

Exploring a Prediction Model Based on Patient Record Data in Carotid Stenosis Risk Assessment

Kjersti Hervik^{1,2}, Tom Wilsgaard³, Truls Myrmed^{1,2}, Knut Eivind Kjørstad¹

¹Department of Heart, Lung and Vascular Surgery, University Hospital of North Norway, Tromsø, Norway; ²The Faculty of Health Sciences, Uit The Arctic University of Norway, Tromsø, Norway; ³Department of Social Sciences, Uit The Arctic University of Norway, Tromsø, Norway

Correspondence: Kjersti Hervik, Department of Heart, Lung and Vascular surgery, University Hospital of North Norway, UNN Tromsø, Hjerter- lunge- og karkirurgisk avdeling, Postboks 100, Tromsø, Norway, 9038, Tel +4741633936, Email Kjersti.Hervik2@unn.no

Introduction: Digitalised patient records represent a large potential source of real-world data. Nevertheless, confidentiality and data protection has made big data extraction from patient journals impossible in the past. Future options for free text reading by artificial intelligence might enable data extraction while maintaining confidentiality. In turn this could enable improvement in risk predictions. However, it remains unclear whether free-text data constitute an appropriate data source for this purpose, or whether this data source can be incorporated into a clinically applicable risk prediction tool using an appropriate statistical model. We aim to investigate this for stroke risk in patients with carotid stenosis.

Methods: A test dataset of patients with carotid stenosis, based on free text data, was manually established. Parameters of choice were extracted from digital patient records to assess their performance in the prediction of individual risk profile. The dataset is applied to evaluate a proposed statistical model including time dependency of specific variables and multiple endpoint analysis, comparing this to the results from a traditional Cox regression analysis.

Results: 392 symptomatic and 471 asymptomatic patients with carotid stenosis were included in the analysis, where respectively 49 repeat events and 56 first stroke events attributed to the carotid stenosis occurred. The suggested extended regression model shows a statistically significant risk of a high degree carotid stenosis lesion on the risk of an attributable stroke, which was not demonstrated by use of a traditional Cox regression analysis. For a stenosis of 70–99%, asymptomatic patients with a right sided carotid stenosis had a hazard ratio (HR) 4.98 for a right sided event, and for left sided carotid stenosis a HR 5.23 for a left sided event. Symptomatic patients were found to have a HR of 2.25 for a recurrent event, as many received protective treatment. This is suggested to imply improved risk prediction by an advanced statistical model.

Discussion: In this article a description of the data set and the methods behind our model choice has been the main emphasis, in order to evaluate its suitability for this purpose. Analysis of model performance by concordance analysis and Brier score has been published in a separate article.

Keywords: carotid stenosis, stroke, carotid endarterectomy, patient journal data, time dependent variables, multiple endpoint analysis

Introduction

Atherosclerotic diseases represent a large group of conditions, where demographic changes impact both development and progression. In the same disease group, diagnostics and treatment options are developing quickly, and keeping track of all aspects can be challenging. Improved prevention strategies could slow down progression of the chronic disease, and prevent or postpone life-threatening acute events such as heart attack, stroke or critical limb ischemia. Still, evidence-based tailoring of therapy is necessary for safe and cost-effective prophylaxis, and minimised impact of adverse treatment effects.

Surgical or endovascular treatment of carotid stenosis in stroke prevention is well known for its inevitable risk of periprocedural strokes. Although the stroke risk in association to treatment is reported to be slightly lower today than 20 years ago,¹ the risk of a procedure-related stroke is still considerable for both endovascular procedures and open endarterectomy.



Risk prediction tools could provide an aid in determining who are likely to benefit the most, and thereby tailor treatment recommendations for each individual patient. Several risk calculators have been developed,^{2–4} but none has yet succeeded in significantly challenging the established guidelines for carotid disease.⁵ In the past, such risk prediction tools have generally been based on the same data material as current guidelines, and the additional value in decision making seems sparse. Risk prediction tools based on randomized controlled trials may be representative at the time of analysis, but necessitate repeated studies to account for demographic changes and improvements in treatment options.

Artificial intelligence (AI) is technology with a broad range of applications in medicine.⁶ The virtual AI branch includes deep learning, which enables data extraction from free text. This could enable real-time risk prediction by allowing continuous updates with current data, thereby facilitating continuous adjustment to evolving demographic characteristics. Risk prediction models based on free text patient record data still lack verification for this purpose. In this study we aim to evaluate the performance of free text patient record data in risk prediction.

We present a new method for evaluating the risk of stroke in patients with carotid stenosis, based on manually extracted patient record data. An extended regression model is applied, to include repeat event analysis and time dependency of key variables. The intention is to explore this data source for risk prediction by applying a suitable statistical model. In this article we will describe the background for the model choice, and the method applied in its development. To our knowledge, the suggested extended regression model has not previously been applied on carotid stenosis data.

The study complies with the Declaration of Helsinki, and has been assessed and approved by the Regional Committee for Medical Research Ethics Northern Norway (REC).

Methods

Data Material

This analysis is performed in order to evaluate the model potential, where a retrospective approach to risk prediction is carried out. The test data set was manually established by the use of “all comer” patient record data from the University Hospital of North Norway. 1428 patients registered with the ICD-10 diagnosis “I65.2 Occlusion and stenosis of carotid artery” since the introduction of an electronic patient record system in 2003, were invited. Both symptomatic and asymptomatic patients were included. Most of the asymptomatic patients had been referred to the out-patient clinic from the Tromsø Study. This is a large population study that has been taking place since 1974, with seven separate studies performed to date. The study consists of questionnaires for all participants, and various examination studies performed on a smaller selection. In the studies Tromsø 4 (1994–95), Tromsø 5 (2001) and Tromsø 6 (2007–2008), a selection of participants were examined by ultrasound for carotid stenosis by an ultrasound technician. Those who had carotid atherosclerosis, were referred to the neurological out-patient clinic for a more detailed ultrasound evaluation. Accordingly, information in relation to this consultation has been documented in the hospital patient record, including ultrasound findings, ongoing medical therapy, smoking history, relevant past medical history, and usually measurement of vital signs. Advice was given in relation to treatment and follow up, and most of these asymptomatic patients had standard medical therapy prescribed. For most patients with a stenosis of more than 50%, follow-up examination to evaluate disease progression was performed.

Inclusion by Consent

All participants referred from the Tromsø Study had previously given a broad consent to the use of Tromsø Study data. The Regional Ethical Committee (REC) advised inclusion of these patients without new consent. For all deceased individuals, REC also advised exemption from consent to access patient record data. Written consent was obtained to permit access to patient record data for all other participants.

Parameters of Consideration

The results from previous carotid studies were used to target patient properties associated with disease progression and/or ischemic events.^{2,7–13} We registered previous medical or surgical treatment at the start of follow up, and separately

registered therapy initiated at the time of diagnosis. Antihypertensive agents, antiplatelets and other anticoagulation therapy, and lipid lowering drugs were registered by type but not dose. If carotid endarterectomy (CEA) or stenting (CAS) was performed at any time during follow-up, preceding events, timing of treatment, method and complications, including postoperative strokes, were registered.

Cerebral Ischemic Events

The primary endpoint was defined as ischemic strokes. All new ischemic cerebral events were registered, until death or censoring at the end of follow up. Transient ischemic attacks (TIAs) were defined as ischemic events with full recovery within 24 hours, where no haemorrhagic or ischemic lesion was detected on CT or MRI. TIAs were treated in the analysis as risk predictors rather than endpoints, as no permanent injury to brain tissue will follow, and because these events are known to frequently precede a stroke. TIAs were the presentation of symptomatic carotid disease in 156 patients, and by that triggered the indication for surgical treatment in 65 of these.

The remaining symptomatic patients had already suffered an ischemic stroke at the time of presentation. For these patients, stroke severity was registered by use of the Modified Rankin Scale (mRS).

Endpoint Classification

Less than 20% of all patients with ischemic strokes are found to have carotid stenosis, and the condition is seen as the cause of the event in only about half of these.¹⁴ As the stroke risk is shown to depend on the degree of stenosis,⁸ it is assumed to increase in the presence of traditional risk factors for atherosclerotic progression if the observation time is long enough. A previous study with a median of 13 years of follow-up still fails to prove this.¹² Stroke by other causes, such as cardiac thromboembolism, is perceived as noise in regards to carotid stenosis risk. Thus, endpoint subclassification is desirable, to pinpoint the risk of strokes caused by the carotid stenosis. A clear cause of cerebral ischemia can be difficult to define, and we aimed for a systematic approach where the subclassification was reproducible and standardised. We have chosen the ASCOD score¹⁵ for this purpose.

Ischemic strokes were registered when a new lesion was detected by CT or MRI, or when symptoms persisted for more than 24 hours and haemorrhage was ruled out. We applied the ASCOD phenotyping to subtype the events.¹⁵ The ASCOD phenotype A1-1, when S, C, O and D are 0 or 3, strongly indicates the carotid stenosis as causative of the event. In the case of A1-2 or -4, or A2-1, where SCOD grade is 0, the causation is also likely. The phenotype A0 makes other causes of the ischemic event more likely. This also applies to A2 and A3 if S, C, O or D grade is 1. When no ASCOD 1 grade is defined, or two or more grade 1 coexist within the phenotypes, the cause is uncertain. In these cases, the most likely cause is evaluated individually for each patient.

The potential source of error due to a degree of objective endpoint evaluation, was explored by performing separate analyses for the endpoints “carotid stroke”, “any ischemic stroke”, and “any ischemic stroke or death”.

Recurrent Events in Survival Analysis

In evaluation of the effect of prophylactic therapy, a long observation time is desirable, as ideally, we would like to reveal the lifetime risk. During a decade or more of follow up, some individuals will experience more than one stroke. These patients could potentially have properties making them particularly vulnerable to new events. Therefore, all new events were registered and included in the model. Multiple endpoint analysis is rarely performed, and existing datasets from previous carotid stenosis studies generally only include time to the first event or exclusion.

Recurrent event models in survival analysis offer statistical methods for evaluating the risk of an event that can occur more than once. Different models exist, and the choice of model depends on whether or not the events are seen as different processes, and on the dependency of risk on the time from entry into the study to the first and recurrent events. In the analysis of stroke, each event is seen as the same process. The time to the first event is assumed to impact the time to a potential recurrent event. The Prentice, Williams, and Peterson Total Time Model (PWP-TT)¹⁶ accounts for this order of events, and is our model of choice.

Time Dependent Variables

Some key variables are known to have a strong impact on the risk of new events, namely the degree of carotid stenosis and new TIAs. We have chosen to include these variables with time dependency to explore the impact of disease progression over time on stroke risk. We do not know of previous datasets with time dependent registrations of carotid stenosis progression and all new TIAs. Treatment of the carotid stenosis is given time dependency by means of a reduction of the stenosis to zero on the treated side, although in some cases this is an assumption, as not all patients received a postoperative control ultrasound.

By the use of an extended Cox model, the effect of a change in a variable after a given time interval can be incorporated in the risk prediction model. It is possible to include both time dependent and time independent variables in the same analysis. The time-independent variables will not change value during the observation time, and is obviously applied for gender. Several of our registered variables in relation to general cardiovascular risk, such as smoking, blood pressure and antihypertensive treatment, could be seen as time dependent. However, for practical reasons, only the predictors expected to have a strong impact on stroke risk were included with time dependency in the test dataset.

The Complexity of Sidedness

Ischemic events can occur in two cerebral hemispheres, each supplied by a separate carotid artery. Intracerebral communications through the Circle of Willis serve as security across the hemispheres, to maintain perfusion if a precerebral artery is compromised. The risk of an embolic stroke affecting one side depends on the degree of carotid stenosis on the ipsilateral side.^{8,17} Still, stroke by hypoperfusion can occur in both hemispheres if the total perfusion to watershed areas of the brain is insufficient due to severe stenosis or occlusion of the precerebral arteries. Each side is therefore considered to carry a separate stroke risk, where the atherosclerotic status of both carotid arteries may contribute to the risk on each side.

If an ischemic event occurs, surgical or endovascular treatment is usually performed on the ipsilateral side. If there is ipsilateral carotid occlusion, treatment might be performed on the contralateral side to prevent hypoperfusion, although there is controversy about this indication for therapy.¹⁸ In our data material, the time and side of treatment is registered, and the degree of stenosis on this side drops to zero after successful recanalization.

As previously described, TIAs are included in the model as time dependent variables. The TIA variable is also subdivided into right and left sides, or unknown if the symptoms are not side specific. These variables are allowed to “count” if several TIAs occur, which means that the excess risk assumed to be associated with having more than one TIA on the same side can be estimated.¹⁹

Software for Data Analysis

Redcap was used for data registration, and data was then exported to SPSS. SPSS was used for simple data analysis and some graphics. Data preparation for a matrix set-up with time intervals to new events and/or changes in time dependent variables was also performed in SPSS, and the matrix dataset was then exported to the SAS software 9.4, SAS Institute Inc., Cary, NC, USA for analysis with time dependent variables and time dependency of endpoints.

Results

1428 patients were registered with the ICD10 diagnosis I65.2 between 2003 and 2019. Of these, 494 were dead at the time of inclusion, and these were included without consent as advocated by REC. Consent was received from 440 patients, either through the Tromsø Study, or by returning a signed consent form. For the individuals where permission to access patient records was granted, the ICD10 diagnosis was individually verified through patient record data. Patients found to have an incorrect diagnosis of carotid stenosis were excluded. A total of 863 patients, 392 symptomatic and 471 asymptomatic, were included in the analysis.

The baseline characteristics were comparable to those seen in previous studies, as shown in [Table 1](#).

Table 1 Baseline characteristics

| | Asymptomatic N=471 | Symptomatic N=392 |
|--|---------------------------|--------------------------|
| Mean age | 69.4 | 68.6 |
| Sex ratio (F:M) | 197:274 (41.8%: 58.2%) | 135:258 (34.4%: 65.6%) |
| Uncorrected hypertension | 60.9% | 64.4% |
| Uncorrected hyperchol (total cholesterol >5) | 44.4% | 51.1% |
| DM | 15.5% | 23.9% |
| Previous stroke/TIA | 33.8% | 29.5% |
| Coronary heart disease | 52.9% | 46.1% |
| AF | 8.5 | 9.9% |
| Smoking (current smoker or history of smoking) | 74.1% | 76.6% |
| APT | | |
| Aspirin | 58.6% | 49.6 |
| Clopidogrel | 4.0% | 2.8% |
| DAPT | 3.4% | 3.6% |
| Warfarin | 5.7% | 4.8% |
| Antihypertensive therapy (any) | 67.9% | 65.9% |

Notes: Baseline characteristics for asymptomatic and symptomatic patients, after data collection from "all comers" patient record data.

Distribution of Carotid Stenosis

At the time of presentation, symptomatic patients had a distribution of carotid disease heavily skewed towards high degree ipsilateral stenosis. The distribution of ipsilateral and contralateral carotid stenosis is shown in [Figures 1 and 2](#).

For the asymptomatic patients, the distribution of right and left sided carotid stenosis was quite similar for the two sides, as shown in [Figures 3 and 4](#).

Effect of Surgical Treatment

We would like to explore the statistical model with time dependent variables and multiple endpoints by comparing the results to those of a traditional Cox regression analysis of the same data set. The standard Cox regression analysis was performed with survival time to the first ipsilateral stroke for symptomatic patients, and time to first event for asymptomatic patients. Among symptomatic patients, a total of 65 new ipsilateral strokes occurred, of which 49 were attributed to carotid stenosis. Including recurrent events, a total of 76 events occurred, of which 61 were ipsilateral to the side of the initial symptoms. Among asymptomatic patients, a total of 87 strokes occurred, of which 56 were attributed to the stenotic carotid artery. Including recurrent events, a total of 74 attributable events occurred.

As shown in [Table 2](#), very little impact on the risk of stroke is seen by individual variables, but a strong risk improvement occurred in patients who had carotid surgery performed after the index event. This is illustrated for the symptomatic group in [Figure 5](#).

A similar result is seen if the endpoint is defined as any ipsilateral stroke (ie ischemic stroke from any cause), but with a weaker protective effect of the carotid surgery, as illustrated in [Figure 6](#).

If the endpoint is defined as any stroke (any side, any cause) or death, the protective effect of surgery after the index stroke is still evident, but less pronounced, as shown in [Figure 7](#).

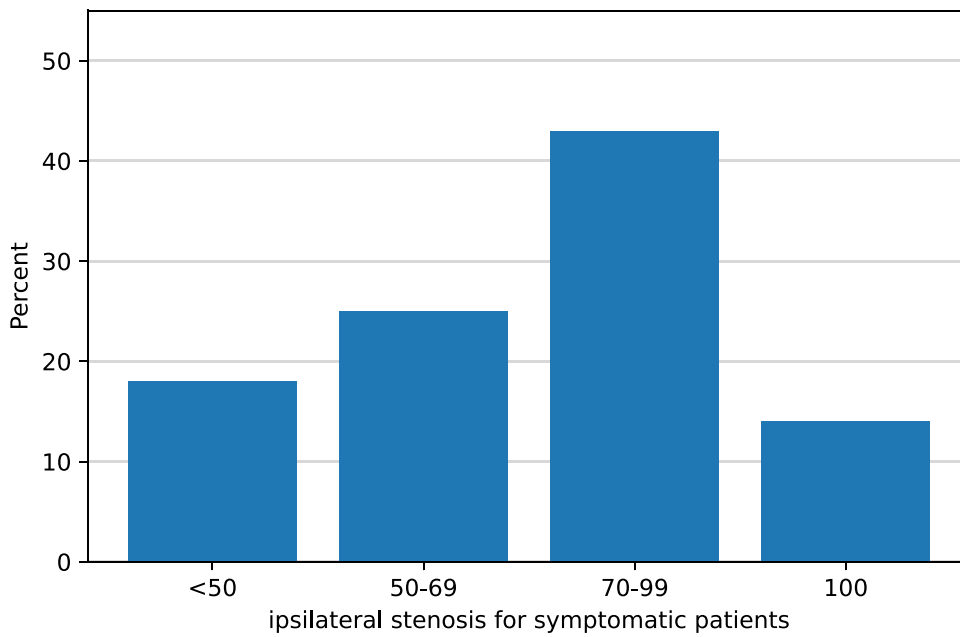


Figure 1 Distribution of ipsilateral carotid stenosis severity in patients included after a cerebral ischemic event.

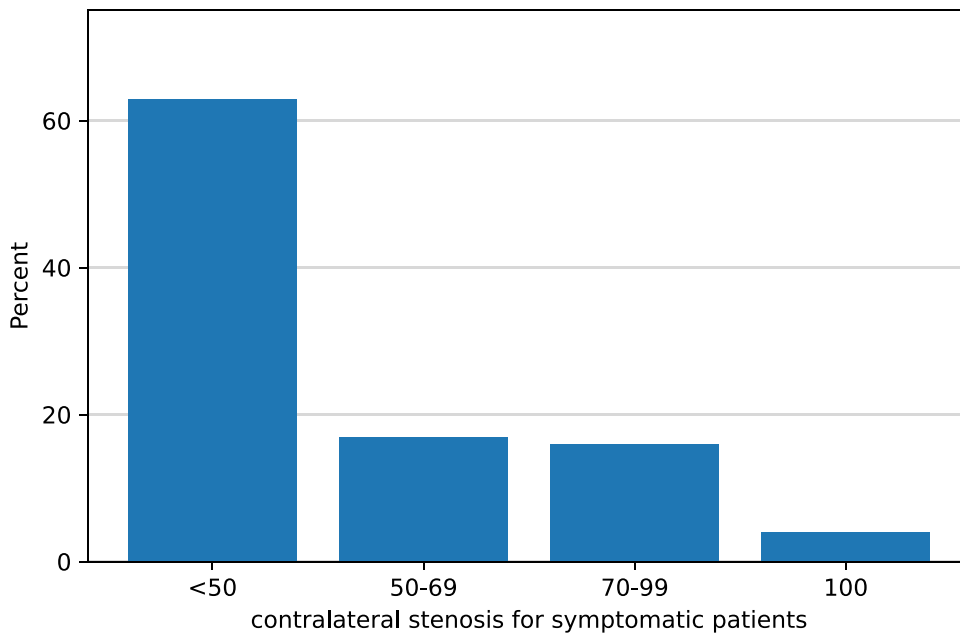


Figure 2 Distribution of contralateral carotid stenosis severity in patients included after a cerebral ischemic event.

When patient record data are analysed, the estimated protective effect of CEA is suspected to be higher than in analyses of RCT data, as patients selected for surgery are believed to have fewer comorbidities, better life expectancy, and lower stroke severity. To illustrate this, we explore the mRS after the index event for those with an index stroke (ie Not TIA), as shown in Figure 8.

This shows that a larger percentage of those who did not undergo carotid surgery had a major index stroke with mRS >2.

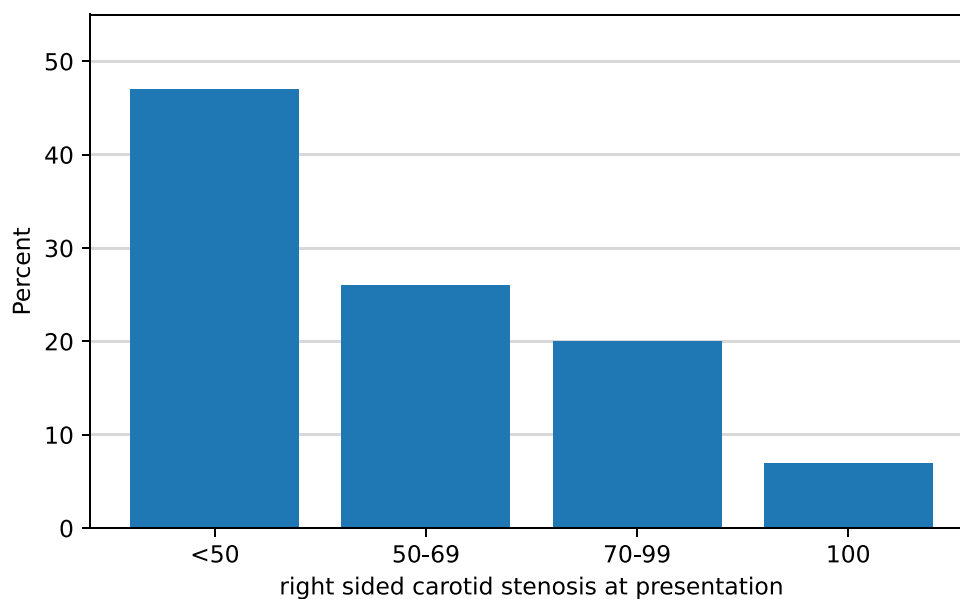


Figure 3 Distribution of right sided carotid stenosis severity in asymptomatic patients.

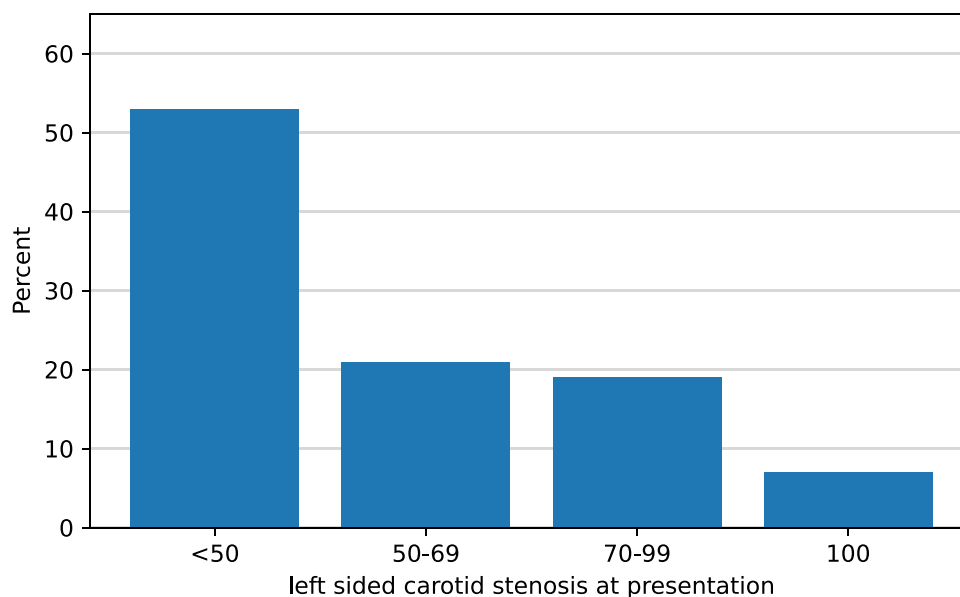


Figure 4 Distribution of left sided carotid stenosis severity in asymptomatic patients.

Surgical Risk

In the 392 patients who were symptomatic at presentation, 194 received surgical treatment for the carotid stenosis. Of these, 10 suffered a postoperative ipsilateral stroke (4,9%).

In the 471 patients who were asymptomatic at the time of diagnosis of the carotid stenosis, 40 underwent surgery for asymptomatic disease, of which four had bilateral surgery. In 44 procedures, two postoperative strokes occurred, one ipsilateral and one contralateral (4,5%, or 2,3% if only ipsilateral strokes are considered). The patient with a contralateral stroke had 80% ICA stenosis and vertebral artery occlusion on the non-operated side.

60 patients that were asymptomatic at the time of diagnosis of the carotid disease, received treatment for symptomatic disease which occurred at a later stage. Of these, 56 had one procedure, three had two procedures and one had three

Table 2 Stroke risk by traditional Cox regression analysis

| Variable | Asymptomatic at Diagnosis, Time to First Event | | Included after Symptoms of a Cerebral Ischemic Event, Time to First New Ipsilateral Event | |
|--|--|------|---|------|
| | HR (95% CI) | p | HR (95% CI) | p |
| Age, years (at time of diagnosis) | 1.02 (0.98, 1.06) | 0.35 | 1.01 (0.98, 1.04) | 0.64 |
| Sex (f) | 0.82 (0.46, 0.48) | 0.51 | 0.96 (0.51, 1.82) | 0.91 |
| Systolic blood pressure at time of diagnosis, mmHg | 1.01 (0.99, 1.02) | 0.34 | 1.01 (1.00, 1.02) | 0.15 |
| Diabetes mellitus | 0.67 (0.32, 1.39) | 0.28 | 0.86 (0.42, 1.77) | 0.68 |
| Current smoker | 0.82 (0.46, 1.45) | 0.49 | 1.34 (0.63, 2.85) | 0.44 |
| No statin therapy after diagnosis | 1.27 (0.64, 2.50) | 0.50 | 1.02 (0.45, 2.32) | 0.97 |
| No anticoagulation after diagnosis | 2.26 (0.85, 6.06) | 0.10 | 4.43 (1.04, 18.54) | 0.04 |
| No antihypertensive treatment | 0.52 (0.28, 0.98) | 0.05 | 0.96 (0.47, 1.97) | 0.92 |
| Carotid stenosis by side | Right side: | | Ipsilateral side: | |
| <50% | ref | | ref | |
| 50-69% | 1.24 (0.60, 2.55) | 0.57 | 0.80 (0.31, 2.10) | 0.65 |
| 70-99% | 2.40 (1.18, 4.89) | 0.16 | 1.85 (0.78, 4.37) | 0.16 |
| 100% | 2.4 (0.96, 5.88) | 0.62 | 1.59 (0.61, 4.15) | 0.34 |
| Carotid stenosis by side | Left side: | | Contralateral side | |
| <50% | ref | | ref | |
| 50-69% | 2.00 (0.97, 4.12) | 0.06 | 0.55 (0.21, 1.43) | 0.22 |
| 70-99% | 2.08 (1.01, 4.28) | 0.46 | 0.58 (0.22, 1.51) | 0.26 |
| 100% | 1.89 (0.67, 5.34) | 0.23 | 0.41 (0.06, 3.22) | 0.41 |
| CEA | 2.04 (0.75, 5.55) | 0.16 | 0.32 (0.14, 0.71) | 0.01 |

Notes: Stroke risk associated with individual risk factors when a traditional cox regression analysis is performed, with separate analyses for patients who were asymptomatic and symptomatic at the time of inclusion.

procedures, of which the last was a CAS. In a total of 65 procedures, seven strokes occurred (10,8%). This incidence is higher than the expected stroke risk in association to the procedure. The CAS for a second procedure on the same side resulted in a stroke, but even with this procedure excluded, the stroke incidence after surgery is 9,4% which is surprisingly high. The reason for this is not known, although the small number of procedures could account for a random outcome without true statistical significance. In a pooled analysis of both patients that were symptomatic at the time of registration and patients who became symptomatic during follow-up, the total number of procedures for symptomatic disease was 258 (excluding the CAS redo). This resulted in a total of 16 postoperative strokes (6,2%) occurring after treatment of symptomatic disease.

The described patient journal data is processed in order to allow analysis of multiple endpoints and time dependent variables, and results are shown in Table 3. When compared with the traditional Cox analysis in Table 2, the introduction of the extended model seems to impact individual risk predictors. Interestingly, the carotid stenosis on the side of the event is not significantly associated with a recurrent event for symptomatic patients in a traditional Cox analysis, but shows significant association when the extended model is applied, with a Hazard ratio (HR) of 2.25 for ipsilateral

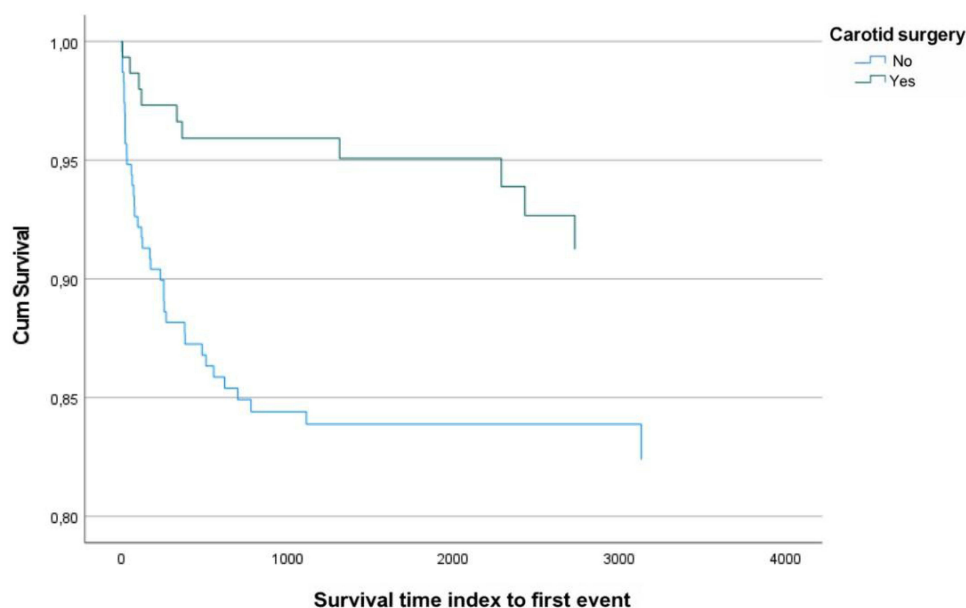


Figure 5 Survival time at mean of covariates in days to the first ipsilateral stroke in symptomatic patients for those who had surgery performed after the index event and those who did not, when the stroke event is suggested most likely to be caused by the carotid stenosis after ASCOD classification.

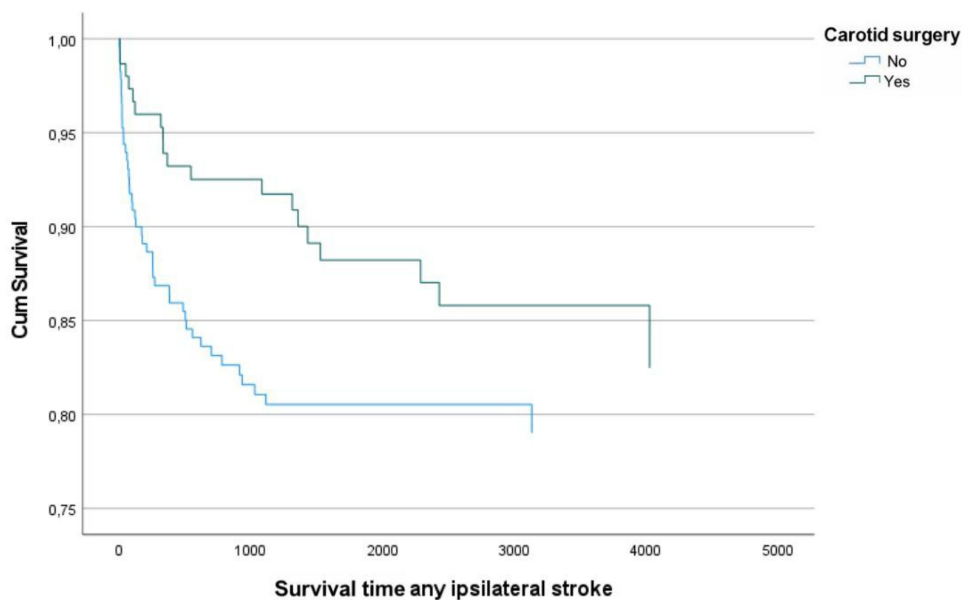


Figure 6 Survival time at mean of covariates in days to the first ischemic ipsilateral stroke in symptomatic patients for those who had surgery performed after the index event and those who did not, when the cause of the stroke is not specified.

stenosis of 70–99%. The same applies to the asymptomatic patients, where a high degree right sided carotid stenosis is significantly associated with a right sided event, HR 4.98, and a high degree left sided carotid stenosis is significantly associated with left sided events, HR 5.23. This is not clearly significant in the traditional Cox analysis, where the according HRs are 2.40 and 2.08, respectively.

This could imply an additional value in risk prediction, but further exploration is considered necessary to evaluate the results. The results have been supported by model tests, which are described in a separate article.

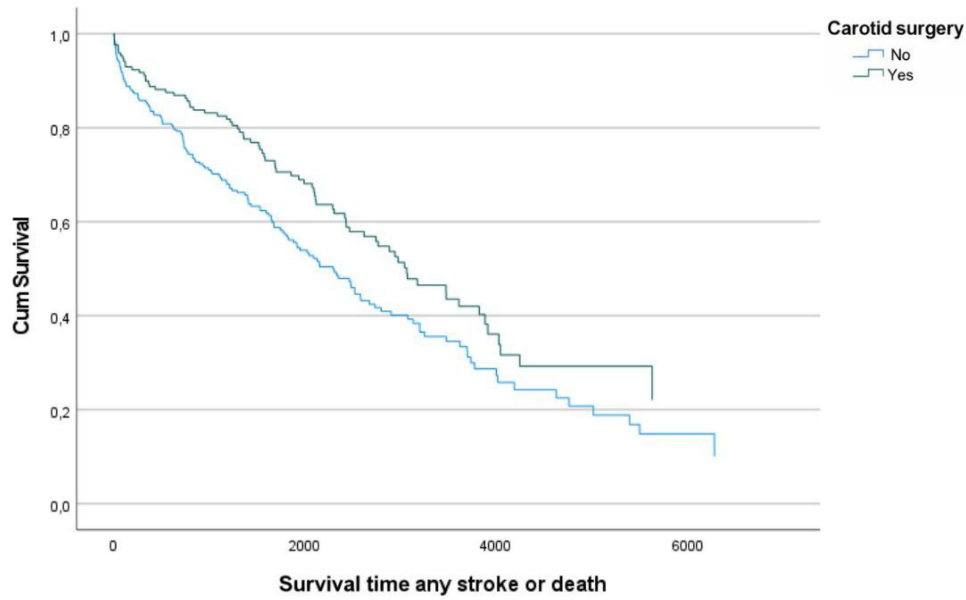


Figure 7 Survival time at mean of covariates in days to stroke or death from any reason in symptomatic patients for those who had surgery performed after the index event and those who did not, when neither the side or cause of the stroke is specified.

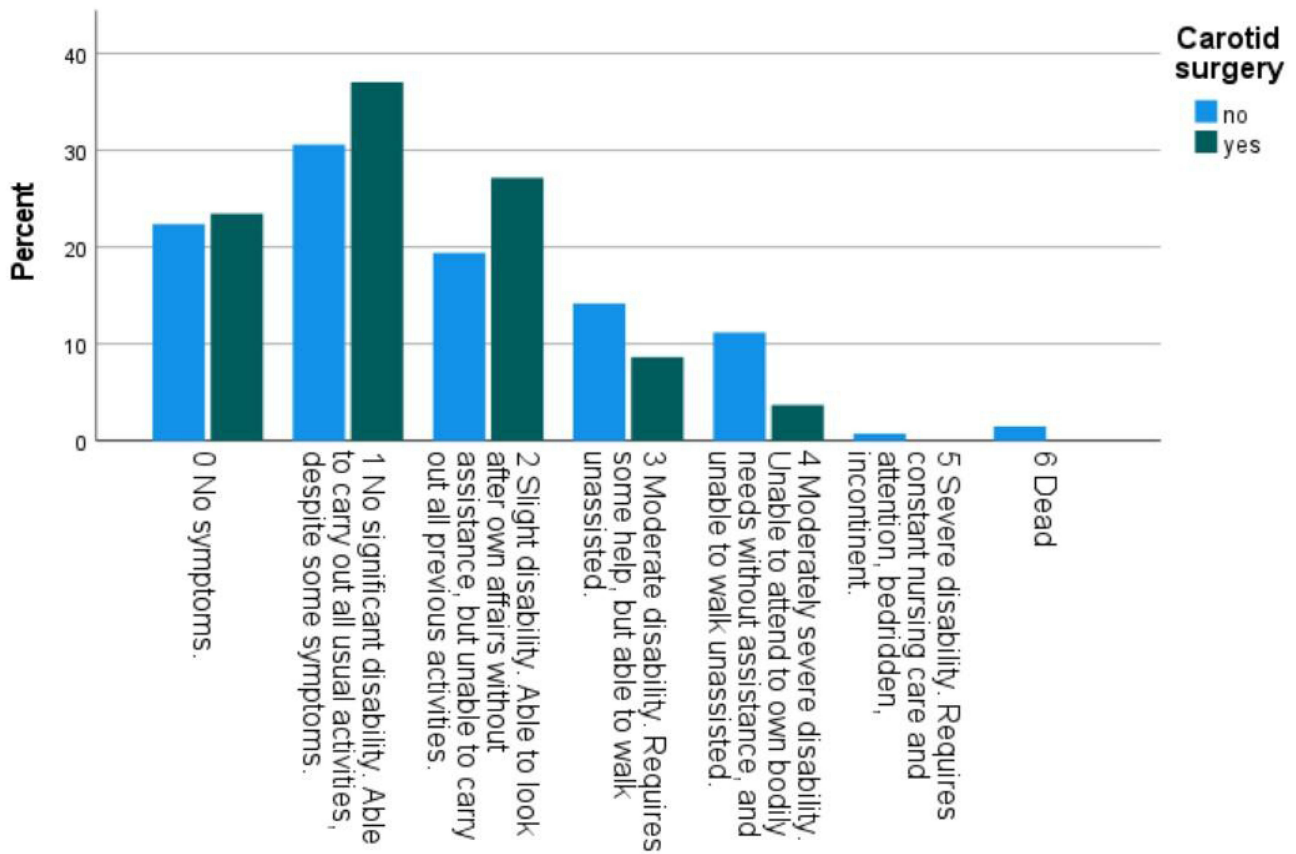


Figure 8 Analysis of stroke severity by mRS for symptomatic patients who had an ischemic stroke as the index event, comparing those who were selected for surgery to those who were not.

Table 3 Survival analysis by an Extended Cox Regression Model

| Asymptomatic at diagnosis | Extended Cox Model, Right Sided Events | | Extended Cox Model, Left Sided Events | |
|--|---|----------|---|----------|
| | HR (95% CI) | p | HR (95% CI) | p |
| Age, years (at time of diagnosis) | 1.02 (0.97, 1.07) | 0.36 | 1.01 (0.95, 1.07) | 0.71 |
| Sex (f) | 0.79 (0.35, 1.77) | 0.56 | 2.16 (1.05, 4.48) | 0.04 |
| Systolic blood pressure (at time of diagnosis), 10mmHg | 1.00 (0.98, 1.01) | 0.87 | 1.10 (0.99, 1.03) | 0.32 |
| Diabetes mellitus | 0.72 (0.23, 2.24) | 0.57 | 2.80 (1.26, 6.25) | 0.01 |
| Current smoker | 1.31 (0.65, 2.65) | 0.45 | 1.67 (0.79, 3.64) | 0.18 |
| No statin therapy after diagnosis | 0.60 (0.23, 1.51) | 0.27 | 0.81 (0.35, 1.86) | 0.62 |
| No anticoagulation after diagnosis | 0.63 (0.21, 1.87) | 0.40 | 0.23 (0.03, 1.81) | 0.16 |
| No antihypertensive treatment | 2.10 (0.80, 5.47) | 0.13 | 1.06 (0.44, 2.55) | 0.89 |
| Carotid stenosis, right side 0 | ref | | ref | |
| <50% | 1.95 (0.57, 6.75) | 0.29 | 1.02 (0.28, 3.65) | 0.98 |
| 50-69% | 1.02 (0.26, 3.97) | 0.03 | 1.27 (0.40, 4.04) | 0.69 |
| 70-99% | 4.01 (1.14, 14.07) | 0.03 | 1.97 (0.58,6.67) | 0.28 |
| 100% | 4.98 (1.39, 17.80) | 0.01 | 0.35 (0.05, 2.27) | 0.27 |
| Carotid stenosis, left side 0 | ref | | ref | |
| <50% | 0.72 (0.22, 2.36) | 0.59 | 3.60 (1.29, 10.08) | 0.04 |
| 50-69% | 1.15 (0.41, 3.24) | 0.79 | 4.18 (1.40, 12.48) | 0.01 |
| 70-99% | 1.18 (0.38, 3.64) | 0.77 | 5.23 (2.04, 13.38) | 0.00 |
| 100% | 3.07 (0.86, 10.94) | 0.08 | 1.96 (0.58, 6.68) | 0.28 |
| Right sided TIA | 1.89 (0.50,7.19) | 0.35 | 1.25 (0.25, 6.25) | 0.78 |
| Left sided TIA | 0.90 (0.28, 2.93) | 0.86 | 2.02 (1.08, 3.78) | 0.03 |
| TIA, unknown side | 1.03 (0.24, 4.42) | 0.96 | Err (few events) | err |
| Included after symptoms of a cerebral ischemic event | Extended cox model, time to ipsilateral events | | Extended cox model, time to contralateral events | |
| Variable | HR (95% CI) | p | HR (95% CI) | p |
| Age, years (at index event) | 1.01 (0.99, 1.03) | 0.41 | 1.01 (0.95, 1.07) | 0.82 |
| Sex (f) | 0.83 (0.47, 1.45) | 0.50 | 0.48 (0.16, 1.46) | 0.20 |
| Systolic blood pressure (at time of diagnosis), 10 mmHg | 1.00 (0.99, 1.01) | 0.70 | 1.22 (1.00, 1.04) | 0.06 |
| Diabetes mellitus | 0.97 (0.51, 1.82) | 0.92 | 0.46 (0.08, 2.60) | 0.38 |
| Current smoker | 0.85 (0.50, 1.47) | 0.57 | 0.99 (0.31, 3.17) | 0.98 |
| No statin therapy after diagnosis | 0.80 (0.35, 1.81) | 0.59 | 0.41 (0.04, 3.97) | 0.44 |
| No anticoagulation after diagnosis | 3.95 (1.03,15.19) | 0.05 | - | - |
| No antihypertensive treatment | 0.99 (0.53, 1.84) | 0.98 | 1.03 (0.32, 3.32) | 0.96 |

(Continued)

Table 3 (Continued).

| | | | | |
|---|-------------------|------|----------------------|------|
| Carotid stenosis ipsilateral to index side 0 | ref | | ref | |
| <50% | 1.23 (0.46, 3.31) | 0.68 | 0.73 (0.06, 8.51) | 0.80 |
| 50-69% | 1.61 (0.70, 3.71) | 0.27 | 1.29 (0.21, 7.96) | 0.78 |
| 70-99% | 2.25 (1.03, 4.94) | 0.04 | 0.52 (0.04, 7.16) | 0.62 |
| 100% | 2.93 (1.45, 3.07) | 0.00 | 3.67 (0.90, 14.90) | 0.07 |
| Carotid stenosis contralateral to index side 0 | ref | | ref | |
| <50% | 1.48 (0.81, 2.73) | 0.20 | 6.30 (0.74, 53.78) | 0.09 |
| 50-69% | 0.66 (0.30, 1.46) | 0.31 | 15.43 (2.22, 107.23) | 0.01 |
| 70-99% | 0.55 (0.21, 1.43) | 0.22 | 12.08 (1.30, 112.26) | 0.03 |
| 100% | 0.49 (0.09, 2.68) | 0.49 | - | - |
| New ipsilateral TIA | 1.13 (0.41, 3.07) | 0.82 | 0.46 (0.17, 1.24) | 0.12 |
| New contralateral TIA | 1.75 (0.62, 4.95) | 0.29 | 4.05 (1.18, 13.88) | 0.03 |
| New TIA unknown side | 1.54 (0.57, 4.13) | 0.40 | 20.86 (3.20, 136.19) | 0.00 |

Notes: Stroke risk associated with individual risk factors when time dependency of key variables and multiple endpoint analysis is included, with separate analyses for patients who were asymptomatic and symptomatic at the time of inclusion. For asymptomatic patients both sides are analysed separately, whereas in the symptomatic patients, the side ipsilateral and contralateral to the index event are separately analysed. For the contralateral side, few events occurred, and confidence intervals are large due to model instability.

Discussion

In the establishment of new risk prediction tools, a large data source is desirable in order to develop a reliable and stable model. The lower the event frequency and the higher the number of variables, the larger the need for power. Our moderate data material is only intended for model exploration, and is not yet applicable in risk prediction.

With the standard use of digitalised patient record systems, data derived from this gold mine of information could play an important role in future prognostic models. Extraction of data from free text information by the use of machine learning could make anonymised data collection possible. Regardless, even for anonymised data, ethical considerations about ownership, data protection and confidentiality represent a controversy for the use of this information in research. The European Data Protection Regulation (GDPR) of 2018 serves to protect the data subject, and by that intensifies the responsibility of the processors. By article 11 of GDPR, processing which does not require identification can still be performed as long as adherence to these regulations is fulfilled, and it seems likely that anonymised data collection could become possible in future research. If so, evaluating whether patient record data are suitable for risk prediction is considered a first step. Evaluating manually extracted data allows for a controlled and verified analysis of this material prior to potential digital data extraction.

Selection bias is a potential source of error in our current data material for individuals included by consent. The Tromso study has a randomised approach to invitation, and randomly assigns participants to the “special studies”, where the carotid ultrasound was performed. There will still be a selection bias in regards to volunteering participants. The bias could be reduced by all-comer analysis, which could be performed if the proposed anonymized free text data extraction is made possible.

In free text patient records, the lack of uniformity of information and tests represent a challenge. Spot parameter measurements, operator dependency, language variability and subjective evaluation could all generate a potential “free text error”. If the prognostic model is to be applied in a clinical setting, the patient record derived data could still be postulated to resemble that of real-life observations in a representative way. Missing data will naturally occur in a real-life dataset, and we have intentionally allowed this to be omitted.

The various causes of cerebral ischemia represent an issue of uncertainty when working with stroke risk. Numerous carotid stenosis studies still present their results without commenting on their endpoint management. The easiest approach would be to include all ischemic strokes as endpoints, and assume an even distribution of strokes from other causes. This seems counterintuitive in cases where the stroke is obviously caused by i.e. cardiac embolization, such as in the presence of bilateral shower embolies. Knowing from previous studies that a stroke is caused by the carotid stenosis in only about half of those who have the condition,¹⁴ at the very least an opinion on stroke subclassification is considered appropriate. Our emphasis is a best effort in standardising stroke subclassification, to predict the actual risk of thromboembolic strokes from carotid lesions, but to complement this by additionally all cause ischemic stroke analyses. Even after ASCOD classification, there is some uncertainty in relation to the causality of the carotid disease when an ischemic stroke occurs. We consider this a challenge which is not yet overcome in stroke research.

The decision to perform CEA is expected to be partly subjective, depending on the surgeon's assessment of patient frailty and the neurologist's evaluation of the cause of the preceding event. This is a key rationale for applying a more suitable statistical model that incorporates recurrent event analysis, as recurrent events while awaiting treatment and in relation to the procedure can be captured, and the long-term effects of successful treatment can be assessed.

Incorporating recurrent event analysis and time dependent variables in the model is considered to be a valuable supplement to traditional survival analyses in cardiovascular risk prediction. In evaluation of any prophylactic treatment, sufficient observation time to estimate the long term effect of the prevention strategy is considered desirable. For any cardiovascular disease, repeat events of a specific condition in the same individual are relatively frequent, and expected to be encountered particularly in data material where the observation time is long. Still, the application of recurrent event analysis in cardiovascular research is quite uncommon. The same applies for time dependent variables, which are also expected to impact risk in relation to recurrent events, i.e. increase in stroke risk with progressive carotid stenosis. We have not found any other studies describing the combination of time dependent variables and recurrent event analysis in the same model. Recurrent event analysis is known to carry a risk of type I error, where a false positive result occurs due to overestimation of risk. It is not known if this can be corrected by the introduction of time dependent variables. Thorough model testing is considered necessary to evaluate model performance, and has been performed in a separate study.

We have selected only a few risk predictors for time dependent analysis for practical reasons, but consider it an interesting next step to include other parameters, such as smoking, biochemical markers, alterations in drug therapy, and plaque property alterations. This data extraction would yield large and complex datasets that could be managed using machine learning methods.

Conclusion

We consider the free text data material as representative when compared to baseline properties from previous carotid studies. The dataset appears sufficient for a study of the performance of a novel risk prediction model, although a larger sample size would be desirable given the relatively low event frequency. The introduction of multiple endpoint analysis and time dependent variables seems to impact individual risk predictors, implying a potential value in risk prediction. Model tests are still necessary to confirm this. This has been performed in a separately published study.

Abbreviations

CEA, Carotid endarterectomy; CAS, Carotid stenting; AI, Artificial intelligence; TIA, Transient ischemic attack; mRS, modified Rankin scale; REC, The Regional Ethical Committee; HR, Hazard Ratio.

Funding

This work was funded by a part time research position at the University Hospital of North Norway for Kjersti Hervik.

Disclosure

The authors have no competing interests in this work.

References

- Munster AB, Franchini AJ, Qureshi MI, Thapar A, Davies AH. Temporal trends in safety of carotid endarterectomy in asymptomatic patients: systematic review. *Neurology*. 2015;85(4):365–372. doi:10.1212/WNL.0000000000001781
- Rothwell PM, Warlow CP; European Carotid Surgery Trialists' Collaborative Group. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. *Lancet*. 1999;353(9170):2105–2110.
- Preiss D, Kristensen SL. The new pooled cohort equations risk calculator. *Can J Cardiol*. 2015;31(5):613–619. doi:10.1016/j.cjca.2015.02.001
- Chaudhry SA, Afzal MR, Kassab A, Hussain SI, Qureshi AI. A new risk index for predicting outcomes among patients undergoing carotid endarterectomy in large administrative data sets. *J Stroke Cerebrovasc Dis*. 2016;25(8):1978–1983. doi:10.1016/j.jstrokecerebrovasdis.2016.01.023
- Dharmarajah B, Thapar A, Salem J, Lane TR, Leen EL, Davies AH. Impact of risk scoring on decision-making in symptomatic moderate carotid atherosclerosis. *Br J Surg*. 2014;101(5):475–480. doi:10.1002/bjs.9461
- Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017;69:S36–S40. doi:10.1016/j.metabol.2017.01.011
- Naylor AR, Sillesen H, Schroeder TV. Clinical and imaging features associated with an increased risk of early and late stroke in patients with symptomatic carotid disease. *Eur J Vasc Endovasc Surg*. 2015;49(5):513–523. doi:10.1016/j.ejvs.2015.01.011
- European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European carotid surgery trial (ECST). *Lancet*. 1998;351(9113):1379–1387. doi:10.1016/S0140-6736(97)09292-1
- Nicolaides AN, Kakkos SK, Kyriacou E, et al. Asymptomatic internal carotid artery stenosis and cerebrovascular risk stratification. *J Vasc Surg*. 2010;52(6):1486. doi:10.1016/j.jvs.2010.07.021
- Halliday A, Harrison M, Hayter E, et al. 10-year stroke prevention after successful carotid endarterectomy for asymptomatic stenosis (ACST-1): a multicentre randomised trial. *Lancet*. 2010;376(9746):1074–1084. doi:10.1016/S0140-6736(10)61197-X
- Cunningham EJ, Bond R, Mehta Z, et al. Long-term durability of carotid endarterectomy for symptomatic stenosis and risk factors for late postoperative stroke. *Stroke*. 2002;33(11):2658–2663. doi:10.1161/01.STR.0000034397.72390.D3
- Singh TD, Kramer CL, Mandrekar J, Lanzino G, Rabinstein AA. Asymptomatic carotid stenosis: risk of progression and development of symptoms. *Cerebrovasc Dis*. 2015;40(5–6):236–243. doi:10.1159/000439179
- Lochen M-L, Gram IT, Mannsverk J, et al. Association of occasional smoking with total mortality in the population-based Tromsø study, 2001–2015. *BMJ Open*. 2017;7(12):e019107. doi:10.1136/bmjopen-2017-019107
- Cheng SF, Brown MM, Simister RJ, Richards T. Contemporary prevalence of carotid stenosis in patients presenting with ischaemic stroke. *Br J Surg*. 2019;106(7):872–878. doi:10.1002/bjs.11136
- Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Wolf ME, Hennerici MG. The ASCOD phenotyping of ischemic stroke (updated ASCO phenotyping). *Cerebrovasc Dis*. 2013;36(1):1–5. doi:10.1159/000352050
- Prentice PRL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68(2):373–379. doi:10.1093/biomet/68.2.373
- Gasecki AP, Eliasziw M, Ferguson GG, Hachinski V, Barnett HJ; North American Symptomatic Carotid Endarterectomy Trial (NASCET) Group. Long-term prognosis and effect of endarterectomy in patients with symptomatic severe carotid stenosis and contralateral carotid stenosis or occlusion: results from NASCET. *J Neurosurg*. 1995;83(5):778–782.
- Turley RS, Freischlag K, Truong T, et al. Carotid stenting and endarterectomy and contralateral carotid occlusion. *J Vasc Surg*. 2019;70(3):824–831. doi:10.1016/j.jvs.2018.12.039
- Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;10:1100–1120.

Vascular Health and Risk Management

Publish your work in this journal

Vascular Health and Risk Management is an international, peer-reviewed journal of therapeutics and risk management, focusing on concise rapid reporting of clinical studies on the processes involved in the maintenance of vascular health; the monitoring, prevention and treatment of vascular disease and its sequelae; and the involvement of metabolic disorders, particularly diabetes. This journal is indexed on PubMed Central and MedLine. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/vascular-health-and-risk-management-journal>

Dovepress
Taylor & Francis Group