Predicting 1- 3-, and 5-year mortality after surgery for colorectal cancer using a Danish quality assurance database

Andreas Weinberger Rosen, Karoline Bendix Bräuner, Mikail Gögenur, Viviane Annabelle Lin, Johan Stub Rønø Clausen, Eldar Allakhverdiiev, Adamantia Tsouchnika, Rasmus Peuliche Vogelsang, Ismail Gögenur

Background

Colorectal cancer (CRC) is the 3rd most common malignant disease and the second most deadly globally, with more than 1.900.000 new cases estimated in 2020 and 935.000 deaths (1). Surgery is the cornerstone of treatment for patients with curable CRC. Additionally, some patients benefit from pre-or postoperative oncological interventions. In the past decades, short-term survival has improved for patients operated for colorectal cancer in Denmark, partly due to standardization and centralization of treatment (2). However, it is becoming apparent that some subgroups need personalization of the treatment to reduce over-or under treatment. Ensuring patients undergoing surgery for CRC get optimal treatment is as important as for all other patients. Still, the combination of treatment options puts patients at risk for several complications and conditions, which can result in morbidity or mortality and reduced quality of life or an escalation of treatments. Due to patients' risk preferences are diverse and the relationships between interventions and important outcomes are complex, a single measurement defining an optimal treatment does not exist. Potentially looking at several outcomes that can describe morbidity, mortality, and quality of life can be used to guide treatment decisions and ensure treatment options are optimized to patients' preferences. This study aims to create a prediction model for long term survival (defined here as 1-, 3-, and 5-years after surgery) for patients operated for colorectal cancer based upon data from a national Danish quality assurance database. The ultimate goal of creating these models is using them together with other models for key outcomes to give clinicians a comprehensive overview of patients' risk profile before surgery for colorectal cancer, thus supporting them in providing treatment recommendations.

Methods

Data from the Danish Colorectal Cancer Group's database, a national quality assurance database started in 2001, which covers >99% of patients diagnosed with colorectal cancer in Denmark and their subsequent treatment were converted to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). The information covered includes demographic information, initial contact and treatment plan, patient characteristics, preoperative and postoperative oncological treatment, surgical treatment, complications, survival, pathology, and radiology findings (3). The study was designed in ATLAS (4) using all patients with a diagnosis of CRC that underwent surgery with at least 365, 1095, or 1825 days of follow-up as the target cohorts. Death from any cause was considered as the outcome event. Date of surgery was used as the index date, and time-at-risk for 1-, 3- and 5- years were defined as 365, 1095, and 1825 days after surgery. Covariates known before surgery or relating to the upcoming surgery were included as potential covariates in the model. A team of medical doctors curated the list to exclude any covariates not considered medically relevant. A LASSO logistic regression model was trained using the Patient-Level-Prediction package (5). The dataset was split into a 75% training set and a 25% test set. The training was done with 3-fold cross-validation. The model was evaluated using the area under the receiver operating characteristic (AUROC), the area under the precision-recall curve (AUPRC), and calibration.

Results

76.828 patients were considered in the entire database. The number of eligible patients for each prediction model is summarized together with the number of outcomes, incidence AUROC and AUPRC in table 1. The total count af variables in each model and the top 5 covariates with the highest and lowest weight for each model are shown in table 2.

Model	Number of patients at risk	Number of patients with outcome	Incidence (%)	AUROC (95% CI)	AUPRC	Brier Score
1 year mortality	**63333**	**10587**	**16.72**	**0.863 (0.855-0.87)**	**0.62**	**0.09**
3 years mortality	**55819**	**18142**	**32.50**	**0.844 (0.837-0.852)**	**0.77**	**0.14**
5 years mortality	**47333**	**20971**	**44.31**	**0.833 (0.826-0.84)**	**0.82**	**0.16**

Table 1. Summarizes the size of the cohort, number of outcomes, and model performance for the different time-at-risks. AUROC: Area under the receiver operating characteristic, CI: Confidence interval, AUPRC: Area under the precision-recall curve.

1-year mortality Total number of covariates: 263		3-years mortality Total number of covariates: 248		5-years mortality Total number of covariates: 203	
Covariate	Weight	Covariate	Weight	Covariate	Weight
Creation of defunctioning ileostomy	1.2491	Endoscopic insertion of permanent colonic stent	2.1279	Endoscopic insertion of permanent colonic stent	2.4433
Only exploratory surgery, diagnostic laparoscopy or exploratory laparotomy	1.1501	Age group: 95-99	1.607	Only exploratory surgery, diagnostic laparoscopy or exploratory laparotomy	2.0941
Age group: 90-94	1.0428	Only exploratory surgery, diagnostic laparoscopy or exploratory laparotomy	1.508	Age group: 95-99	1.7757
ASA Score 4	0.9658	WHO performance status 4	1.3672	Age group: 90-94	1.5283
Endoscopic insertion of permanent colonic stent	0.9259	Age group: 90-94	1.2988	Construction of stoma, defunctioning ileostomy, colostomy or internal shunt procedure	1.3505
Age group: 45-49	-0.7202	MRI Tx	-0.855	MRI Tx	-1.4963
Age group: 40-44	-0.7174	Intermediate to high grade histologic differentiation	-0.8316	Age group: 45-49	-0.6796
Age group: 55-59	-0.6718	Age group: 25-29	-0.6487	Age group: 30-34	-0.6687
Excision of liver metastasis	-0.5612	Age group: 45-49	-0.6044	Age group: 35-39	-0.5111
Age group: 50-54	-0.5519	Exicion of secondary	-0.4902	Age group: 50-54	0.4923

		malignant neoplasm			

Table 2. Shows the total count of covariates and the top 5 covariates with highest and lowest weight for the 1-year, 3-years, and 5-years mortality model.

The models were generally well-calibrated. Figure 1 shows the calibration plot of the test set of the 5-years mortality model as an example.

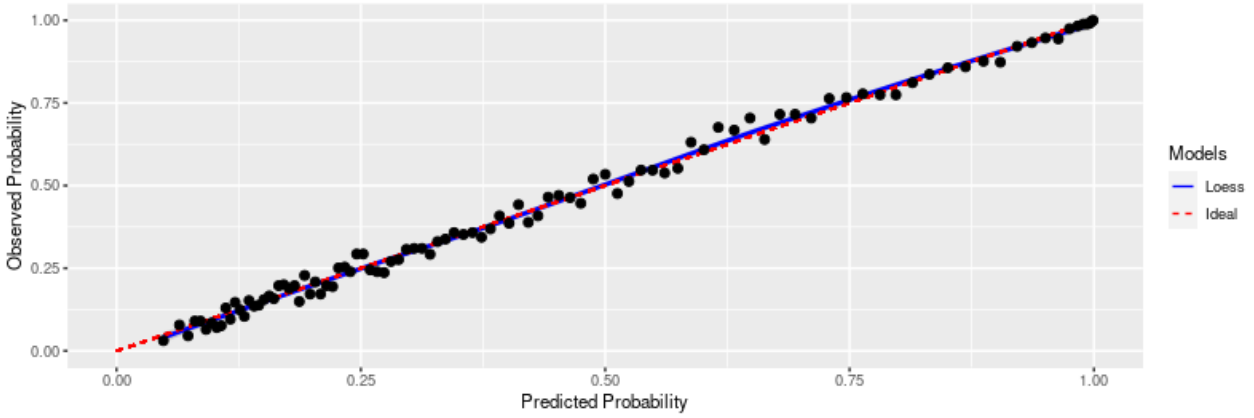

Figure 1. *Shows the calibration plot for the LASSO regression with 5-years time-at-risk for the test set.*

Conclusion

We showed that it is possible to create a reasonable prediction model for long-term postoperative mortality for patients undergoing surgery for colorectal cancer with good discrimination and calibration, using data from a Danish quality assurance database. The database offers highly granular information about patients around the time of operation but limited information about medical history and virtually no information about physiological measurements or exposure to drugs. However, data sources that describe these domains exist (6–8), and due to the unique civil registration number for each participant in the Danish Civil System, possible to merge these sources easily. Representing new areas of patients' health status has the possibility to improve the model performance (9), but the output of risk of death needs to be interpreted holistically, taking other important factors into accounts, such as complications or recurrence, along with patients' preferences. In the future, it could be imagined that a suite of prediction models could be used to support clinical decisions. In the present study, we investigated the case of using variables known before surgery, but additional prediction models at different points in the patient trajectory could help facilitate the contentious evaluation of the optimal strategy, e.g., highlighted in a recent study (10). However, further studies using causal inference methodologies are needed to investigate alternative and optimal treatment (11). In conclusion, using observational Danish healthcare data with the OMOP CDM and OHDSI tools are a promising framework for generating applicable clinical prediction models.

References/Citations

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Iversen LH, Ingeholm P, Gogenur I, Laurberg S. Major reduction in 30-day mortality after elective colorectal cancer surgery: a nationwide population-based study in Denmark 2001-2011. Ann Surg Oncol. 2014/03/04. 2014;21(7):2267–73.
3. Marc Overhage J, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Informatics Assoc. 2012;19(1):54–60.
4. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. 2015;574–8.
5. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Informatics Assoc. 2018;25(8):969–75.
6. Sandegaard JL, Schmidt SAJ, Sørensen HT, Pedersen L, Ehrenstein V, Schmidt M. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol [Internet]. 2015;7:449. Available from: http://dx.doi.org/10.2147/CLEP.S91125
7. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. Int J Epidemiol [Internet]. 2017;46(3):798-798f. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27789670%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5837522
8. Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K. Existing data sources in clinical epidemiology: Laboratory information system databases in Denmark. Clin Epidemiol. 2020;12:469–75.
9. Nielsen AB, Thorsen-Meyer H-C, Belling K, Nielsen AP, Thomas CE, Chmura PJ, et al. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records. Lancet Digit Heal [Internet]. 2019;1(2):e78–89. Available from: https://linkinghub.elsevier.com/retrieve/pii/S258975001930024X
10. Vogelsang RP, Bojesen RD, Hoelmich ER, Orhan A, Buzquurz F, Cai L, et al. Prediction of 90-day mortality after surgery for colorectal cancer using standardized nationwide quality-assurance data. BJS open. 2021;5(3).
11. Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. Chance. 2019;32(1):42–9.