test statistic is standardized by the variance of the estimator $\hat{\theta}$, $\widehat{var}(\hat{\theta})$, which takes into account an exchangeable relation between the ranks of the subunits of each cluster. The estimator $\hat{\theta}$ is combined across clusters of different sizes.[76] The approach assumes that no correlation exists between clusters.[75,76] A part of the derivation of the test statistic W^* can be found below.

Group 1 is indicated by δ= 1 and group 2 by δ = 0. The cluster size (or: the number of subunits within a cluster) is indicated with (superscript) g and $1 \leq g \leq g_{\max}$. $g_{\max}$ is the maximum cluster size (< ∞). $N^{(g)}$ indicates the number of clusters with size g. [75,76] The following hypothesis will be tested:

$$H_0: F_x^{(g)} = F_y^{(g)} \text{ ,g } = 1 \ldots g_{max}$$
$$H_1: F_x^{(g)} \neq F_y^{(g)} \text{ ,g } = 1 \ldots g_{max}$$

Where g = *1* … g_{max}. $F_x^{(g)}$ represents the cumulative distribution fraction in group 1 and $F_y^{(g)}$ represents the cumulative distribution fraction in group 2.[75]

The overall test statistic W^* asymptomatically converges to a standard normal distribution under mild assumptions as the total number of clusters $N = \sum_g^{g_{max}} N^{(g)} \to \infty$.

W^* is given by:[75,76]

$$W^* = \frac{\left(\hat{\theta} - \frac{1}{2}\right)}{\sqrt{\{\widehat{var}(\hat{\theta})\}}}$$

In this equation estimator $\hat{\theta} = \frac{\sum_{g=1}^{gmax} w^{(g)}\hat{\theta}^{(g)}}{\sum_{g=1}^{gmax} w^{(g)}}$ and $\widehat{var}(\hat{\theta}) = 1/\sum_{g=1}^{gmax} w^{(g)}$, where $w^{(g)} = 1/var(\hat{\theta}^{(g)})$. The parameter $\hat{\theta}^{(g)}$ is based on the related Mann-Whitney U statistic for clusters of size g and equals $Pr(Z_{ij}^{(g)} > Z_{kl}^{(g)})$, when $\delta_{ij}^{(g)} = 1$ and $\delta_{kl}^{(g)} = 0$.[75] $\delta_{ij}^{(g)}$ is only 1 if the jth subunit in the ith cluster of size g belongs to group 1.

The estimated $var(\hat{\theta}^{(g)})$ is obtained by an extensive and involved equation. The equation includes the outcome of a two stage randomization distribution of the rank statistic $w_c^{(g)} = \sum_{i=1}^{N}\sum_{j=1}^{g}\delta_{ij}R_{ij}$ at the cluster level. Again, $\delta_{ij} = 1$ when the jth subunit of the ith cluster belongs to group 1, otherwise $\delta_{ij} = 0$. Furthermore, R_{ij} is the rank of the observation, g is the cluster size and N equals the number of clusters. Moreover, $var(\hat{\theta}^{(g)})$ takes into account $N^{(g)}$ and the correlation between the ranks of the subunits. The intra-cluster correlation is assumed to be exchangeable.[75,76]

The complete derivation of the Wilcoxon rank sum test statistic for clusters of varying size with grouping at the subunit level can be found in Rosner et al (2006).[75]

3. Principal component analysis

Principal component analysis is a technique that is able to reduce the dimensionality of the data.[126] With this technique new variables, the principle components (PCs), are obtained. The PCs are a linear combination of the original variables.[127] They each explain a proportion of the variance of the data. The first PC explains the largest proportion, the second PC explains the second largest proportion etc. The PC's are orthonormal to each other and thus uncorrelated. PCA inherently assumes that the most important dynamics of the data are captured by the largest variances in the data.[128]

PCA requires data pre-processing. The data should be organized in a m x n matrix $\boldsymbol{X}$, where m is the number of observations and n is the number of variables. Each observation can thus be seen as an n-dimensional vector.[128] Matrix **Y** can be obtained from matrix **X** if the variable mean is subtracted from every column (in R, this is specified by the parameter "center" [129]). Additionally, the variables are divided by their standard deviations (in R, this is specified by the parameter "scale" [129]).[128,129] Matrix **Y** is now a scaled, zero-centred m x n representation of the data. The PCA technique used in this thesis relies on the singular value decomposition of **Y**.[129]

The following equation is used:

$$\boldsymbol{Y} = \boldsymbol{USV}^T$$

Here, **U** is a *m x n* matrix of which the columns form an orthogonal basis for the range of **Y** and represent the left singular vectors. **S** is a diagonal matrix of the form *n x n* and the nonzero elements are the singular values that are organized in a decreasing order. ***V*** is an orthogonal *n* x *n* matrix of which the columns are called the right singular vectors[130]. The right singular vectors are the principle components of matrix **Y**. The proportion of variance explained by the principle component is found by taking the square of the corresponding singular value.[128,131]

4.1 The logistic regression model

A logistic regression model is a linear model can be used to determine the effect of the predictor variables on the outcome. The outcome is a binary variable that can only take on the value 0 or 1. For example, presence of rupture= 1, absence of rupture = 0. Due to the binary nature of the outcome, the outcome itself cannot be used as the dependent variable. Instead, the continuous variable probability (p) is introduced. A cut-off value is specified for the probability above which the predicted outcome will be 1.[73]

The link function is the function that describes the relationship between the predictor variables and the dependent variable.[132] A link function widely used in logistic regression (and also the default link function for logistic regression in R[133]) is the logit link:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

In this equation, $\text{logit}(p)$ scales the probability to a value between 0 and 1. $\ln\left(\frac{p}{1-p}\right)$ represents the logarithmic of the odds (log odds) of the outcome that is marked as 1. Additionally, β_0 is the intercept term, $x_1, x_2 \ldots x_m$ are the predictor variables and $\beta_1, \beta_2 \ldots \beta_m$ are the estimated regression coefficients that determine the size of effect of the predictor variables.[134] In a logistic regression model, the log odds of the outcome marked as 1 is a linear combination of the predictor variables.

The exponential of a regression coefficient gives the odds ratio associated with a one-unit increase in the corresponding predictor variable.[85] The odds ratio (OR) compares the odds of the outcome (case (1) /non-case (0)) in the presence of exposure to the odds of the outcome without the exposure:[85]

$$OR = \frac{number\ of\ exposed\ cases / number\ of\ exposed\ non-cases}{number\ of\ non-exposed\ cases / number\ of\ non-exposed\ non-cases}$$

A non-linear transformation of the predictor variables complicates the interpretation of this "one-unit" term. In this thesis, multiple variables underwent a *ln(x+1)* transformation. Another way was found to express the effect of these parameters, namely by filling in the fitted equation. We can retain the odds of the outcome labelled as 1 by exponentiating the log odds of the outcome labelled as 1, which gives:[135]

$$odds = \frac{p}{1-p} = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m)}$$

An example:

Let x_1 be the ln(x+1) transformed variable *calcium*. Let β_1 be the corresponding coefficient. The odds ratio of rupture (marked as 1) belonging to plaque-bins containing 20% calcium (x=20) relative to plaque-bins containing 0% (x=0) calcium can be calculated as follows:

$$0\% \text{ calium content: } x_1 = \ln(x+1) = Ln(0+1) = 0$$

$$20\% \; calcium \; content: \; x_1 = ln(x+1) = Ln(20+1) = ln(21)$$

$$odds \; of \; rupture \; 0\% \text{ calcium } = e^{(\beta_0+0\beta_1+\beta_2 x_2+\cdots+\beta_m x_m)}$$

$$odds \; of \; rupture \; 0\% \text{ calcium } = e^{(\beta_0+ln(21)\beta_1+\beta_2 x_2+\cdots+\beta_m x_m)}$$

$$OR = \frac{odds \; 20\% \; calcium}{odds \; 0\% \; calcium} = \frac{e^{(\beta_0+ln(21)\beta_1+\beta_2 x_2+\cdots+\beta_m x_m)}}{e^{(\beta_0+0\beta_1+\beta_2 x_2+\cdots+\beta_m x_m)}} = \frac{e^{(ln(21)\beta_1)}}{e^{(0)}} = e^{(ln(21)\beta_1)}$$

This is the odds ratio when the other characteristics of the bins are held fixed.

4.2. GEE as an extension of the logistic regression model

In 1986, Zeger and Liang proposed a methodology for cluster analysis called generalized estimating equations (GEE). The GEE method is based on the quasi-likelihood approach. The quasi-likelihood approach specifies the variance of the outcome variable as a known function of the expectation.[136] The GEE method can be used as an extension of a logistic regression model and other generalized linear models as a way to deal with clusters in the data.

In this thesis the GEE method is applied, because there are clusters in the dataset. There is a dependency between the sections of the same patients and between the bins of the same slides. Clusters are assigned at the patient level and the subunits in the clusters are the bins.

Suppose that there are $i = 1, \ldots, K$ patients that contribute $b = 1, .., n_i$ bins. Let Y_{ib} the outcome variable (rupture (1), non-rupture (0)) of bin b in patient i. Let X_{ib} be a $p \; x \; 1$ vector of p covariates.

Define μ_{ib} to be the expectation of Y_{ib} and relate it to the vector of covariates X_{ib} by function g: [93,137,138]

$$E(Y_{ib}) = \mu_{ib} = g(X_{ib}^T\beta)$$

The inverse of g is called the link function l.

$$l(\mu_{ib}) = X_{ib}^T\beta$$

Here, β is a $p\ x\ 1$ vector of regression coefficients. The logit link is used in the case of a binary outcome.[93]

Now, let Y_i be the be $n_i\ x\ 1$ vector of outcomes and X_i be the $n_i\ x\ p$ matrix for the ith subject. Define μ_i to be the expectation of Y_i. Again, let β be the $p\ x\ 1$ vector of regression coefficients. The generalized link function $l(\mu_i) = \left(l(\mu_{i1}), \dots, l(\mu_{in_i})\right)^T$ links the mean response vector μ_i to the covariates:[93,139]

$$E(Y_i) = \mu_i$$

$$l(\mu_i) = X_i\beta$$

Based on the quasi-likelihood approach, it is assumed that the variance of the outcome Y_{ib} depends upon the expectation via function V_i:

$$var(Y_{ib}) = V_i(\mu_i)\phi$$

where ϕ is a constant scale parameter. If the outcome is binary, ϕ is taken as 1.[81,93] The working covariance matrix V_i is given by:[136,138]

$$V_i = A_i^{\frac{1}{2}}R_i(\alpha)A_i^{\frac{1}{2}}$$

In this function, A_i is an $n_i\ x\ n_i$ diagonal matrix with entries $V(\mu_{ib})$. Because Y_{ib} is binary, $V(\mu_{ib}) = \mu_{ib}(1-\mu_{ib})$. $R_i(\alpha)$ is the $n_i\ x\ n_i$ working correlation matrix that defines the relation between the bins that are obtained from the same patient.[136] The available working correlation matrices in the R "geepack" are found in Table 1.[138] Additionally, users can define their own working correlation matrix.

Table 1 Working correlation matrices available in R package "geepack".[138]

Working correlation matrix	Intracluster correlation	
independence	$\mathrm{Cor}(\mathrm{Y_{ib}}, Y_{ib'}) = 0,$	$b \neq b'$
exchangeable	$\mathrm{Cor}(\mathrm{Y_{ib}}, Y_{ib'}) = \alpha,$	$b \neq b'$
ar1 (autoregressive)	$\mathrm{Cor}(\mathrm{Y_{ib}}, Y_{ib'}) = \alpha^{\lvert b-b' \rvert}$	
unstructured	$\mathrm{Cor}(\mathrm{Y_{ib}}, Y_{ib'}) = \alpha_{bb'},$	$b \neq b'$

The estimate of β $(p\ x\ 1)$ can be found by solving the following equation via an iterative algorithm: [136,138]

$$\sum_{i=1}^{K} \left(\frac{\delta \mu_i}{\delta \beta}\right)^T V_i^{-1}\left(Y_i - \mu_i(\beta)\right) = 0$$

Here, $\frac{\delta \mu_i}{\delta \beta}$ is a $n_i\ x\ p$ matrix whose bth row corresponds to $\frac{\mu_{ib}}{\beta^T}$, V_i^{-1} is an $n_i\ x\ n_i$ matrix and $Y_i - \mu_i$ are $n_i\ x\ 1$ vectors.

The fitting algorithm is defined by the following steps:[93,137]
1. An initial β is estimated under the assumption of an independent intracluster correlation (independent working correlation matrix). By this assumption, the GEE logistic regression reduces to an ordinary logistic regression.
2. Based on the standardized residuals and the assumed working correlation matrix R, α_1 is estimated at the current β.
3. The estimated working covariance matrix V_i is evaluated at α_1. β_1 can be updated while keeping R(α_1) fixed.
Step 2 and 3 are iterated until convergence.

The covariance matrix of $\hat{\beta}$, and thereby the standard errors of the regression coefficients, are estimated with the covariance sandwich estimator Σ_e.[82,87,137]

$$\Sigma_e = I_0^{-1} I_1 I_0^{-1}$$

Where $I_0 = \sum_{i=1}^{K} \left(\frac{\delta \mu_i}{\delta \beta}\right)^T V_i^{-1} \left(\frac{\delta \mu_i}{\delta \beta}\right) = 0$ and $I_1 = \sum_{i=1}^{K} \left(\frac{\delta \mu_i}{\delta \beta}\right)^T V_i^{-1} cov(Y_i) V_i^{-1} \left(\frac{\delta \mu_i}{\delta \beta}\right) = 0$

Here, β, α and ϕ are replaced by estimates and $cov(Y_i)$ is estimated by$(Y_i - \mu_i(\hat{\beta})(Y_i - \mu_i(\hat{\beta})^T$.[137,138] Especially in large-sample settings this estimator is robust, even if the working correlation matrix is mis-specified ($cov(Y_i) \neq V_i^{-1}$). The approach might, however, lead to underestimation of the variances of the regression coefficients in the case of a low sample size

and rare events. Over the couple of years, corrections have been proposed for these situations, but literature does not universally agree upon a solution.[112]

The Wald test statistic is used to compare two nested models (including and excluding a selected predictor variable). The Wald test statistic has a chi-square distribution and takes into account the robust sandwich estimate of the covariance matrix of $\hat{\beta}$. A significant result indicates that the parameter has a significant relation to the outcome variable.[138,140] The regression coefficients have the same population-averaged interpretation as the regression coefficients obtained from a standard logistic regression.[86]

The Akaike Information Criteria (AIC) is widely known for its ability to determine the relative quality of a statistical model for a certain dataset and can be used as a criterion for model selection (e.g. in a logistic regression model). The AIC is defined as:[82]

$$AIC = -2\,LL + 2p$$

In this equation, *LL* is the log-likelihood of the fitted model and *p* represents the number of parameters in de model. The quasi-likelihood under the independence model criterion (QIC) is based on the AIC and can be used to compare the fit of the GEE logistic regression model that assumed different working correlation matrices. In the QIC, the log-likelihood function is changed into the quasi-likelihood function and the penalty term is altered.

$$QIC = -2\,QL\left(\hat{\beta}(R); I\right) + 2trace\left(\widehat{\Omega}^{-1}\hat{V}_R\right)$$

In this equation, *QL* is the quasi-likelihood constructed under the working independence model. In the case of a binary outcome it is often specified that $QL = Y\,ln\left(\mu/(1-\mu)\right) + ln(1-\mu)$ and $\hat{\mu} = g\left(x\hat{\beta}\right)$. The function g( ) is the inverse link function. [82,141] The estimate $\hat{\beta}$ and the covariance estimator $\hat{V}_R$, are obtained from the model with the assumed working correlation matrix $R(\alpha)$. Empirical estimator $\widehat{\Omega}^{-1}$ is the negative Hessian (square matrix of second order derivatives) of $QL\left(\hat{\beta}; I\right)$ under the independence working correlation matrix.[82,141,142] The model with the smallest QIC value is preferred.[79,82,141]

When the GEE method is used as an extension to the logistic regression model, the following assumptions need validation: (1) the predictor variables should be linearly related to the logit of the outcome[143] (Appendix C2.3) (2) correlation between observations of the same cluster is allowed, however, there should not be a dependency between the observations across clusters,[79,143] (3) the number of clusters should not be too low (not less than 10).[143]

Appendix C: Results

1. Spatial analysis

The maximum value in the heatmap of *necrotic core* is 7 (Figure 1). Based on the colour distribution, the necrotic core showed a high variety in sizes and shapes. Large areas, as well as many small and filamentous areas are observed. Calcification often occurred in the necrotic core, transforming almost the complete necrotic core into calcium. The filamentous and small shapes indicate (the part of) the necrotic core that was not calcified. The larger areas showed larger uncalcified necrotic cores, or necrotic cores that were calcified to a lower extent.

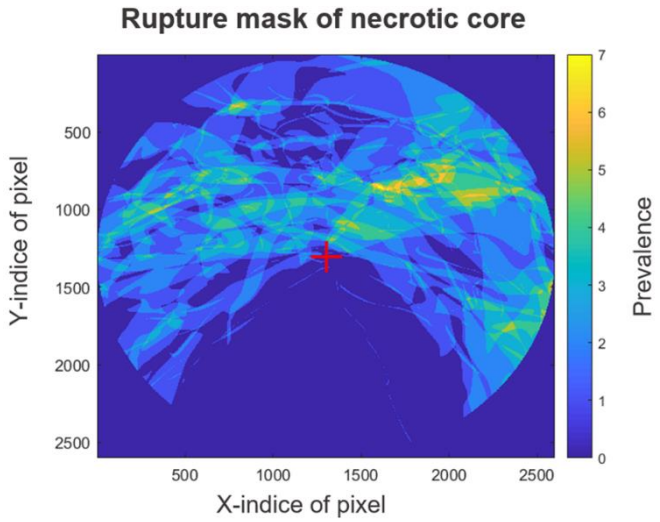


Figure 1 Rupture heatmap of the variable *necrotic core* (n=23)

2. Statistical analysis

2.1 HISTOGRAMS

The variable histograms of the 168 observations can be found in Figure 2. All variables, except for *plaque*, were expressed as the percentage relative to the plaque area in the bin. *Plaque* was expressed as the area in the bin measured by the number of pixels. It can be observed that the distribution of all variables, except *plaque* was fairly right-skewed. Additionally, all variables, except *plaque* show measurements that are equal to zero, indicating that the tissue class was not present in the bin. In order to increase the chance of obtaining useful output from the principle component analysis[128], as well as to meet the linearity assumption of the GEE logistic regression more closely[79], all variables, except *plaque*, were transformed by a $log(x+1)$ transformation. This transformation is often used for data that is right skewed and that contains zero values.[144] Here, x represents the value of the respective variable. The result of this transformation can be seen in Figure 3. The transformation led to a slightly "improved" distribution, because the distribution more closely resembled a normal distribution. The measurements equal to zero, evidently, still had a value of zero after the transformation.

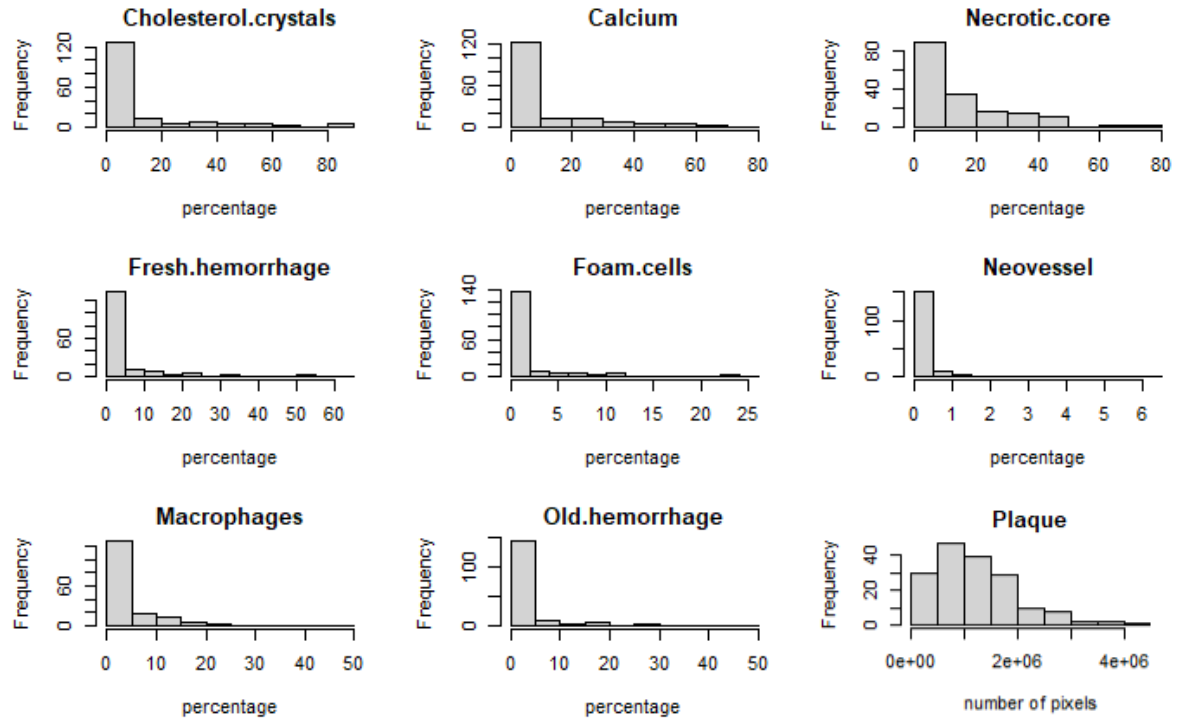


Figure 2 Histograms of the variables measured in the digitized tissue sections (n=168 bins). All variables, except *plaque*, are expressed as the percentage relative to the plaque area in the bin. The x-axis indicates the percentage, the y-axis indicates the frequency of the measurement. In the case of plaque, the x-axis shows the number of pixels present in the bin. The variable *plaque* is expressed as the number of pixels.

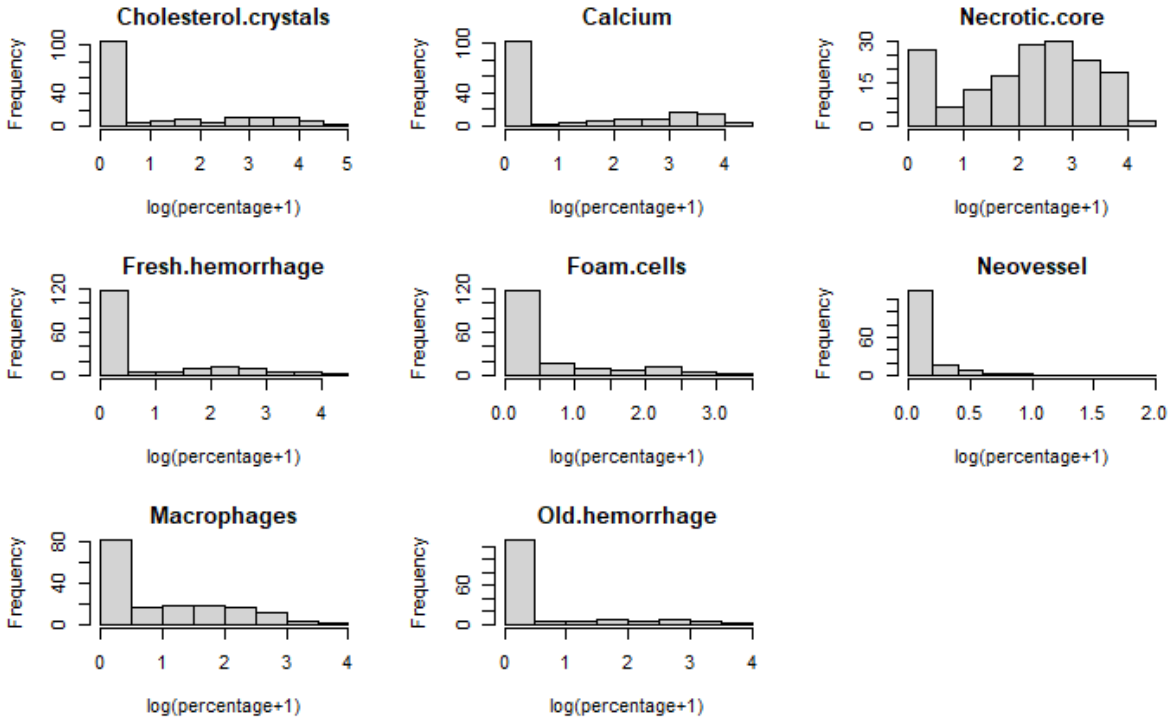


Figure 3 Histograms of the variables measured in the digitized tissue sections (n=168 bins). All variables, except *plaque,* are expressed as the log(x+1) transformation of the percentage relative to the plaque area in the bin. The x-axis indicates the (transformed) percentage, the y-axis indicates the frequency of the measurement.

2.2 PRINCIPAL COMPONENT ANALYSIS

The results of the principle component analysis can be found in Table 1. For this analysis, the *log(x+1)*-transformed variables (except for *plaque*) were used. Each of the variables contributes to every principle component, as indicated by the non-zero variable loadings. The variables do not show obvious clusters throughout the data. The first principal component (PC1) explains most of the variance in the data, namely 24.4%. PC2, PC3, PC4, PC5, PC6 and PC7 explain 21.5%, 17.3%, 12.5%, 10.5%, 8.9% and 4.9% of the variance, respectively. This is shown in Figure 4. All PCs contributed a fair amount to the explanation of the data variances. It was not evident which variable clusters could be formed in order to reduce the dimensionality of the data.

Table 1 Outcome of the principle component analysis. The variable loadings of the principle components (PCs) are indicated below heading PC1…PC7. The standard deviations of the PCs are found in the row titled "standard deviation". "Proportion of variance" indicates the proportion of the variance in the data explained per principal component. In the "cumulative proportion" row the proportion of variances are summed.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Log(Neovessels +1)	0.27	-0.42	0.45	-0.32	0.25	-0.59	0.19
Log(Foam cells +1)	0.22	-0.34	0.30	0.83	-0.11	0.12	0.15
Log(Macrophages +1)	0.52	-0.22	-0.074	-0.22	0.46	0.64	-0.061
Plaque	0.52	0.34	0.33	-0.046	-0.37	-0.07	-0.60
Log(Cholesterol crystals+1)	0.53	0.38	-0.31	-0.0017	-0.20	-0.13	0.65
Log(Calcium+1)	-0.23	0.28	0.69	-0.23	-0.13	0.41	0.39
Log(Necrotic core+1)	-0.020	0.57	0.14	0.31	0.72	-0.18	-0.056
Standard deviation	1.31	1.23	1.10	0.94	0.86	0.79	0.59
Proportion of variance	0.24	0.22	0.17	0.13	0.10	0.089	0.049
Cumulative proportion	0.24	0.46	0.63	0.76	0.86	0.95	1.00

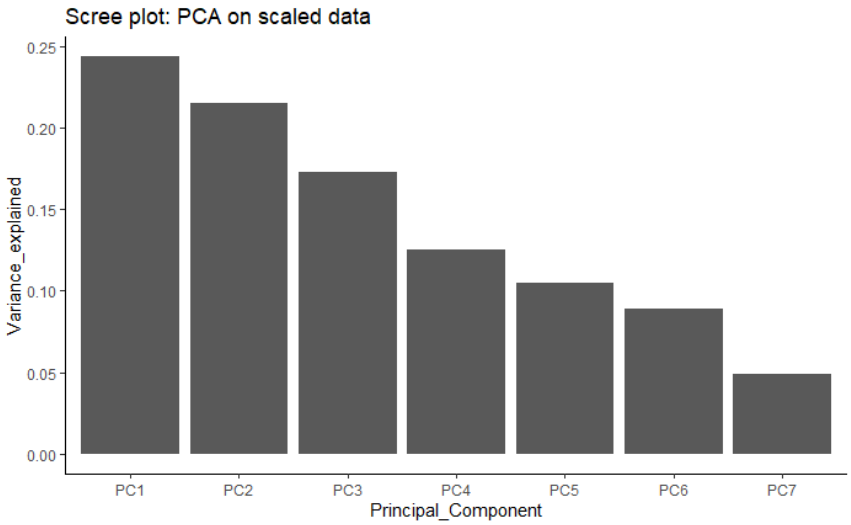


Figure 4 This plot shows the proportion of the variance explained per principal component (PC). If all bars are added, the height of the bar equals 1.

2.3 GEE LOGISTIC REGRESSION

The GEE logistic regression model, just like a normal logistic regression, assumes that the predictor variables are linearly related to the logit of the outcome variable. Scatterplots can indicate whether the assumption is validated.

Figure 5,6 and 7 show scatterplots in which the relation between the predictor variables and the predicted logit of the outcome can be studied. Figure 5 shows the scatterplots belonging to the GEE logistic regression (working correlation matrix: independence) with untransformed variables *calcium, cholesterol crystals, macrophages*, *necrotic core, neovessels* and *plaque*. The fit of this model, expressed in QIC was 137.29. Figure 6 shows the scatterplots belonging to the GEE logistic regression (working correlation matrix: exchangeable) with log(x+1)-transformed variables *calcium, cholesterol crystals, macrophages*, *necrotic core, neovessels* and untransformed variable *plaque*. The fit of this model, expressed in QIC was 136.42. Lastly, Figure 7 shows the scatterplots belonging to the GEE logistic regression (working correlation matrix: independence) with log(x+1)-transformed variables calcium, cholesterol crystals, macrophages, necrotic core, neovessels and untransformed variable plaque (Pl). The fit of this model, expressed in QIC was 134.37.

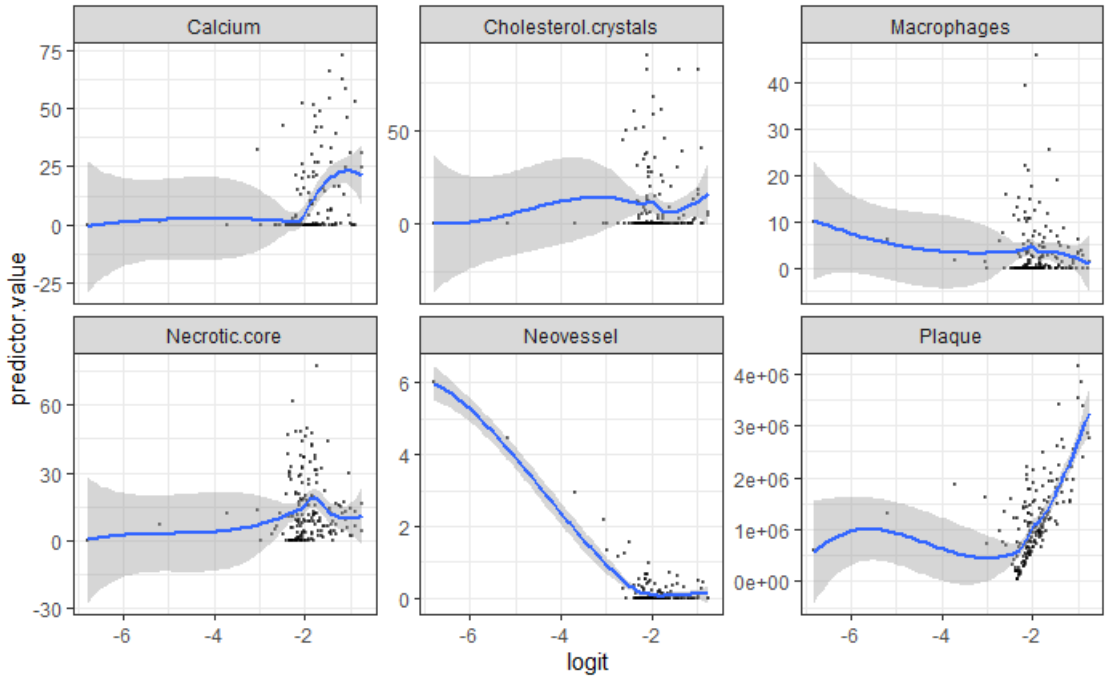


Figure 5 The scatter plots show the relation between the predictor values and the logit of the predicted outcome (rupture/no rupture) in the GEE logistic regression model (working correlation matrix: independence). The model included untransformed variables *calcium, cholesterol crystals, macrophages, necrotic core, neovessels* and *plaque*.

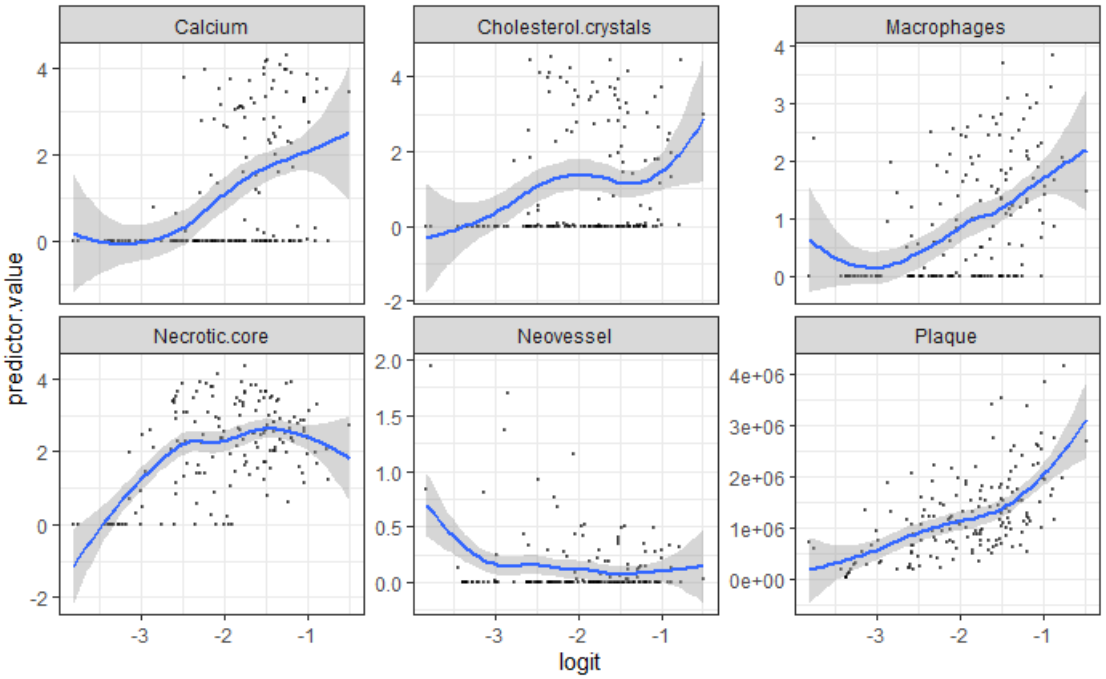


Figure6 The scatter plots show the relation between the predictor values to the logit of the outcome (rupture/no rupture) in the GEE logistic regression model (working correlation matrix: exchangeable). This model included log(x+1)-transformed variables *calcium, cholesterol crystals, macrophages, necrotic core, neovessels* and untransformed variable *plaque*.

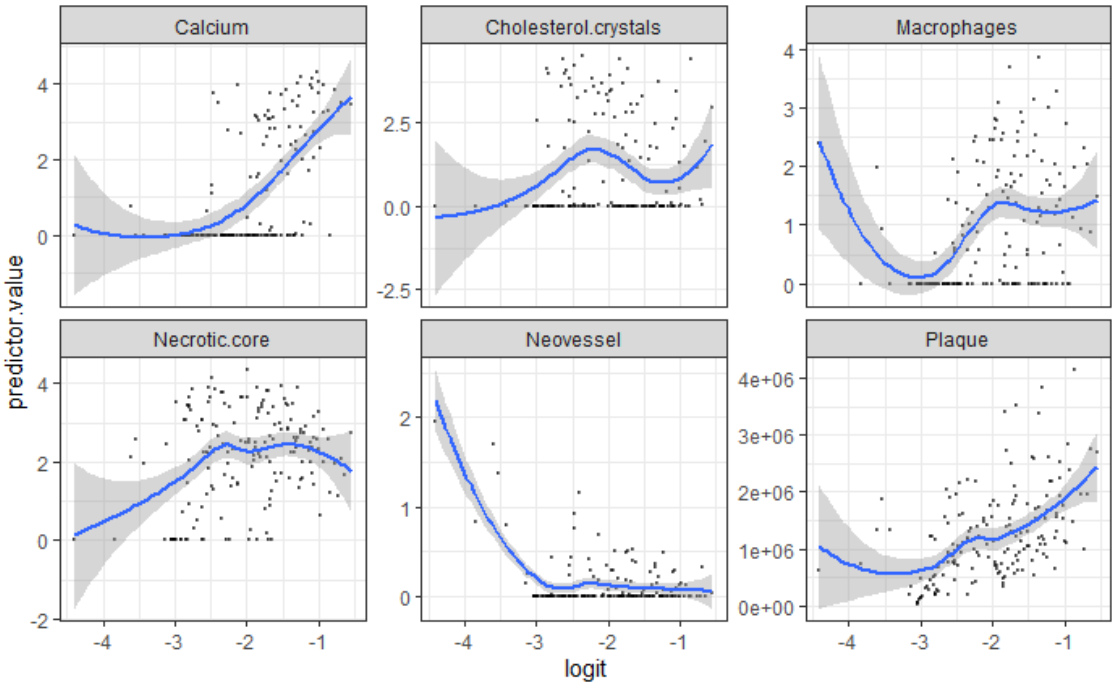


Figure 7 The scatter plots show the relation between the predictor values to the logit of the outcome (rupture/no rupture) in the GEE logistic regression model (working correlation matrix: independence). This model included log(x+1)-transformed variables *calcium, cholesterol crystals, macrophages, necrotic core, neovessels* and untransformed variable *plaque*.

3. Sensitivity analysis

3.1 2-BIN ANALYSIS

The outcome of two-sided tests for the 2-bin analysis can be found in Table 2. The distributions of the variables *plaque, macrophages* and *fresh haemorrhage* were significantly different in the ruptured bins and the non-ruptured bins. These parameters tended to take on higher values in the ruptured bins.

Table 2 Outcome of two-sided clustered Wilcoxon Rank Sum test (rgl method) for the 2-bin analysis. Group 1: ruptured bins (n=19), group 2: non-ruptured bins (n=19). Clusters are defined at the patient level. The Z-value is the test statistic. Significant p-values ($\leq$0.05) are indicated with *.

Variables	Z-value	P-value
Cholesterol Crystals	0.84	0.40
Plaque	2.50	0.012 *
Calcium	-0.25	0.80
Necrotic core	-0.28	0.78
Fresh hemorrhage	3.75	0.00018 *
Foam Cells	1.30	0.20
Neovessels	1.89	0.059
Macrophages	2.01	0.044 *
Old hemorrhage	1.87	0.062

The results of the GEE logistic regression analysis with the exchangeable working correlation matrix are shown in Table 3. *Plaque* is significantly associated with the rupture. The QIC of this model equalled 49.14. The estimated intracluster correlation parameter alpha is -0.32.

Table 3 Outcome of 2-bin GEE logistic regression model, working correlation matrix: exchangeable. "Estimate" indicates the parameter estimates, "std.error" indicates the standard error of the parameter, estimated with the robust (sandwich) estimator. "Wald" indicates the Wald test statistic. "Pr(>|W|)" indicates the corresponding p-value. A p-value $\leq$0.05 is considered significant, indicated with *.

Variables	Estimate	Std. error	Wald	Pr(>|W|)
Intercept	-4.02	2.30	3.07	0.080
Log(Cholesterol crystals+1)	0.32	0.19	2.93	0.087
Plaque/10^5	0.033	0.017	3.87	0.049 *
Log(Calcium+1)	0.28	0.18	2.45	0.12
Log(Necrotic core+1)	0.40	0.41	0.92	0.34
Log(Neovessels +1)	2.13	1.17	3.30	0.069
Log(Macrophages +1)	0.076	0.20	0.14	0.70

The QIC of the model with the independence working correlation model equalled 52.72 and was therefore not preferred.

3.2 4-BIN ANALYSIS

The outcome of two-sided tests for the 4-bin analysis can be found in Table 4. The distributions of the variables *plaque* and *fresh haemorrhage* were significantly different in the ruptured bins and the non-ruptured bins. These parameters tended to take on higher values in the ruptured bins.

Table 4 Outcome of clustered two-sided Wilcoxon Rank Sum test (rgl method) for the 4-bin analysis. Group 1: ruptured bins (n=23), group 2: non-ruptured bins (n=61). Clusters are defined at the patient level. The Z-value is the test statistic. Significant p-values (≤0.05) are indicated with *.

Variables	Z-value	P-value
Cholesterol Crystals	1.26	0.21
Plaque	2.03	0.042 *
Calcium	1.69	0.091
Necrotic core	-0.079	0.94
Fresh hemorrhage	5.68	1.36×10^{-8} *
Foam Cells	1.85	0.065
Neovessels	1.70	0.089
Macrophages	1.59	0.11
Old hemorrhage	1.71	0.088

The results of the GEE logistic regression analysis with the independence working correlation matrix are shown in Table 5. *Plaque* and *calcium* were associated with rupture. The QIC of this model equalled 96.70.

Table 5 Outcome of 4-bin GEE logistic regression model, working correlation matrix: independence. "Estimate" indicates the parameter estimates, "std.error" indicates the standard error of the parameter, estimated with the robust (sandwich) estimator. "Wald" indicates the Wald test statistic. "Pr(>|W|)" indicates the corresponding p-value. A p-value ≤0.05 is considered significant, indicated with *.

Variables	Estimate	Std. error	Wald	Pr(>|W|)
Intercept	-2.30	0.74	9.57	0.0020 *
Log(Cholesterol crystals+1)	-0.10	0.21	0.23	0.63
Plaque/10^5	0.036	0.018	4.22	0.040 *
Log(Calcium+1)	0.28	0.070	16.16	5.80×10^{-5} *
Log(Necrotic core+1)	-0.09	0.27	0.11	0.74
Log(Neovessels +1)	-1.55	0.99	2.47	0.12
Log(Macrophages +1)	0.47	0.29	2.59	0.11

The QIC of the model with the exchangeable working correlation model equalled 100.25 and was therefore not preferred. The estimated correlation parameter alpha was -0.10.

3.3 16-BIN ANALYSIS

The outcome of two-sided tests for the 16-bin analysis can be found in Table 6. The distributions of the variables *calcium* and *fresh haemorrhage* were significantly different in the ruptured bins and the non-ruptured bins. These parameters tended to take on higher values in the ruptured bins.

Table 6 Outcome of two-sided clustered Wilcoxon Rank Sum test (rgl method) for the 16-bin analysis. Group 1: ruptured bins (n=23), group 2: non-ruptured bins (n=313). Clusters are defined at the patient level. The Z-value is the test statistic. Significant p-values ($\leq$0.05) are indicated with *.

Variables	Z-value	P-value
Cholesterol Crystals	0.79	0.43
Plaque	1.63	0.10
Calcium	2.28	0.022 *
Necrotic core	0.31	0.76
Fresh hemorrhage	7.37	1.75×10^{-13}*
Foam Cells	-0.029	0.98
Neovessels	-0.031	0.98
Macrophages	0.96	0.34
Old hemorrhage	0.85	0.39

The results of the GEE logistic regression analysis with the independence working correlation matrix are shown in Table 7. *Calcium* was significantly associated with rupture. The QIC of this model equalled 170.17.

Table 7 Outcome of 16-bin GEE logistic regression model, working correlation matrix: exchangeable. "Estimate" indicates the parameter estimates, "std.error" indicates the standard error of the parameter, estimated with the robust (sandwich) estimator. "Wald" indicates the Wald test statistic. "Pr(>|W|)" indicates the corresponding p-value. A p-value $\leq$0.05 is considered significant, indicated with *.

Variables	Estimate	Std. error	Wald	Pr(>|W|)
Intercept	-3.06	0.37	67.09	2.20×10^{-16} *
Log(Cholesterol crystals+1)	0.0095	0.17	0.00	0.96
Plaque/10^5	0.058	0.040	2.12	0.15
Log(Calcium+1)	0.30	0.096	9.94	0.0016*
Log(Necrotic core+1)	-0.19	0.19	1.06	0.30
Log(Neovessels +1)	-1.23	1.20	1.05	0.31
Log(Macrophages +1)	0.19	0.28	0.45	0.50

The QIC of the model with the independence working correlation model equalled 171.71 and was therefore not preferred. The estimated correlation parameter alpha was -0.031.

Appendix D: Python, R & MATLAB scripts

In addition to the tissue structures mentioned in the main text of this thesis, two other tissue structures were segmented in the MeVisLab software: *Desorientated SMC*" and *"Loose Fibrous Tissue Macrophages"*. The Python- and R scripts include these tissue structures and in the Histology Atlas they can be found as well. The tissue structures were, however, not included in the statistical analyses, because they were present in less than 10 of the 168 observations.

1. Python scripts

1.1 HISTOLOGYSECTION

The function "HistologySection" is provided below. This function extracts the locations of the different tissue structures from the NIFTII files that were a product of the segmentation in MeVisLab. Binary masks, which had the same size of the original image, were created per tissue structure. The pixel of the mask was given a value of 1 if the tissue structure was present at the location, otherwise the pixel value remained zero.

Multiple tissue structures could be present at the same location, i.e. the same pixel. In some cases, the assignment of multiple tissue structures to pixels with similar indices was not preferred. Therefore, priority rules were added that enabled the adjustment of the binary masks if certain combinations of tissue structures were present:
Example 1: The pixels of the *plaque* mask with value 1 were set to 0 if the *lumen* mask contained a 1 at the pixels with similar indices. *Plaque* was actually segmented as the outer wall and the *lumen* had to be extracted from the *plaque* binary mask in order to represent *plaque.*
Example 2: The pixels of the *necrotic core* mask and *cholesterol crystal*s mask with value 1 were set to 0 if the *calcium* mask contained a 1 at the pixels with similar indices. Calcium was often recognizable as calcified necrotic core or calcified cholesterol crystals.

Additional priority rules can be found in the script. Moreover, the "HistologySection" function includes the binning procedure. It also determines the number of pixels belonging to the different tissues structures per bin.

```
# -*- coding: utf-8 -*-
"""
Created on Tue Jul 28 17:55:42 2020

@author: Suze-Anne Korteland, Rachel Oversier, Adriana Lentz
"""


import SimpleITK as sitk
import os
import numpy as np
from enum import Enum
import re
import cv2
import matplotlib.pyplot as plt
```

```
import matplotlib as mpl
from scipy.spatial import cKDTree
#from skimage.feature import canny
import math
import logging
from PIL import Image
from io import BytesIO
from scipy import sparse

# set font for matplotlib
mpl.rcParams['font.family'] ='Arial'

#test = []

class HistologySection:
   # use an enum to represent the different tissue types
   class Tissue(Enum):
    # assign Enum(keyword, number, description, colour, drawing priority)
      """ Enum representing the different tissue types. """

      LIPID = 0, 'Cholesterol crystals', '#e0ff17', 6
      PLAQUE = 1, 'Plaque', '#ffa1c0', 2
      LUMEN = 2, 'Lumen', '#e6fffc', 10
      CALC = 3, 'Calcium', '#911eb4', 8
      CORE = 4, 'Necrotic core', '#1ae82c', 8
      HEM = 5, 'Fresh hemorrhage', '#f20202',8
      FOAM = 6, 'Foam cells', '#0de0d6', 8
      ART = 7, 'Artefact', '#ffa500' ,8
      RUP = 8, 'Rupture', '#050000', 50
      NEO = 9, 'Neovessel', '#a9a9a9', 8
      EXTRA = 10, 'Extra, not used here', '#ffffff',8
      MP = 11, 'Macrophages', '#1732bd',50
      SMC = 12, 'Inflamed and disorientated smooth muscle cells', '#4dab49', 8
      OLDHEM = 13, 'Old hemorrhage', '#f032e6',8
      LFT = 14, 'Loose fibrous tissue with macrophages', '#237d2b', 8



      def __new__(cls, code, title, color, priority):
         obj = object.__new__(cls)
         obj._value_ = code
         obj.title = title
         obj.color = color
         obj.priority = priority
         return obj

      def __str__(self):
         return self.title

      @classmethod
      def get_sorted_valuelist(cls, reverse=False):
         """ Return list of the Tissues, sorted by value. """
         return sorted([t.value for t in cls], reverse=reverse)

      @classmethod
      def get_sorted_titles(cls, reverse=False):
         """ Return list of the Tissues, sorted by value. """
         return sorted([t.title for t in cls], reverse=reverse)

      @classmethod
      def get_colorlist(cls):
         """ Return list of Tissue colors. """
         return [t.color for t in cls]
```

```
  def __init__(self, image_path, segmentation_path, nbins=1):
    """
    Constructor for the Histology Section class.

    Parameters:
      image_path (str):
        Full path to the histology image file. The image should be in NIFTI format.
        The image filename should contain a substring with the section nr of the image,
        preceded by the letter 'B'.
        For example: 'P1_B02_S12_EvG - 2018-01-12 11.39.42_map2_original_registered.nii'
        T
      segmentation_path (str):
        Full path to the segmentation corresponding to the histology image.
        The segmentation image should be in NIFTI format.
        The segmentation image filename should contain a substring with the
        section nr. of the image, preceeded by the letter 'B'.
        For example: 'P1_B10_S12_EvG_map2_seg_original_registered.nii'

      nbins (int):
        The number of bins to use for analyzing the segmentation (default = 1).

    """
    # Create logger.
    self.logger = logging.getLogger(__name__)
    # Get the section nr from the image path and segmentation path.
    searchobj_img = re.search('(?:B)(\d+)', image_path)
    searchobj_seg = re.search('(?:B)(\d+)', segmentation_path)
    if searchobj_img and searchobj_seg:
      # Get the slice id from the filenames.
      slice_id_img = int(searchobj_img.group(1)) -1 #subtract 1 to start at slice 0
      slice_id_seg = int(searchobj_seg.group(1)) -1 #subtract 1 to start at slice 0
      if slice_id_img == slice_id_seg:
        # ok!
        self.framenr = slice_id_img
      else:
        # The provided images do not correspond, they do not have the same frame nr.
        self.logger.error('Histology image and segmentation have different section numbers')
        ValueError('Histology image from section {} and segmentation from section {} do not
correspond'.format(slice_id_img, slice_id_seg))
    else:
      self.logger.error('Incorrect filename format for file {} or file {}'.format(image_path, segmentation_path))
      raise ValueError('Filenames should contain the character B followed by a number to indicate the section number')

    # img_extension = os.path.splitext(image_path)[1]
    # seg_extension = os.path.splitext(segmentation_path)[1]
    self.ImgName = image_path[-27:-8]
    # Load the image.
    self.image = sitk.ReadImage(image_path)
    # Load the segmentation.
    self.segmentation = sitk.ReadImage(segmentation_path)

    # Check image and segmentation have the same voxel dimensions in x,y.
    if self.image.GetSize()[0:2] != self.segmentation.GetSize()[0:2]:
      self.logger.error('Histology image and segmentation have different dimensions')
      raise ValueError('Histology image and segmentation have different dimensions')



    # Construct voxel to world matrix.
    self.voxtoworld = self._construct_voxtoworld(self.image)

    # Create segmentation masks for plotting and analysis.
```

```
    self.segmentationmasks = self._construct_masks(self.segmentation)

    # Center of image,  take the center of the voxels.
    self.center = self._get_img_middle(self.image)
    # The number of radial bins to subdivide the segmentation into.
    self.nbins = nbins
    # Initialize bin grouping.
    self.bin_grouping = self._bin_pixels()
    # Initialize plaque_thickness params to None (expensive to compute, so only compute it
    # the first time it is needed).
    self.plaque_thickness_mean = None
    self.plaque_thickness_min = None
    self.plaque_thickness_max = None
    self.plaque_thickness_sd = None
    # Initialize distance to lumen to None (expensive to compute, so only
    # compute it the first time it is needed).
    self.dist_to_lumen = None

  def _get_img_middle(self,image):
    isLUMEN = self.Tissue.LUMEN.value
    print(isLUMEN)
    LUMEN_array = np.transpose(self.segmentationmasks[isLUMEN])
    print(LUMEN_array)
    try:
      y, x = np.nonzero(LUMEN_array)
    except:
      y, x, z = np.nonzero(LUMEN_array)

    x_center = np.sum(x)/len(x)
    y_center = np.sum(y)/len(y)

    #plt.figure()
    #plt.imshow(LUMEN_array)
    #plt.scatter(x_center, y_center, s=20, c='b', marker='o')

    LUMEN_middle = [x_center, y_center, 0.5]

    return LUMEN_middle

  def _construct_voxtoworld(self, img):
    # Construct 4x4 voxel to world matrix from the sitk image data.
    dim = img.GetDimension()
    if dim == 2:
      # Generate the 3rd dimension.
      D = np.eye(4)
      # Construct world-to-vox matrix from image info
      # direction matrix.
      D[:2,:2] = np.array(img.GetDirection()).reshape((2,2))
      # Get spacing.
      s = np.diag([img.GetSpacing()[0], img.GetSpacing()[1],1,1])
      # Get origin (add z-direction, origin is at -0.5).
      origin = np.array([img.GetOrigin()[0],img.GetOrigin()[1],0.5])
    elif dim == 3:
      # Construct world-to-vox matrix from image info
      # direction matrix.
      D = np.eye(4)
      D[:3,:3] = np.array(img.GetDirection()).reshape((3,3))
      # Get spacing.
      s = np.diag([img.GetSpacing()[0],img.GetSpacing()[1],img.GetSpacing()[2],1])
      # Get origin.
      origin = np.array(img.GetOrigin())
    elif dim == 4:
      # Construct world-to-vox matrix from image info
```

```
        # direction matrix.
        D = np.array(img.GetDirection()).reshape((4,4))
        # Get spacing.
        s = np.diag([img.GetSpacing()[0], img.GetSpacing()[1],img.GetSpacing()[2],1])
        # Get origin.
        origin = np.array(img.GetOrigin())[:3]
    else:
        self.logger.error('Invalid data dimension')
        raise ValueError('Invalid data dimension of {} when loading histology image data'.format(dim))

    # Construct vox-to-world matrix.
    W = D.dot(s)
    W[:3,3] = origin
    return W

def get_imagedata(self,img):
    """
    Get a numpy array from a SimpleITK image

    Args:
        img (SimpleITK image): the simpleITK image for which you want to get the image data.

    Returns:
        numpy array with the image data. The order of the returned array is [x,y],
        or in the case of color data, [x,y,RGB color].

    """
    # Copy the data so it can be manipulated without changing the source.
    img_data = np.copy(sitk.GetArrayFromImage(img))

    # Swap some axis as sitk and numpy axes order is different.
    if img_data.ndim == 2:
        # 2d data.
        # Swap axis, numpy image data is display y,x.
        img_data = np.swapaxes(img_data,0,1)
        # Select 2d slice.
        img_slice = img_data
    elif img_data.ndim == 3:
        # 3d data.
        # Swap axis, numpy image data is display z,y,x.
        img_data = np.swapaxes(img_data,0,2)
        # Select 2d slice.
        img_slice = img_data[:,:,0]
    elif img_data.ndim == 4:
        # 3d data + RGB color
        # swap axis,  numpy img data is c, z, y, x
        # we want x,y,z,c
        img_data = np.swapaxes(img_data,1,2)
        img_data = np.swapaxes(img_data,0,3)
        #select 2d slice
        img_slice = img_data[:,:,0,:]
    else:
        self.logger.error('Invalid data dimension')
        raise ValueError('Invalid data dimension  of {} when loading histology image data'.format(img_data.ndim))
    return img_slice

def _construct_masks(self, segmentation):
    # Separate the segmentation image into masks, one for each tissue type.
    # This is useful in plotting and analysis. The masks are binary images with
    #0's and 1's. 1 indicates the tissue type is present, 0 indicates it is absent.
    masks = {}
    # Split the segmentation data into masks.
```

```
    seg_data = self.get_imagedata(segmentation)
    # The segmentation is split through a right shift + subtraction.
    # For the right shift + subtraction to work, the loop over the segmentation types
    # needs to be in descending order.
    tissue_list = self.Tissue.get_sorted_valuelist(reverse=True)
    tissue_list_names = ["LFT", "OLDHEM", "SMC", "MP", "EXTRA", "NEO", "RUP", "ART", "FOAM", "HEM",
            "CORE", "CALC", "LUMEN", "PLAQUE", "LIPID"]

    tissue_list_names.reverse()

    for i in tissue_list[0:]:
        # Shift bits to the right (equivalent to dividing by 2**i).
        # If the current type is present in a pixel, right shifting i times
        # will return 1.
        masks[i] = np.right_shift(seg_data,i)
        # Subtract the current tissue type value from the image data.
        seg_data[masks[i]>0] = seg_data[masks[i]>0] - 2**i
        # Set PLAQUE pixels to zero if they are also LUMEN
        if i == self.Tissue.PLAQUE.value:
            islumen = np.logical_and(masks[self.Tissue.PLAQUE.value], masks[self.Tissue.LUMEN.value])
            masks[self.Tissue.PLAQUE.value][islumen] = 0

    # Additional priority rules
    for i in tissue_list[0:]:
        # priority rules for Artefact incorporation
        if i == self.Tissue.ART.value and i != self.Tissue.PLAQUE.value:
            for x in tissue_list:
                if x != self.Tissue.ART.value and x != self.Tissue.PLAQUE.value:
                    # controle
                    name = tissue_list_names[x]
                    try:
                        isboth = np.logical_and(masks[self.Tissue.ART.value], masks[x])
                        masks[x][isboth] = 0
                    except:
                        print("No " + name + " in this mask")
                else:
                    pass
        # Artefact should not have priority over plaque, extra check.
        elif i == self.Tissue.PLAQUE.value:
            isboth = np.logical_and(masks[self.Tissue.ART.value], masks[self.Tissue.PLAQUE.value])
            masks[self.Tissue.PLAQUE.value][isboth] = 1
        # Cholesterol crystals have priority over the necrotic core (they often are next to each other).
        elif i == self.Tissue.LIPID.value:
            isboth = np.logical_and(masks[self.Tissue.LIPID.value], masks[self.Tissue.CORE.value])
            masks[self.Tissue.CORE.value][isboth] = 0
        # Macrophages and Foam cells are often close. The borders of the foam cells are better visible.
        elif i == self.Tissue.FOAM.value:
            isboth = np.logical_and(masks[self.Tissue.FOAM.value], masks[self.Tissue.MP.value])
            masks[self.Tissue.FOAM.value][isboth] = 0
        # Areas with Macrophages and Foam cells contain neovessels. These should be removed from the NEO/FOAM area
        elif i == self.Tissue.NEO.value:
            isboth1 = np.logical_and(masks[self.Tissue.NEO.value], masks[self.Tissue.FOAM.value])
            isboth2 = np.logical_and(masks[self.Tissue.NEO.value], masks[self.Tissue.MP.value])
            masks[self.Tissue.FOAM.value][isboth1] = 0
            masks[self.Tissue.MP.value][isboth2] = 0
        # Blood in the necrotic core can age. Because this can be distinguished we classify them separately.
        elif i == self.Tissue.OLDHEM.value:
            isboth = np.logical_and(masks[self.Tissue.OLDHEM.value], masks[self.Tissue.CORE.value])
            masks[self.Tissue.CORE.value][isboth] = 0
        # Calficiation is often calcified necrotic core or calcified cholesterol crystals. However, we segment it as calcification
        elif i == self.Tissue.CALC.value:
            isboth1 = np.logical_and(masks[self.Tissue.CALC.value], masks[self.Tissue.CORE.value])
            isboth2 = np.logical_and(masks[self.Tissue.CALC.value], masks[self.Tissue.LIPID.value])
```

```
            masks[self.Tissue.CORE.value][isboth1] = 0
            masks[self.Tissue.LIPID.value][isboth2] = 0

    return masks

def _bin_pixels(self):
    # Bin the pixels into nbin bins by assigning a bin group number to each pixel.
    # The bins are generated starting with the first bin aligned along the
    # x-axis of the world coordinate system of the image (specified by
    # the voxtoworld matrix).

    # Create a grid with the voxel indices.
    grid = np.indices(self.get_imagedata(self.image).shape[:2])
    # Convert the grid to a list with voxel coordinates, x,y,z.
    vox_coords = np.column_stack((grid[0].flatten(),
                          grid[1].flatten(),
                          np.ones_like(grid[0].flatten())*self.framenr,
                          np.ones_like(grid[0].flatten())))
    # Get the center of the image.
    #masks =
    vox_center = self.center[:]
    # Compute angle w.r.t. image center for every voxel.
    vox_angle = np.arctan2(vox_coords[:,1] - vox_center[1], vox_coords[:,0] - vox_center[0])
    # Reshape to image dimensions.
    vox_angle = np.reshape(vox_angle, grid[0].shape)

    # Compute the start angle of the first bin, which is the rotation angle
    # of the vox-to-world matrix along the z dimension.
    # To find the angle decompose the vox-to-world affine
    # matrix into a scaling and a rotation (assuming no shear).
    # The entry at voxtoworld[0,0] is scale_x * cos(theta) and the entry
    # at voxtoworld[1,0] is scale_x * sin(theta) so compute
    # theta from voxtoworld[1,0]/voxtoworld[0,0] = sin(theta)/cos(theta) =>
    # theta = arctan(sin(theta)/cos(theta))
    theta = np.arctan2(self.voxtoworld[1,0] , self.voxtoworld[0,0])

    # Rotate the voxel angles to align with world axis, so add theta.
    # Add pi to make the angles in the range [0,2pi], add theta and get the modulus to get
    # the result after adding in the range [0,2pi]. Subtract pi to go back to the range [-pi,pi]
    # NOTE: for historical reasons, rotate in the opposite direction, therefore add theta.
    vox_angle = (vox_angle + math.pi + theta) % (2*math.pi) - math.pi

    bin_grouping = np.zeros(self.get_imagedata(self.image).shape[0:2])
    sample_bin_start = np.zeros(self.nbins)
    sample_bin_end = np.zeros(self.nbins)
    radial_bin = np.linspace(-math.pi, math.pi, self.nbins+1)
    for j in range(self.nbins):
        sample_bin_start[j] = radial_bin[j]
        sample_bin_end[j] = radial_bin[j+1]
        #  Find indices of voxels that fall between the start angle and end angle of bin j.
        group = np.where(np.all([vox_angle >= sample_bin_start[j], vox_angle < sample_bin_end[j]], axis=0))
        # assign bin_grouping at these voxel indices to group j.
        bin_grouping[group] = j
    return bin_grouping

def plot_binning(self, savetofile = True, dirpath='D:/Test/Test3/', tissuelist=None):
    """
    Generate plot of the bins overlaying the histology and segmentation images.

    If savetofile is set to True and a directory path for the output image is specified,
    the image will be saved to file as png and tiff and closed.

    Args:
```

```
        savetofile (bool, optional): boolean specifying whether to save plots to file (True)
                or not (False). Plots are saved as .png and as .tiff in the specified output directory.
                Default is False.
        dirpath (str, optional): full path to the directory where to save the plots if savetofile
                is set to True. Default is ''
        tissuelist (list of Tissue enum types): list of tissue enum types to plot.
              Default is None.

    Files created:
        .png and .tiff images (600 dpi) of the plot. The image will be put in the
        output directory specified in 'dirpath' with a filename of the form:
          'B[SECTIONNR]_[STAINING]_bins.png'

    """
    # Plot the segmentation.
    self.plot_segmentation(tissuelist=tissuelist)
    ax = plt.gca()
    # Overlay a white semi-trasparent layer.
    ax.imshow(np.ones_like(self.bin_grouping).T, alpha=0.4, cmap='Greys', zorder=1000)
    # Compute image rotation and convert to degrees.
    thetadeg = math.degrees(np.arctan2(self.voxtoworld[1,0] , self.voxtoworld[0,0]))
    # Get the number of degrees per bin.
    dtheta = 360/self.nbins
    # Generate patch for each bin
    r = np.max(self.image.GetSize())
    firstbin  = -180 # subtract 180 degrees as the first bin should start on the opposite side
    for i in range(self.nbins):
        # NOTE: for historical reasons, rotate in the opposite direction,
        # so subtract thetadeg.
        binstart = firstbin - thetadeg + i*dtheta
        binend = firstbin - thetadeg + (i+1) * dtheta
        # Draw wedge shaped patch centered at self.center[:2] with radius r that sweeps binstart to binend (in degrees)
        w = mpl.patches.Wedge(self.center[:2], r, binstart, binend, facecolor='none', edgecolor = 'k', zorder=2000)
        ax.add_patch(w)
        # Add text with bin number.
        x = self.center[0] + r/4 * math.cos(math.radians(binstart + 0.5*dtheta))
        y = self.center[1] + r/4 * math.sin(math.radians(binstart + 0.5*dtheta))
        plt.text(x,y,str(i+1), zorder=2000, fontsize=10)
    plt.axis('off')

    if savetofile:
        # Check if path exists.
        if os.path.isdir(dirpath):
            # Save figure to file and close figure.
            png1 = BytesIO()
            plt.savefig(png1,format='png',dpi=600, bbox_inches= 'tight')
            # load image into PIL
            png2 = Image.open(png1)
            # save as tiff
            png2.save(os.path.join(dirpath, 'B{:02d}_{}_bins.tiff'.format(self.framenr+1, self.ImgName)), dpi=(600.0,600.0))
            plt.close()
        else:
            # Incorrect path, cannot save figure.
            self.logger.warning('Incorrect directory path. Cannot save figure.')


def plot_segmentation(self, savetofile = True, dirpath ="D:/Test/Test4/", tissuelist=None):
    """
    Generate plot of the tissue segmentation overlaying the histology image.

    If savetofile is set to True and a directory path for the output image is specified,
    the image will be saved to file as png and closed.
```

```
    Args:
        savetofile (bool, optional): boolean specifying whether to save plots to file (True)
                or not (False). Plots are saved as .png in the specified output directory.
                Default is False.
        dirpath (str, optional): full path to the directory where to save the plots if savetofile
                is set to True. Default is ''.
        tissuelist (list of Tissue enum types, optional): list of tissue enum types to plot. Default is None.

    Files created:
        .png images (300dpi) of the plot. The image will be put in the
        output directory specified in 'dirpath' with a filename of the form:
            'B[SECTIONNR]_[STAINING]_segmentation.png'
    """
    # Get list of colors from the Tissue enum.
    colorlist = self.Tissue.get_colorlist()
    # Add transparent to the colorlist so that background is zero.
    # TODO: it would be better to include this in the tissue type or so, as background...
    # However it is a little complicated because of including zero, while the LIPID type has value 0, since 2**0 = 1.
    colorlist.insert(0,'#ffffff00')
    # Generate colormap from the list of colors.
    cmap = mpl.colors.ListedColormap(colorlist)
    # Define edges of bins for the colormap based on the Tissue segmentation values.
    # Data falling within a bin is mapped to the color with the same index.
    bounds = 0
    bounds = np.append(bounds, 2**np.copy(self.Tissue.get_sorted_valuelist()))
    bounds = np.append(bounds,max(bounds)*2)
    # Subtract 0.5 to make sure the segemntation values fall inside the correct interval.
    bounds = bounds-0.5
    # Generate colormap index based on the bounds.
    norm = mpl.colors.BoundaryNorm(bounds,cmap.N)

    plt.figure()
    plt.title('Block {:02d}'.format(self.framenr+1),fontsize=12)
    show_legend = False
    # Plot the histology image.
    a1 =1 # fully opaque.
    a2 = 0.6 # slightly transparent.
    # Plot gray image.
    plt.imshow(self.get_imagedata(self.image).mean(axis=2).T, alpha=a1, cmap='gray')
    show_legend = True

    # Plot the different tissue segmentation masks.
    colors = []
    types_present = self.get_tissuetypes()
    if tissuelist:
        # Get all the types that are in tissuelist and present in the segmentation.
        # Note: if the tissuelist does not contain any present tissue types
        # returns an empty set and no segmentation will be plotted.
        types_present = list(set(tissuelist).intersection(types_present))

    for tissue in types_present:
        # Get copy of mask for current tissue.
        mask = np.copy(self.segmentationmasks[tissue.value])
        # Replace 1's in mask with 2**(tissue enum code value) to assign correct color
        mask[mask>0] = 2**tissue.value
        plt.imshow(mask.T, cmap=cmap, norm=norm, alpha=a2, zorder=tissue.priority)
        # Get the color of the value and add it to colors list (needed to plot legend).
        colors.append(cmap(norm(2**tissue.value)))

    if show_legend:
        # Add legend, adapted from stackoverflow question
        # https://stackoverflow.com/questions/40662475/matplot-imshow-add-label-to-each-color-and-put-them-in-legend.
        # Create a patch (proxy artist) for every color.
```

```
        patches = [mpl.patches.Patch(color=colors[i], label=types_present[i].title) for i in range(len(types_present)-1,-1,-1)]
        # Put those patched as legend-handles into the legend.
        lgd = plt.legend(handles=patches, bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0. ,fontsize=10)

    plt.axis('off')

    if savetofile:
        # Check if path exists.
        if os.path.isdir(dirpath):
            # Save figure to file and close figure.
            if show_legend:
                plt.savefig(os.path.join(dirpath,
                                         'B{:02d}_{}_segmentation.png'.format(self.framenr+1, self.ImgName)),
                                        dpi=300, bbox_extra_artists=(lgd,),
                                        bbox_inches='tight')
            else:
                plt.savefig(os.path.join(dirpath,
                                         'B{:02d}_{}_segmentation.png'.format(self.framenr+1, self.ImgName)),
                                        dpi=300,
                                        bbox_inches='tight')
            plt.close()

        else:
            # Incorrect path, cannot save figure.
            self.logger.warning('Incorrect directory path. Cannot save figure')

def get_tissuetypes(self):
    """
    Return the tissue types present in the segmentation.

    Tissue is present if in the tissue's mask, any pixel is 1.
    """

    return [self.Tissue(t) for t, m in self.segmentationmasks.items() if 1 in m]

def compute_tissue_area(self, tissue, binid=None, depth=None):
    """
    Compute tissue area (mm^2).

    Args:
        binid (int, optional):
            If specified, compute the tissue area for the radial bin with index
            binid. Default is None. If None, compute tissue area for the
            whole segmentation.
        depth (float, optional): The depth (in mm) into the vessel wall for which to check whether the
            tissue is present.

    Returns:
        Tissue area (mm^2)
    """
    if (depth and (self.dist_to_lumen is None)):
        # dist_to_lumen has not yet been computed, so compute it first.
        self.dist_to_lumen = self.compute_dist_to_lumen()



    # TODO: add check for valid binid number
    pixel_area = self.image.GetSpacing()[0] * self.image.GetSpacing()[1]
    if self.is_tissue_present(tissue, binid,depth):
        if binid is None:
            if depth:
                pixel_count = np.count_nonzero(self.segmentationmasks[tissue.value][self.dist_to_lumen<=depth])
            else:
```

```
                pixel_count = np.count_nonzero(self.segmentationmasks[tissue.value])
            else:
                if depth:
                    mask = np.logical_and(self.segmentationmasks[tissue.value], self.bin_grouping == binid)
                    pixel_count = np.count_nonzero(mask[self.dist_to_lumen<=depth])
                else:
                    pixel_count = np.count_nonzero(np.logical_and(self.segmentationmasks[tissue.value], self.bin_grouping ==
binid))
            tissue_area = pixel_count * pixel_area
        else:
            if binid is None:
                if depth:
                    self.logger.warning('Tissue type {} is not present in segmentation within {} mm from lumen. Cannot compute
area.'.format(tissue.name, binid, depth))
                else:
                    self.logger.warning('Tissue type {} is not present in segmentation. Cannot compute area.'.format(tissue.name,
binid))
            else:
                if depth:
                    self.logger.warning('Tissue type {} is not present in bin {} within {} mm from lumen. Cannot compute
area.'.format(tissue.name, binid, depth))
                else:
                    self.logger.warning('Tissue type {} is not present in bin {}. Cannot compute area.'.format(tissue.name, binid))
            tissue_area = np.nan
        return tissue_area

    def is_tissue_present(self, tissue, binid=None, depth=None):
        """
        Return True if a tissue is present, False if not.

        Args:
            tissue (Tissue enum): The tissue enum type for which to test its presence.
            binid (int, optional): If binid is specified, check the presence of the tissue
            in the radial bin with index binid. Default is None. If None, check the
            presence of the tissue in the entire segmentation.
            depth (float, optional): The depth (in mm) into the vessel wall for which to check whether the
                tissue is present.


        Returns:
            Boolean, True if the tissue is present, False if the tissue is not.
        """
        if (depth and (self.dist_to_lumen is None)):
            # dist_to_lumen has not yet been computed, so compute it first.
            self.dist_to_lumen = self.compute_dist_to_lumen()


        # TODO: add check for valid binid number.
        if tissue.value in self.segmentationmasks:
            if binid is None:
                if depth:
                    return np.any(self.segmentationmasks[tissue.value][self.dist_to_lumen<=depth])
                else:
                    return np.any(self.segmentationmasks[tissue.value])
            else:
                if depth:
                    mask = np.logical_and(self.segmentationmasks[tissue.value], self.bin_grouping == binid)
                    return np.any(mask[self.dist_to_lumen<=depth])
                else:
                    return np.any(np.logical_and(self.segmentationmasks[tissue.value], self.bin_grouping == binid))
        else:
            return False
```

1.2 MASTERSCRIPT

The "masterscript" was created to automatedly process the NIFTII files that were present in the same folder. This script retrieves the output of “HistologSection”. Additionally, this script converts the number of pixels of the tissue structures per bin to the percentage of the tissue structure relative to the plaque size (per bin). The results are written to an excel file. Moreover, an image of the segmentation and an image of the segmentation overlapped with the bins are plotted. The preferred number of bins can be adjusted in this script.

```
# -*- coding: utf-8 -*-
"""
Created on Thu Aug 27 10:38:53 2020

@author: Rachel Oversier & Adriana Lentz
"""
import glob
import os
import csv
import numpy as np
import matplotlib.pyplot as plt
'--- Check working directory ---'
print("Previous Working Directory", os.getcwd())

'--- Change working directory to file location ---'
os.chdir("D:/Test/")
# change path if necessary
print("Changed Working Directory", os.getcwd())

basepath = "D:/Test/"
# change path if necessary
write_path = "D:/Test/Test1/ "
# for masks

'--- From working directory import binning-, segmentation- & area calculation- script ---'
# SimpleITK needs to be installed (use anaconda prompt)
# Opencv needs to be installed (use anaconda prompt)
from Thesis_Python_segmentation import HistologySection

niifiles = []
for file in glob.glob("*.nii"):
    niifiles.append(file)

segfiles = []
imgfiles = []

for file in niifiles:
    if 'img' in file:
        imgfiles.append(file)
    elif 'seg' in file:
        segfiles.append(file)

lenfiles = len(segfiles)

tissues = []
for tissue in HistologySection.Tissue:
    tissues.append(tissue)
```

```
BinsAantal = 8  # Adjust for number of bins that you want for the segmentation

for nr in range(lenfiles):
   img_he_path = basepath + imgfiles[nr]
   seg_he_path = basepath + segfiles[nr]
   filename = img_he_path[-27:-8]
   # function Histology section with new parameters overwrites input parameters of the other Python file that contains the
function Histologysection
   if img_he_path[-27:-8] == seg_he_path[-27:-8]:
      histo_he = HistologySection(img_he_path, seg_he_path, nbins=BinsAantal)

      temp_masks = HistologySection._construct_masks(histo_he, histo_he.segmentation)

      HistologySection.plot_binning(histo_he)
      histo_he.plot_segmentation()

      '--- Write mask arrays separately ---'
      keys = temp_masks.keys()
      for key in keys:
         np.savez(write_path+img_he_path[-27:-8]+'_'+str(key),temp_masks[key])

      '--- Normalization & write object to excel---'

      # Compute tissue area per bin and per tissue
      tissue_area_dict = {}
      for tissue in tissues:
         nbin_dict = {}
         for nbin in range(BinsAantal):
            nbin_dict[nbin]=HistologySection.compute_tissue_area(histo_he, tissue, binid=nbin)

         tissue_area_dict[tissue] = nbin_dict

      # Divide value per bin per tissue by plaque area per bin (except for plaque and lumen)
      Ptissue_area_dict = {}

      for tissue in tissues:
         nbin_dict = {}
         PLAQUE = tissues[1]
         LUMEN = tissues[2]
         for nbin in range(BinsAantal):
            if tissue != PLAQUE and tissue != LUMEN:
               value = tissue_area_dict[tissue][nbin]
               divide_amount = tissue_area_dict[PLAQUE][nbin]
               nbin_dict[nbin] = value/divide_amount
            elif tissue == LUMEN:
               nbin_dict[nbin]=tissue_area_dict[LUMEN][nbin]
            elif tissue == PLAQUE:
               nbin_dict[nbin]=tissue_area_dict[PLAQUE][nbin]
            else:
               print('ERROR tissue type')

         Ptissue_area_dict[tissue]=nbin_dict

      # Write to Excel
      # Change path if necessary
      Excel_path = 'D:/Test/Test2/'+ filename +'.csv' # For excel

      with open(Excel_path, 'w', encoding = "utf-8") as csvfile:
         fn = ['binid']
         for tissue in tissues:
            fn.append(tissue)

         csvwriter = csv.DictWriter(csvfile, fieldnames = fn, delimiter = ';', lineterminator='\n')
```

```
        csvwriter.writeheader()

        for binid in range(BinsAantal):
          csvwriter.writerow({fn[0]:binid, fn[1]: Ptissue_area_dict[tissues[0]][binid], fn[2]:
Ptissue_area_dict[tissues[1]][binid],
                              fn[3]: Ptissue_area_dict[tissues[2]][binid], fn[4]: Ptissue_area_dict[tissues[3]][binid],
                              fn[5]: Ptissue_area_dict[tissues[4]][binid], fn[6]: Ptissue_area_dict[tissues[5]][binid],
                              fn[7]: Ptissue_area_dict[tissues[6]][binid], fn[8]: Ptissue_area_dict[tissues[7]][binid],
                              fn[9]: Ptissue_area_dict[tissues[8]][binid], fn[10]: Ptissue_area_dict[tissues[9]][binid],
                              fn[11]: Ptissue_area_dict[tissues[10]][binid], fn[12]: Ptissue_area_dict[tissues[11]][binid],
                              fn[13]: Ptissue_area_dict[tissues[12]][binid], fn[14]: Ptissue_area_dict[tissues[13]][binid],
                              fn[15]: Ptissue_area_dict[tissues[14]][binid]})

      '--- Free-up memory ---'
      # can be uncommented if necessary
      del histo_he
      del temp_masks
      del tissue_area_dict
      del Ptissue_area_dict
      #plt.close('all')

    else:
      print("Img and seg name don't match")
```

---

2. MATLAB script

The following script was used to clip out the region of interest (a circular region around the luminal rupture site with a radius of 1300 pixels, equal to 2.40mm) of binary *plaque*, *cholesterol crystals*, *calcium* and *necrotic core* masks. The output of the script was a *square-shaped* mask that contained the circular region of interest.

In another script (not shown here) this mask was rotated, so that a vertical line could be drawn between the rupture and the centre of the lumen. The lumen was on the lower half of the circle. In order to achieve this rotation, the MATLAB function imrotate(*square*, *degrees*, 'crop') was used. The mask had the same size after the rotation, because of the last argument. The 23 masks of each tissue type were summed. Using the MATLAB function imagesc(*summedmask*), a rupture heat map was created for the parameters *plaque*, *cholesterol crystals*, *calcium* and *macrophages*.

```
mask = (readNPY('Path_to_NPY_mask')).'; % load m x n mask
filename = 'filename'; % filename of the histology section
mask = padarray(mask, [1000,1000]); % pad array with zeroes so that the mask gets "bigger", mask size is (m+2000) x
(n+2000)

%%
figure;
D = imshow(mask, [0 1]);
impixelinfo(D);
uint_mask=uint8(mask); % uint 16 to uint 8

x_y_rupture = ['x' 'y'];  % change for each section/rupture, indicates the x&y coordinate of the middle of the rupture
```

```
radius_rupture = 'constant' ;  % choose the same radius for every section, this radius will be used to define the region of interest
x_rupture = x_y_rupture(:,1);
y_rupture = x_y_rupture(:,2);

roi = images.roi.Circle(gca,'Center', x_y_rupture,'Radius',radius_rupture); % define region of interest
Round_ROI_mask = double(roi.createMask());  % creates a round mask from region of interest
uint_Round_mask= uint8(Round_ROI_mask); % double to uint 8
Round_mask_around_rupture = uint_Round_mask.* uint_mask; % now only the region of interest has the values of the original mask. Outside this circle, all pixels have value 0. Size is still (m+2000) x (n+2000).

figure, imshow(Round_mask_around_rupture, [0 1]);
impixelinfo;
hold on
plot(x_rupture, y_rupture, 'r+', 'Markersize', 20, 'Linewidth', 2)

%%
left_corner_x = x_rupture - radius_rupture; % location point for square
left_corner_y = y_rupture - radius_rupture; % location point for square
ROI_mask =
Round_mask_around_rupture([left_corner_y:left_corner_y+2*radius_rupture],[left_corner_x:left_corner_x+2*radius_rupture]);  % now, a square is cut out, exactly around the region of interest. The size of the ROI mask is (2*radius_rupture) x (2*radius_rupture)
figure
imshow(ROI_mask , [0 1]); %
hold on
plot(1301, 1301, 'r+', 'Markersize', 20, 'Linewidth', 2)

save(filename); % save the workspace
```

3.R scripts

3.1 8-BIN ANALYSIS

The following script was used for the statistical analysis of the digitized histological sections, when they were divided into eight bins. The script provides histograms, correlation matrices, the outcomes of (clustered) Wilcoxon rank sum tests, the principal component analysis, linearity checks and GEE logistic regression models with an exchangeability working correlation matrix and an independence working correlation matrix.

General

```
set.seed(3)
rm(list = ls())
```

Load packages

```
### Start###
library(readr)
library(tidyverse)
library(psych)
library(geepack)
library(corrplot)
library(reshape2)
library(ggcorrplot)
library(coin)
library(clusrank)
```

Read data

```
# load data
data <- read.csv(file="Path_to_header , header = TRUE, dec = ',', sep = ";")
temp = list.files(pattern="*.csv")
files <- list.files(path = "Path_to_8bin_files", pattern = "*_.csv", full.names = T)
files <- data.frame(files)

for (file in (files$files)){
 print(file)
 data_read <- read.csv(file=file, header = TRUE, dec = ',', sep = ";")
 data_read$filename <- file
 data <- rbind(data, data_read)
}
```

Clean data

```
# new column headers
colnames(data) <- c('Binid', 'Cholesterol.crystals', 'Plaque', 'Lumen', 'Calcium', 'Necrotic.core', 'Fresh.hemorrhage', 'Foam.cells', 'Artefact', 'Rupture', 'Neovessel', 'extra', 'Macrophages', 'Desorientated.SMC', 'Old.hemorrhage', 'Loose.Fibrous.Tissue.Macrophages', 'filename')

# check data is not loaded multiple times
data <- unique(data)

#Patient number as a variable (name of file is of the form "P..B.....", P is patient number, B contains the section number
data$filename <- str_split(data$filename, "/P", simplify = T)[,2]
data$filename <- str_split(data$filename, "B", simplify = T)[,1]
data$filename <- as.numeric(data$filename)

# characters to numeric
data <- as.data.frame(apply(data, 2, as.numeric))

# NaN's to 0 (indicating tissue is not present)
data[is.na(data)] <- 0

# Make rupture variable categorical and make it a factor and a numeric data type.
data[data$Rupture > 0,]$Rupture <- 1
data$Rupture <- as.factor(data$Rupture)
data$Rupture_num <- as.numeric(data$Rupture) -1

# correction for pixels and percentages (scaling factor induced by sitk module in Python is 14.28483007)
data$Cholesterol.crystals = data$Cholesterol.crystals*100
data$Plaque = data$Plaque * 14.28483007
data$Calcium = data$Calcium*100
data$Necrotic.core = data$Necrotic.core*100
data$Fresh.hemorrhage = data$Fresh.hemorrhage*100
data$Foam.cells = data$Foam.cells*100
data$Artefact = data$Artefact*100
data$Neovessel = data$Neovessel *100
data$extra = data$extra * 100
data$Macrophages = data$Macrophages *100
data$Desorientated.SMC = data$Desorientated.SMC *100
data$Old.hemorrhage = data$Old.hemorrhage*100
data$Loose.Fibrous.Tissue.Macrophages = data$Loose.Fibrous.Tissue.Macrophages*100
```

Explore data

Correlation plots

```
# histogram of variables, untrantransformed and log(x+1) transformed
multi.hist <- function(x) {nvar <- dim(x)[2]  #number of variables
    nsize=trunc(sqrt(nvar))+1   #size of graphic
    old.par <- par(no.readonly = TRUE) # all par settings which can be changed
    par(mfrow=c(nsize,nsize))      #set new graphic parameters
```

```
    for (i in 1:nvar) {
    name=names(x)[i]              #get the names for the variables
    hist(x[,i],main=name,xlab="percentage") } #draw the histograms for each variable
    on.exit(par(old.par))   #set the graphic parameters back to the original
    }
multi.hist(data[c(2,5:8,11,13,15)])
multi.hist(log(data[c(2,5:8,11,13,15)]+1))
multi.hist(data[c(3)])

# dataset for correlation
datacor<- (data[c(2:3,5:8,11,13,15)])
colnames(datacor) <- c('CC', 'PI', 'Ca', 'NC', 'FH', 'FC', 'Nv', 'Mp', 'OH')
datacor$Rupture_num = data$Rupture_num

# summary of the data
summary(datacor[datacor$Rupture_num == 0, [c(1:10)])
summary(datacor[datacor$Rupture_num == 1, [c(1:10)])
summary(datacor[c(1:10)])

# Function to get correlations and their p values
corr.data = function(x) {
 # Get correlations
 cor.vals = cor(datacor[c(1:9)], method = "kendall")
 # Get p-values
 cor.p = cor.mtest(datacor[c(1:9)], conf.level = 0.95, method = "kendall")$p
 rownames(cor.p) = rownames(cor.vals)
 colnames(cor.p) = colnames(cor.vals)
 cbind(rowvars=rownames(cor.vals), data.frame(cor.vals)) %>%
  gather(colvars, corr, -rowvars) %>%
  left_join(cbind(rowvars=rownames(cor.p), data.frame(cor.p)) %>%
            gather(colvars, p.value, -rowvars))
}

# Create plot
corr.data(datacor[c(1:9)]) %>%
 ggplot(aes(colvars, fct_rev(rowvars)))+
  geom_tile(colour="grey70", fill=NA) +
  geom_text(aes(label=sprintf("%1.2f", corr)), position=position_nudge(y=0.2),
                size=3, colour="grey20") +
  geom_text(aes(label=paste0("(",sprintf("%1.2f", p.value),")")),
  position=position_nudge(y=-0.2),
  colour="grey20", size=2.5) +
  labs(x="", y="") +
  theme_classic() +
  coord_fixed()

# nicer corrplot with ggcorrplot -> corrplot transformed data = corrplot untransformed data
corr<- cor(datacor[c(1:9)], method = "kendall")
p.mat <- cor_pmat(datacor[c(1:9)], method = "kendall", conf.level = 0.95)
ggcorrplot(
 corr,
 type = "lower",
 lab=TRUE,
 outline.color = "white",
 digits =2
)
corr<- cor((datacor[datacor$Rupture_num == 1, c(1:9)]), method = "kendall")
p.mat <- cor_pmat((datacor[datacor$Rupture_num == 1, c(1:9)]), method = "kendall", conf.level = 0.95)
ggcorrplot(
 corr,
 type = "lower",
 lab = TRUE,
 outline.color = "white",
```

```
 digits =2
)
```

Wilcoxon rank sum test

```
# the wilcoxon rank sum test, grouping by rupture level
levels(data$Rupture)
datacor$Rupture <- factor(data$Rupture, levels=c("1", "0"))
wilcox.test(data$Cholesterol.crystals  ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Plaque ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Calcium ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Necrotic.core~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Fresh.hemorrhage ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Foam.cells ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Neovessel ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Macrophages ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)
wilcox.test(data$Old.hemorrhage ~ data$Rupture,
        distribution = "exact", conf.int = TRUE, conf.level=0.95)

# Wilcoxon rank sum test for clustered data
clusWilcox.test(data$Cholesterol.crystals ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Plaque ~ Rupture_num  + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Calcium ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Necrotic.core~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Fresh.hemorrhage ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Foam.cells ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Neovessel ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Macrophages ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Old.hemorrhage ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
```

Analysis

PCA

```
# log(x+1) version of data
data_log <- data
data_log$Macrophages <- log(data$Macrophages + 1)
data_log$Foam.cells <- log(data$Foam.cells + 1)
data_log$Neovessel <- log(data$Neovessel + 1)
data_log$Calcium <- log(data$Calcium +1)
data_log$Cholesterol.crystals <- log(data$Cholesterol.crystals +1)
data_log$Necrotic.core <- log(data$Necrotic.core +1)

# PCA of all variables
pca_everything <- prcomp(data_log[, c("Neovessel", "Foam.cells", "Macrophages", "Plaque", "Cholesterol.crystals", "Calciu
```

```
m", "Necrotic.core")], scale = TRUE)
pca_everything
summary(pca_everything)

# biplot of the PCA
biplot(pca_everything, scale = 0 )
biplot(pca_everything, choices=c(1,3), scale = 0 )
biplot(pca_everything, choices=c(2,3), scale = 0 )

# plot of variance explained
plot(pca_everything, type ="l")
var_explained_df <- data.frame("Principal_Component"= paste0("PC",1:7),
Variance_explained=(pca_everything$sdev)^ 2/sum((pca_everything$sdev)^ 2))

var_explained_df %>%
 ggplot (aes(x=Principal_Component,y=Variance_explained))+
  geom_col ()+
  labs (title="Scree plot: PCA on scaled data")

# can PC1 and PC2 predict rupture?
data_pca = cbind(data_log, pca_everything$x[,1:7])
ggplot(data_pca, aes(PC1, PC2, col = Rupture, fill=Rupture)) +
 stat_ellipse(geom = "polygon", col = "black", alpha=0.5) +
 geom_point(shape=21, col="black"
```

GEE models

```
#gee model, first model with untransformed data

data$Plaque_GEE = data$Plaque/100000
gee_begin <- geeglm(Rupture_num ~ Neovessel+Calcium+ Necrotic.core + Macrophages+Cholesterol.crystals + Plaque_GEE
, data=data, family="binomial", corstr = "exchangeable", id = filename)
gee_begin
summary(gee_begin)
#assumption check
probabilities <- predict(gee_begin, type = "response")
theme_set(theme_classic())
# select numeric predictors
mydata <- datacor[c(1,2,3,4,7,8)] %>%
 dplyr::select_if(is.numeric)
predictors <- colnames(mydata)
# Bind the logit and tidying the data for plot, test linearity
mydata <- mydata %>%
 mutate(logit = log(probabilities/(1-probabilities))) %>%
 gather(key = "predictors", value = "predictor.value", -logit)
ggplot(mydata, aes(logit, predictor.value))+
 geom_point(size = 0.5, alpha = 0.5) +
 geom_smooth(method = "loess") +
 theme_bw() +
 facet_wrap(~predictors, scales = "free_y")

# GEE models of transformed data
data$Plaque_GEE = data$Plaque/100000
gee_trans_exch <- geeglm(Rupture_num ~ log(Neovessel+1)+log(Calcium+1)+ log(Necrotic.core+1) + log(Macrophages+1)+l
og(Cholesterol.crystals+1)+ Plaque_GEE, data=data, family="binomial", corstr = "exchangeable", id = filename)
gee_trans_exch
summary(gee_trans_exch)
exp(gee_trans_exch$coefficients)
gee_trans_ind <- geeglm(Rupture_num ~ log(Neovessel+1)+log(Calcium+1)+ log(Necrotic.core+1) + log(Macrophages+1)+lo
g(Cholesterol.crystals+1)+ Plaque_GEE, data=data, family="binomial", corstr = "independence", id = filename)
summary(gee_trans_ind)
exp(gee_trans_ind$coefficients)

# QIC values
```

```r
QIC(gee_begin)
QIC(gee_trans_exch)
QIC(gee_trans_ind)

# asssumption check, repeat for gee_trans_exch
probabilities <- predict(gee_trans_ind, type = "response")
theme_set(theme_classic())
# select numeric predictors
mydata <- data_log[c(2,3,5,6,11,13)] %>%
 dplyr::select_if(is.numeric)
predictors <- colnames(mydata)
# Bind the logit and tidying the data for plot, test linearity
mydata <- mydata %>%
 mutate(logit = log(probabilities/(1-probabilities))) %>%
 gather(key = "predictors", value = "predictor.value", -logit)
ggplot(mydata, aes(logit, predictor.value))+
 geom_point(size = 0.5, alpha = 0.5) +
 geom_smooth(method = "loess") +
 theme_bw() +
 facet_wrap(~predictors, scales = "free_y")
```

3.2 SENSITIVITY ANALYSIS

The script for the sensitivity analysis of the bin size can be found below. This script included the clustered Wilcoxon rank sum tests and the GEE logistic regression models with an exchangeable and an independence working correlation matrix. The histological sections were divided into 2-,4- and 16- bins and then analysed.

General

```r
set.seed(3)
rm(list = ls())
```

load packages

```r
### Start ###
library(readr)
library(tidyverse)
library(psych)
library(geepack)
library(corrplot)
library(clusrank)
```

Read data

```r
### load data
data <- read.csv(file="Path_to_header", header = TRUE, dec = ',', sep = ";")
temp = list.files(pattern="*.csv")
files <- list.files(path = "Path_to_2bin_4bin _16bin_files", pattern = "*.csv", full.names = T)
files <- data.frame(files)

for (file in (files$files)){
 print(file)
 data_read <- read.csv(file=file, header = TRUE, dec = ',', sep = ";")
 data_read$filename <- file
 data <- rbind(data, data_read)
}
```

Clean data

```
# new column headers
colnames(data) <- c('Binid', 'Cholesterol.crystals', 'Plaque', 'Lumen', 'Calcium', 'Necrotic.core', 'Fresh.hemorrhage', 'Foam.cells', 'Artefact', 'Rupture', 'Neovessel', 'extra', 'Macrophages', 'Desorientated.SMC', 'Old.hemorrhage', 'Loose.Fibrous.Tissue.Macrophages', 'filename')



# check data is not loaded multiple times
data <- unique(data)

# Patient number as extra variable
data$filename <- str_split(data$filename, "/P", simplify = T)[,2]
data$filename <- str_split(data$filename, "B", simplify = T)[,1]
data$filename <- as.numeric(data$filename)

# characters to numeric
data <- as.data.frame(apply(data, 2, as.numeric))

# NaN's to 0 (indicating tissue is not present)
data[is.na(data)] <- 0

# Make rupture variable categorical and make it a factor and a numeric data type.
data[data$Rupture > 0,]$Rupture <- 1
data$Rupture <- as.factor(data$Rupture)
data$Rupture_num <- as.numeric(data$Rupture) -1

# correction for pixels and percentages (scaling factor induced by sitk module in Python is 14.28483007)
data$Cholesterol.crystals = data$Cholesterol.crystals*100
data$Plaque = data$Plaque * 14.28483007
data$Calcium = data$Calcium*100
data$Necrotic.core = data$Necrotic.core*100
data$Fresh.hemorrhage = data$Fresh.hemorrhage*100
data$Foam.cells = data$Foam.cells*100
data$Artefact = data$Artefact*100
data$Neovessel = data$Neovessel *100
data$extra = data$extra * 100
data$Macrophages = data$Macrophages *100
data$Desorientated.SMC = data$Desorientated.SMC *100
data$Old.hemorrhage = data$Old.hemorrhage*100
data$Loose.Fibrous.Tissue.Macrophages = data$Loose.Fibrous.Tissue.Macrophages*100
data$Plaque_GEE = data$Plaque/100000
```

Clustered Wilcoxon/GEE for 2/4/16 bins

```
#  Clustered Wilcoxon
clusWilcox.test(data$Cholesterol.crystals ~ Rupture_num  + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Plaque ~ Rupture_num  + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Calcium ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Necrotic.core~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Fresh.hemorrhage ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Foam.cells ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Neovessel ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Macrophages ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
clusWilcox.test(data$Old.hemorrhage ~ data$Rupture_num + cluster(data$filename),
        data=data, method = "rgl")
```

```
## gee -> 2 bin
gee_trans_exch <- geeglm(Rupture_num ~ log(Neovessel+1)+log(Calcium+1)+ log(Necrotic.core+1) + log(Macrophages+1)+log(Cholesterol.crystals+1) + Plaque_GEE, data=data, family="binomial", corstr = "independence", id = filename)
gee_trans_exch
summary(gee_trans_exch)
QIC(gee_trans_exch)

gee_trans_ind <- geeglm(Rupture_num ~ log(Neovessel+1)+log(Calcium+1)+ log(Necrotic.core+1) + log(Macrophages+1)+log(Cholesterol.crystals+1) + Plaque_GEE, data=data, family="binomial", corstr = "exchangeable", id = filename)
summary(gee_trans_ind)
QIC(gee_trans_ind)
```

Bibliography

1. World Health Organisation. Cardiovascular diseases (CVDs). https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Published 2017.
2. Fairman RM. Management of asymptomatic carotid atherosclerotic disease. In: *Up To Date*. ; 2017. https://www-uptodate-com.ezproxy.leidenuniv.nl:2443/contents/management-of-asymptomatic-carotid-atherosclerotic-disease?search=carotid artery stenosis&source=search_result&selectedTitle=2~150&usage_type=default&display_rank=2.
3. Zhao X. Pathogenesis of atherosclerosis. In: *Up To Date*. ; 2019.
4. Akyildiz AC, Speelman L, Gijsen FJH. Mechanical properties of human atherosclerotic intima tissue. *J Biomech*. 2014;47(4):773-783. doi:10.1016/j.jbiomech.2014.01.019
5. Fairman RM. Management of symptomatic carotid atherosclerotic disease. https://www-uptodate-com.ezproxy.leidenuniv.nl:2443/contents/management-of-symptomatic-carotid-atherosclerotic-disease?search=management of carotid atherosclerotic disease&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1. Published 2017.
6. Boekhoven RW. *Mechanical Characterisation of Healthy and Diseased Carotid Arteries Using Ultrasound*.; 2015.
7. White CJ, Ramee SR, Collins TJ, et al. Carotid stents to prevent stroke: A nonsurgical option. *Ochsner J*. 2003;5(1):18-23.
8. O'Brien M, Chandra A. Carotid revascularization: Risks and benefits. *Vasc Health Risk Manag*. 2014;10:403-416. doi:10.2147/VHRM.S48923
9. Arroyo LH, Lee RT. Mechanisms of plaque rupture: Mechanical and biologic interactions. *Cardiovasc Res*. 1999;41(2):369-375. doi:10.1016/S0008-6363(98)00308-3
10. Cicha I, Wörner A, Urschel K, et al. Carotid plaque vulnerability: A positive feedback between hemodynamic and biochemical mechanisms. *Stroke*. 2011;42(12):3502-3510. doi:10.1161/STROKEAHA.111.627265
11. Barrett HE, Cunnane EM, Kavanagh EG, Walsh MT. On the effect of calcification volume and configuration on the mechanical behaviour of carotid plaque tissue. *J Mech Behav Biomed Mater*. 2016;56:45-56. doi:10.1016/j.jmbbm.2015.11.001
12. Akyildiz AC, Speelman L, van Brummelen H, et al. Effects of intima stiffness and plaque morphology on peak cap stress. *Biomed Eng Online*. 2011;10:1-13. doi:10.1186/1475-925X-10-25
13. Richardson PD. Biomechanics of plaque rupture: Progress, problems, and new frontiers. *Ann Biomed Eng*. 2002;30(4):524-536. doi:10.1114/1.1482781

14. Kwak BR, Bäck M, Bochaton-Piallat ML, et al. Biomechanical factors in atherosclerosis: Mechanisms and clinical implications. *Eur Heart J*. 2014;35(43):3013-3020. doi:10.1093/eurheartj/ehu353

15. Barrett HE, Mulvihill JJ, Cunnane EM, Walsh MT. Characterising human atherosclerotic carotid plaque tissue composition and morphology using combined spectroscopic and imaging modalities. *Biomed Eng Online*. 2015;14(Suppl 1):S5. doi:10.1186/1475-925X-14-S1-S5

16. Volokh KY. Comparison of biomechanical failure criteria for abdominal aortic aneurysm. *J Biomech*. 2010;43(10):2032-2034. doi:10.1016/j.jbiomech.2010.03.024

17. Holzapfel GA, Sommer G, Regitnig P. Anisotropic mechanical properties of tissue components in human atherosclerotic plaques. *J Biomech Eng*. 2004;126(5):657-665. doi:10.1115/1.1800557

18. Bentzon JF, Otsuka F, Virmani R, Falk E. Mechanisms of plaque formation and rupture. *Circ Res*. 2014;114(12):1852-1866. doi:10.1161/CIRCRESAHA.114.302721

19. Teng Z, Zhang Y, Huang Y, et al. Material properties of components in human carotid atherosclerotic plaques: A uniaxial extension study. *Acta Biomater*. 2014;10(12):5055-5063. doi:10.1016/j.actbio.2014.09.001

20. Finet G, Ohayon J, Rioufol G. Biomechanical interaction between cap thickness, lipid core composition and blood pressure in vulnerable coronary plaque: Impact on stability or instability. *Coron Artery Dis*. 2004;15(1):13-20. doi:10.1097/00019501-200402000-00003

21. Loree HM, Kamm RD, Stringfellow RG, Lee RT. Effects of fibrous cap thickness on peak circumferential stress in model atherosclerotic vessels. *Circ Res*. 1992;71(4):850-858. doi:10.1161/01.RES.71.4.850

22. Doradla P, Otsuka K, Nadkarni A, et al. Biomechanical Stress Profiling of Coronary Atherosclerosis: Identifying a Multifactorial Metric to Evaluate Plaque Rupture Risk. *JACC Cardiovasc Imaging*. 2020;13(3):804-816. doi:10.1016/j.jcmg.2019.01.033

23. Dolla WJS, House JA, Marso SP. Stratification of risk in thin cap fibroatheromas using peak plaque stress estimates from idealized finite element models. *Med Eng Phys*. 2012;34(9):1330-1338. doi:10.1016/j.medengphy.2011.12.024

24. Cilla M, Peña E, Martínez MA, Kelly DJ. Comparison of the vulnerability risk for positive versus negative atheroma plaque morphology. *J Biomech*. 2013;46(7):1248-1254. doi:10.1016/j.jbiomech.2013.02.012

25. Huang H, Virmani R, Younis H, Burke AP, Kamm RD, Lee RT. The impact of calcification on the biomechanical stability of atherosclerotic plaques. *Circulation*. 2001;103(8):1051-1056. doi:10.1161/01.CIR.103.8.1051

26. Ohayon J, Finet G, Gharib AM, et al. Necrotic core thickness and positive arterial remodeling index: Emergent biomechanical factors for evaluating the risk of plaque rupture. *Am J Physiol - Hear Circ Physiol*. 2008;295(2):717-727. doi:10.1152/ajpheart.00005.2008

27. Teng Z, Brown AJ, Calvert PA, et al. Coronary plaque structural stress is associated with plaque composition and subtype and higher in acute coronary syndrome: The BEACON i (Biomechanical Evaluation of Atheromatous Coronary Arteries) study. *Circ Cardiovasc Imaging*. 2014;7(3):461-470. doi:10.1161/CIRCIMAGING.113.001526

28. Wong KKL, Thavornpattanapong P, Cheung SCP, Sun Z, Tu J. Effect of calcification on the mechanical stability of plaque based on a three-dimensional carotid bifurcation model. *BMC Cardiovasc Disord*. 2012;12. doi:10.1186/1471-2261-12-7

29. Li ZY, Howarth S, Tang T, Graves M, U-King-Im J, Gillard JH. Does calcium deposition play a role in the stability of atheroma? Location may be the key. *Cerebrovasc Dis*. 2007;24(5):452-459. doi:10.1159/000108436

30. Mulvihill JJ, Cunnane EM, McHugh SM, Kavanagh EG, Walsh SR, Walsh MT. Mechanical, biological and structural characterization of in vitro ruptured human carotid plaque tissue. *Acta Biomater*. 2013;9(11):9027-9035. doi:10.1016/j.actbio.2013.07.012

31. Cunnane EM, Mulvihill JJE, Barrett HE, et al. Mechanical, biological and structural characterization of human atherosclerotic femoral plaque tissue. *Acta Biomater*. 2015;11(1):295-303. doi:10.1016/j.actbio.2014.09.024

32. Cunnane EM, Mulvihill JJE, Barrett HE, Hennessy MM, Kavanagh EG, Walsh MT. Mechanical properties and composition of carotid and femoral atherosclerotic plaques: A comparative study. *J Biomech*. 2016;49(15):3697-3704. doi:10.1016/j.jbiomech.2016.09.036

33. Lendon CL, Davies MJ, Born GVR, Richardson PD. Atherosclerotic plaque caps are locally weakened when macrophages density is increased. *Atherosclerosis*. 1991;87(1):87-90. doi:10.1016/0021-9150(91)90235-U

34. Itskovich V V., Samber DD, Mani V, et al. Quantification of human atherosclerotic plaques using spatially enhanced cluster analysis of multicontrast-weighted magnetic resonance images. *Magn Reson Med*. 2004;52(3):515-523. doi:10.1002/mrm.20154

35. Garcìa-Garcìa HM, Gogas BD, Serruys PW, Bruining N. IVUS-based imaging modalities for tissue characterization: Similarities and differences. *Int J Cardiovasc Imaging*. 2011;27(2):215-224. doi:10.1007/s10554-010-9789-7

36. Redgrave JN, Gallagher P, Lovett JK, Rothwell PM. Critical cap thickness and rupture in symptomatic carotid plaques: The oxford plaque study. *Stroke*. 2008;39(6):1722-1729. doi:10.1161/STROKEAHA.107.507988

37. Narula J, Nakano M, Virmani R, et al. Histopathologic Characteristics of Atherosclerotic Coronary Disease and Implications of the Findings for the Invasive and Noninvasive Detection of Vulnerable Plaques. *J Am Coll Cardiol*. 2013;61(10):1041-1051. doi:10.1016/j.jacc.2012.10.054.Histopathologic

38. Moreno PR, Purushothaman KR, Fuster V, O'Connor WN. Intimomedial interface damage and adventitial inflammation is increased beneath disrupted atherosclerosis in the aorta: Implications for plaque vulnerability. *Circulation*. 2002;105(21):2504-2511. doi:10.1161/01.CIR.0000017265.52501.37

39. Virmani R, Burke AP, Farb A, Kolodgie FD. Pathology of the Vulnerable Plaque. *J Am Coll Cardiol*. 2006;47(8 SUPPL.):0-5. doi:10.1016/j.jacc.2005.10.065

40. Davies MJ, Richardson PD, Woolf N, Katz DR, Mann J. Risk of thrombosis in human atherosclerotic plaques: Role of extracellular lipid, macrophage, and smooth muscle cell content. *Br Heart J*. 1993;69(5):377-381. doi:10.1136/hrt.69.5.377

41. Mauriello A, Servadei F, Sangiorgi G, et al. Asymptomatic carotid plaque rupture with unexpected thrombosis over a non-canonical vulnerable lesion. *Atherosclerosis*. 2011;218(2):356-362. doi:10.1016/j.atherosclerosis.2011.06.056

42. Carr SC, Farb A, Pearce WH, Virmani R, Yao JST. Activated inflammatory cells are associated with plaque rupture in carotid artery stenosis. *Surgery*. 1997;122(4):757-764. doi:10.1016/S0039-6060(97)90084-2

43. Tavora FR, Ripple M, Li L, Burke AP. Monocytes and neutrophils expressing myeloperoxidase occur in fibrous caps and thrombi in unstable coronary plaques. *BMC Cardiovasc Disord*. 2009;9:1-7. doi:10.1186/1471-2261-9-27

44. Moreno PR, Purushothaman KR, Fuster V, et al. Plaque neovascularization is increased in ruptured atherosclerotic lesions of human aorta: Implications for plaque vulnerability. *Circulation*. 2004;110(14):2032-2038. doi:10.1161/01.CIR.0000143233.87854.23

45. McCarthy MJ, Loftus IM, Thompson MM, et al. Angiogenesis and the atherosclerotic carotid plaque: An association between symptomatology and plaque morphology. *J Vasc Surg*. 1999;30(2):261-268. doi:10.1016/S0741-5214(99)70136-9

46. Abela GS, Aziz K. Cholesterol crystals rupture biological membranes and human plaques during acute cardiovascular events - A novel insight into plaque rupture by scanning electron microscopy. *Scanning*. 2006;28(1):1-10. doi:10.1002/sca.4950280101

47. Abela GS, Aziz K, Vedre A, Pathak DR, Talbott JD, DeJong J. Effect of Cholesterol Crystals on Plaques and Intima in Arteries of Patients With Acute Coronary and Cerebrovascular Syndromes. *Am J Cardiol*. 2009;103(7):959-968. doi:10.1016/j.amjcard.2008.12.019

48. Littlewood TD, Bennett MR. Apoptotic cell death in atherosclerosis. *Curr Opin Lipidol*. 2003;14(5):469-475. doi:10.1097/00041433-200310000-00007

49. Bennett RB, Sinha S, Owens GK. Vascular smooth muscle cells in atherosclerosis. *Circ Res*. 2017;118(4). doi:10.1161/CIRCRESAHA.115.306361.

50. Myasoedova VA, Chistiakov DA, Grechko A V., Orekhov AN. Matrix metalloproteinases in pro-atherosclerotic arterial remodeling. *J Mol Cell Cardiol*. 2018;123(August):159-167. doi:10.1016/j.yjmcc.2018.08.026

51. Sluimer, Kolodgie, Bijnens, et al. Thin-walled microvessels in human coronary atherosclerotic plaques show incomplete endothelial junctions: relevance of compromised structural integrity for intraplaque microvascular leakage. *J Am Coll Cardiol*. 2009;53(17). doi:10.1016/j.jacc.2008.12.056.

52. Kumamoto M, Nakashima Y, Sueishi K. Intimal neovascularization in human coronary atherosclerosis: Its origin and pathophysiological significance. *Hum Pathol*. 1995;26(4):450-456. doi:10.1016/0046-8177(95)90148-5

53. Kolodgie FD, Gold HK, Burke AP, et al. Intraplaque Hemorrhage and Progression of Coronary Atheroma. *N Engl J Med*. 2003;349(24):2316-2325. doi:10.1056/NEJMoa035655

54. Daemen MJ, Ferguson MS, Gijsen FJ, et al. Carotid plaque fissure: An underestimated source of intraplaque hemorrhage. *Atherosclerosis*. 2016;254:102-108. doi:10.1016/j.atherosclerosis.2016.09.069

55. Ross MH, Pawlina W. *Histology, A Text and Atlas*. Sixth. (Taylor C, ed.). Baltimore: Wolters Kluwer Health; 2011.

56. Bancroft JD, Layton C. *Bancroft's Theory and Practice of Histological Techniques, Chapter: 10 The Hematoxylins and Eosin*. Eighth Edi. Elsevier; 2019. doi:10.1016/B978-0-7020-6864-5.00010-4

57. Schick F. Tissue segmentation: a crucial tool for quantitative MRI and visualization of anatomical structures. *Magn Reson Mater Physics, Biol Med*. 2016;29(2):89-93. doi:10.1007/s10334-016-0549-0

58. Buscema M, Hieber SE, Schulz G, et al. Ex vivo evaluation of an atherosclerotic human coronary artery via histology and high-resolution hard X-ray tomography. *Sci Rep*. 2019;9(1):1-13. doi:10.1038/s41598-019-50711-1

59. Römer TJ, Brennan JF, Fitzmaurice M, et al. Histopathology of human coronary atherosclerosis by quantifying its chemical composition with Raman spectroscopy. *Circulation*. 1998;97(9):878-885. doi:10.1161/01.CIR.97.9.878

60. Thorp E, Li G, Seimon TA, Kuriakose G, Ron D, Tabas I. Reduced Apoptosis and Plaque Necrosis in Advanced Atherosclerotic Lesions of Apoe-/- and Ldlr-/- Mice Lacking CHOP. *Cell Metab*. 2009;9(5):474-481. doi:10.1016/j.cmet.2009.03.003

61. House WJ. Atlas EOSIN Y.

62. Manolakou P, Angelopoulou R, Bakoyiannis C, et al. Cellular proliferation in complicated versus uncomplicated atherosclerotic lesions: Total cell population, foam cells and newly formed microvessels. *Tissue Cell*. 2009;41(6):408-413. doi:10.1016/j.tice.2009.05.003

63. Wihastuti TA, Sargowo D, Heriansyah T, et al. The reduction of aorta histopathological images through inhibition of reactive oxygen species formation in hypercholesterolemia rattus norvegicus treated with polysaccharide peptide of Ganoderma lucidum. *Iran J Basic Med Sci*. 2015;18(5):514-519. doi:10.22038/ijbms.2015.4416

64. Teng Z, He J, Degnan AJ, et al. Critical mechanical conditions around neovessels in carotid atherosclerotic plaque may promote intraplaque hemorrhage. *Atherosclerosis*. 2012;223(2):321-326. doi:10.1016/j.atherosclerosis.2012.06.015

65. Milei, J ; Parodi, J C ; Fernandez Alonso, G ; Barone, A ; Beigelman, R ; Ferreira, L M ;

Arrigoni, G ; Matturri L. Carotid atherosclerosis. Immunocytochemical analysis of the vascular and cellular composition in endarterectomies. *Cardiologia*. 1996;41(6):535-542.

66. Dantas ML, de Oliveira JMGC, Carvalho L, et al. Comparative analysis of the tissue inflammatory response in human cutaneous and disseminated leishmaniasis. *Mem Inst Oswaldo Cruz*. 2014;109(2):202-209. doi:10.1590/0074-0276130312

67. Manduteanu I, Simionescu M. Inflammation in atherosclerosis: A cause or a result of vascular disorders? *J Cell Mol Med*. 2012;16(9):1978-1990. doi:10.1111/j.1582-4934.2012.01552.x

68. Chu B, Kampschulte A, Ferguson MS, et al. Hemorrhage in the atherosclerotic carotid plaque: A high-resolution MRI study. *Stroke*. 2004;35(5):1079-1084. doi:10.1161/01.STR.0000125856.25309.86

69. Steinmetz N, Appelhoff S, Rossant C. npy-matlab. GitHub. https://github.com/kwikteam/npy-matlab. Published 2018.

70. The Pennsylvania State University. 18.3 - Kendall Tau-b Correlation Coefficient. https://online.stat.psu.edu/stat509/node/158/. Published 2018.

71. Llukan Puka. *International Encyclopedia of Statistical Science*. (Lovric M, ed.). Berlin Heidelberg: Springer-Verlag Berlin; 2011. doi:DOI 10.1007/978-3-642-04898-2

72. RDocumentation. Correlation, Variance And Covariance (Matrices). https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor. Accessed October 12, 2020.

73. Petrie A, Sabin C. *Medical Statistics at a Glance*. Third edit. Wiley-Blackwell; 2011.

74. Tattar BN, Ramaiah S, Manjunath BG. *A Course in Statistics with R*. John Wiley and Sons Ltd; 2016.

75. Rosner B, Glynn RJ, Lee MT. Extension of the Rank Sum Test for Clustered Data : Two-Group Comparisons with Group Membership Defined at the Subunit Level. 2006;(December):1251-1259. doi:10.1111/j.1541-0420.2006.00582.x

76. Jiang Y, Lee MT, Rosner B. Wilcoxon Rank-Based Tests for Clustered Data with R Package clusrank.

77. Yujing J, Ting Lee M-L, Yan J. *Package ' Clusrank ': Wilcoxon Rank Sum Test for Clustered Data.*; 2018.

78. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22. doi:10.1093/biomet/73.1.13

79. Wilson J, Lorenz K. Generalized Estimating Equations Logistic Regression. 2015:281. doi:10.1007/978-3-319-23805-0

80. Holzapfel GA, Sommer G, Gasser CT, Regitnig P. Determination of layer-specific mechanical properties of human coronary arteries with nonatherosclerotic intimal thickening and related constitutive modeling. *Am J Physiol - Hear Circ Physiol*. 2005;289(5 58-5):H2048-H2058. doi:10.1152/ajpheart.00934.2004

81. Xu C, Li Z, Xue Y, Zhang L, Wang M. An R package for model fitting, model selection and the simulation for longitudinal data with dropout missingness. *Commun Stat Simul Comput*. 2019;48(9):2812-2829. doi:10.1080/03610918.2018.1468457

82. Pan W. Akaike's Informat ion Criterion in Generalized Estimating Equations. *Biometrics*. 2001;(March):120-125.

83. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ*. 2015;351:7-11. doi:10.1136/bmj.h3868

84. Boyle JJ. Association of coronary plaque rupture and atherosclerotic inflammation. *J Pathol*. 1997;181(1):93-99. doi:10.1002/(SICI)1096-9896(199701)181:1<93::AID-PATH696>3.0.CO;2-H

85. Szumilas M. Statistical question: Odds ratios. *J Can Acad Child Adolesc Psychiatry*. 2010;19(3):227-229. doi:10.1136/bmj.c4414

86. Zhang D. EPID 766 : Analysis of Longitudinal Data from Epidemiologic Studies. *EPID 766 Anal Longitud Data from Epidemiol Stud Grad Summer Sess Epidemiol NC State Univ Dep Stat*. 2016:205. http://www4.stat.ncsu.edu/~dzhang2/epid766/766note.pdf.

87. Højsgaard S, Halekoh U. On the usage of the geepack Citing geepack Simulating a dataset To illustrate the usage of the waves argument and the zcor argument together with. 2019;(2006):1-6.

88. Boyle J. Macrophage Activation in Atherosclerosis: Pathogenesis and Pharmacology of Plaque Rupture. *Curr Vasc Pharmacol*. 2005;3(1):63-68. doi:10.2174/1570161052773861

89. Akoglu H. User's guide to correlation coefficients. *Turkish J Emerg Med*. 2018;18(3):91-93. doi:10.1016/j.tjem.2018.08.001

90. Teng Z, Sadat U, Brown AJ, Gillard JH. Plaque hemorrhage in carotid artery disease: Pathogenesis, clinical and biomechanical considerations. *J Biomech*. 2014;47(4):847-858. doi:10.1016/j.jbiomech.2014.01.013

91. Milei J, Parodi JC, Ferreira M, Barrone A, Grana DR, Matturri L. Atherosclerotic plaque rupture and intraplaque hemorrhage do not correlate with symptoms in carotid artery stenosis. *J Vasc Surg*. 2003;38(6):1241-1247. doi:10.1016/S0741-5214(03)00910-8

92. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*. 2010;21(4):467-474. doi:10.1097/EDE.0b013e3181caeb90

93. Laird NM. Longitudinal Data Analysis for Counts and Binary Outcomes: Generalized Estimating Equations (GEE). 1989;(1988):96-109.

94. Mulvihill JJ, Cunnane EM, McHugh SM, Kavanagh EG, Walsh SR, Walsh MT. Mechanical, biological and structural characterization of in vitro ruptured human carotid plaque tissue. *Acta Biomater*. 2013;9(11):9027-9035. doi:10.1016/j.actbio.2013.07.012

95. Grønholdt ML, Dalager-Pedersen S, Falk E. Coronary atherosclerosis: determinants of

plaque rupture. *Eur Heart J*. 1998;19 Suppl C:C24—9. http://europepmc.org/abstract/MED/9597422.

96. Mccann MT. Tools for Automated Histology Image Analysis. 2015.

97. Tarkin JM, Dweck MR, Evans NR, et al. Imaging Atherosclerosis. *Circ Res*. 2016;118(4):750-769. doi:10.1161/CIRCRESAHA.115.306247

98. Yabushita H, Bouma BE, Houser SL, et al. Characterization of human atherosclerosis by optical coherence tomography. *Circulation*. 2002;106(13):1640-1645. doi:10.1161/01.CIR.0000029927.92825.F6

99. Nishimura S, Ehara S, Hasegawa T, Matsumoto K, Yoshikawa J, Shimada K. Cholesterol crystal as a new feature of coronary vulnerable plaques: An optical coherence tomography study. *J Cardiol*. 2017;69(1):253-259. doi:10.1016/j.jjcc.2016.04.003

100. Crochan J O'Sullivan, David Tüller RZ and FRE. Intravascular Photoacoustic Imaging of Vulnerable Atherosclerotic Plaque. *ICR; Interv Cardiol Rev*. 2016:27-32. doi:10.15420/icr.2016

101. Daghem M, Bing R, Fayad ZA, Dweck MR. Noninvasive Imaging to Assess Atherosclerotic Plaque Composition and Disease Activity: Coronary and Carotid Applications. *JACC Cardiovasc Imaging*. 2020;13(4):1055-1068. doi:10.1016/j.jcmg.2019.03.033

102. Den Hartog AG, Bovens SM, Koning W, et al. Current status of clinical magnetic resonance imaging for plaque characterisation in patients with carotid artery stenosis. *Eur J Vasc Endovasc Surg*. 2013;45(1):7-21. doi:10.1016/j.ejvs.2012.10.022

103. Naim C, Douziech M, Therasse É, et al. Vulnerable atherosclerotic carotid plaque evaluation by ultrasound, computed tomography angiography, and magnetic resonance imaging: An overview. *Can Assoc Radiol J*. 2014;65(3):275-286. doi:10.1016/j.carj.2013.05.003

104. Johri AM, Nambi V, Naqvi TZ, et al. Recommendations for the Assessment of Carotid Arterial Plaque by Ultrasound for the Characterization of Atherosclerosis and Evaluation of Cardiovascular Risk: From the American Society of Echocardiography. *J Am Soc Echocardiogr*. 2020;33(8):917-933. doi:10.1016/j.echo.2020.04.021

105. Rothwell PM, Eliasziw M, Gutnikov SA, et al. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet*. 2003;361(9352):107-116. doi:10.1016/S0140-6736(03)12228-3

106. Howard DPJ, Van Lammeren GW, Rothwell PM, et al. Symptomatic carotid atherosclerotic disease: Correlations between plaque composition and ipsilateral stroke risk. *Stroke*. 2015;46(1):182-189. doi:10.1161/STROKEAHA.114.007221

107. Hellings WE, Moll FL, Vries J-PPM De, et al. Atherosclerotic Plaque Composition and Occurrence of Restenosis After Carotid Endarterectomy. *JAMA*. 2008;299(5):547-554.

108. Moreno PR, Purushothaman KR, Fuster V, et al. Plaque neovascularization is increased in ruptured atherosclerotic lesions of human aorta: Implications for plaque

vulnerability. *Circulation*. 2004;110(14):2032-2038. doi:10.1161/01.CIR.0000143233.87854.23

109. Mauriello A, Sangiorgi GM, Virmani R, et al. A pathobiologic link between risk factors profile and morphological markers of carotid instability. *Atherosclerosis*. 2010;208(2):572-580. doi:10.1016/j.atherosclerosis.2009.07.048

110. Konishi T, Funayama N, Yamamoto T, et al. Pathological quantification of carotid artery plaque instability in patients undergoing carotid endarterectomy. *Circ J*. 2018;82(1):258-266. doi:10.1253/circj.CJ-17-0204

111. Michel JB, Virmani R, Arbustini E, Pasterkamp G. Intraplaque haemorrhages as the trigger of plaque vulnerability. *Eur Heart J*. 2011;32(16):1977-1985. doi:10.1093/eurheartj/ehr054

112. Rogers P, Stoner J. Modification of the Sandwich Estimator in Generalized Estimating Equations with Correlated Binary Outcomes in Rare Event and Small Sample Settings. *Am J Appl Math Stat 2015*. 2015;3(6):243-251.

113. Hansa G, Bhargava K, Bansal M, Tandon S, Kasliwal RR. Carotid intima-media thickness and coronary artery disease: An Indian perspective. *Asian Cardiovasc Thorac Ann*. 2003;11(3):217-221. doi:10.1177/021849230301100308

114. Chapter 22 Cardiovascular System: Vessels and Circulation. http://droualb.faculty.mjc.edu/Lecture Notes/Unit 4/cardiovascular-blood vessels with figures.htm.

115. Robertson AM, Watton PN. *Mechanobiology of the Arterial Wall*. Elsevier Inc.; 2013. doi:10.1016/B978-0-12-415824-5.00008-4

116. Moreno PR, Purushothaman KR, Sirol M, Levy AP, Fuster V. Neovascularization in human atherosclerosis. *Circulation*. 2006;113(18):2245-2252. doi:10.1161/CIRCULATIONAHA.105.578955

117. Zhao X. Pathogenesis of atherosclerosis. In: *Up To Date*. ; 2019. https://www-uptodate-com.ezproxy.leidenuniv.nl:2443/contents/pathogenesis-of-atherosclerosis?search=atherosclerosis&sectionRank=1&usage_type=default&anchor=H27901102&source=machineLearning&selectedTitle=1~150&display_rank=1#H27901102.

118. Bergheanu SC, Bodde MC, Jukema JW. Pathophysiology and treatment of atherosclerosis: Current view and future perspective on lipoprotein modification treatment. *Netherlands Hear J*. 2017;25(4):231-242. doi:10.1007/s12471-017-0959-2

119. Milutinović A, Šuput D, Zorc-Pleskovič R. Pathogenesis of atherosclerosis in the tunica intima, media, and adventitia of coronary arteries: An updated review. *Bosn J Basic Med Sci*. 2020;20(1):21-30. doi:10.17305/bjbms.2019.4320

120. Doran AC, Meller N, McNamara CA. Role of smooth muscle cells in the initiation and early progression of atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2008;28(5):812-819. doi:10.1161/ATVBAHA.107.159327

121. Loughrey CM, Young IS. *Clinical Biochemistry of the Cardiovascular System*. Third Edit.

Elsevier Ltd.; 2014. doi:10.1016/B978-0-7020-5140-1.00038-9

122. Otsuka F, Sakakura K, Yahagi K, Joner M, Virmani R. Has our understanding of calcification in human coronary atherosclerosis progressed? *Arterioscler Thromb Vasc Biol*. 2014;34(4):724-736. doi:10.1161/ATVBAHA.113.302642

123. Wild C. *The Wilcoxon Rank-Sum Test*.; 1997.

124. Neuhäuser M. *International Encyclopedia OfStatistical Science*. (Lovric M, ed.). Berlin Heidelberg: Springer-Verlag; 2011. doi:DOI 10.1007/978-3-642-04898-2

125. Chicken M, Douglas H, Eric AW. *Nonparametric Statistical Methods*. Third. (Balding DJ, Cressie NA., Fitzmaurice CM, et al., eds.). Hoboken, New Jersey.: John Wiley & Sons, Inc; 2015.

126. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscip Rev Comput Stat*. 2010;2(4):433-459. doi:10.1002/wics.101

127. Bro R, Smilde AK. Principal component analysis. *Anal Methods*. 2014;6(9):2812-2831. doi:10.1039/c3ay41907j

128. Shlens J. A Tutorial on Principal Component Analysis. 2014. http://arxiv.org/abs/1404.1100.

129. RDocumentation. Principle Component Analysis. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html.

130. Mathworks. Singular value decomposition. https://www.mathworks.com/help/matlab/ref/double.svd.html;jsessionid=314ca0129f261777a44fbb697153#bu2_0hq-S. Published 2020. Accessed October 1, 2020.

131. Wall ME, Rechtsteiner A, Rocha LM. Singular Value Decomposition and Principal Component Analysis. *A Pract Approach to Microarray Data Anal*. 2005:91-109. doi:10.1007/0-306-47815-3_5

132. The Pennsylvania State University. 6.1 - Introduction to Generalized Linear Models. https://online.stat.psu.edu/stat504/node/216/. Published 2018.

133. Kabacoff RI. Generalized Linear Models. https://www.statmethods.net/advstats/glm.html. Published 2017. Accessed September 28, 2020.

134. MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. *Fundamental Principals of Statistical Inference*.; 2018. doi:10.1016/b978-0-12-407197-1.00004-1

135. Schrader R. Logistic Regression – Interpreting Parameters. *Logist Regres – Interpret Parameters*.:112-123. http://www.unm.edu/~schrader/biostat/bio2/Spr06/lec11.pdf.

136. Zeger SL, Liang KY. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*. 1986;42(1).

137. The GENMOD Procedure. SAS Institute Inc. https://www.sfu.ca/sasdoc/sashtml/stat/chap29/sect38.htm#:~:text=Working Correlation Matrix&text=is the true correlation matrix,true covariance matrix of Y.&text=If you specify the working,to the independence estimating equation.

Published 2019. Accessed September 28, 2020.

138. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. *J Stat Softw*. 2006;15(2):1-11. doi:10.18637/jss.v015.i02

139. The Pennsylvania State University. 12.1 - Introduction to Generalized Estimating Equations. https://online.stat.psu.edu/stat504/node/180/. Published 2018. Accessed September 28, 2020.

140. Pekár S, Brabec M. Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology*. 2018;124(2):86-93. doi:10.1111/eth.12713

141. Cui J. QIC program and model selection in GEE analyses. *Stata J*. 2007;7(2):209-220. doi:10.1177/1536867x0700700205

142. Cui J, Qian G. Selection of working correlation structure and best model in GEE analyses of longitudinal data. *Commun Stat Simul Comput*. 2007;36(5):987-996. doi:10.1080/03610910701539617

143. Ghisletta P, Spini D. An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *J Educ Behav Stat*. 2004;29(4):421-437. doi:10.3102/10769986029004421

144. Kirchner JW. Data Analysis Toolkit # 3 : Tools for Transforming Data. *Analysis*. 2001;(x):1-2.