

BMJ Open Hospital readmission risk prediction based on claims data available at admission: a pilot study in Switzerland

Beat Brüngger,[Ⓢ] Eva Blozik

To cite: Brüngger B, Blozik E. Hospital readmission risk prediction based on claims data available at admission: a pilot study in Switzerland. *BMJ Open* 2019;**9**:e028409. doi:10.1136/bmjopen-2018-028409

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-028409>).

Received 6 December 2018
Revised 30 May 2019
Accepted 31 May 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Department of Health Sciences, Helsana Group, Zurich, Switzerland

Correspondence to

Dr Eva Blozik;
eva.blozik@helsana.ch

ABSTRACT

Objectives Evaluating whether future studies to develop prediction models for early readmissions based on health insurance claims data available at the time of a hospitalisation are worthwhile.

Design Retrospective cohort study of hospital admissions with discharge dates between 1 January 2014 and 31 December 2016.

Setting All-cause acute care hospital admissions in the general population of Switzerland, enrolled in the Helsana Group, a large provider of Swiss mandatory health insurance.

Participants The mean age of 138 222 hospitalised adults included in the study was 60.5 years. Patients were included only with their first index hospitalisation. Patients who deceased during the follow-up period were excluded, as well as patients admitted from and/or discharged to nursing homes or rehabilitation clinics.

Measures The primary outcome was 30-day readmission rate. Area under the receiver operating characteristic curve (AUC) was used to measure the discrimination of the developed logistic regression prediction model. Candidate variables were theory based and derived from a systematic literature search.

Results We observed a 30-day readmission rate of 7.5%. Fifty-five candidate variables were identified. The final model included pharmacy-based cost group (PCG) cancer, PCG cardiac disease, PCG pain, emergency index admission, number of emergency visits, costs specialists, costs hospital outpatient, costs laboratory, costs therapeutic devices, costs physiotherapy, number of outpatient visits, sex, age group and geographical region as predictors. The prediction model achieved an AUC of 0.60 (95% CI 0.60 to 0.61).

Conclusions Based on the results of our study, it is not promising to invest resources in large-scale studies for the development of prediction tools for hospital readmissions based on health insurance claims data available at admission. The data proved appropriate to investigate the occurrence of hospitalisations and subsequent readmissions, but we did not find evidence for the potential of a clinically helpful prediction tool based on patient-sided variables alone.

INTRODUCTION

Hospital readmissions describe events in which patients have to return to the same or another hospital shortly after the discharge.

Strengths and limitations of this study

- This study helps to tailor future research related to the development of prediction tools for hospital readmissions.
- It extends the existing literature by focusing on potential predictors available in claims data and at the time of the hospitalisation.
- The study population was not limited to specific conditions or age groups.
- Eligible potential predictors were selected based on a systematic literature search.
- Only a small part of the predictors used in existing models could be extracted from Swiss claims data and were available at the time of hospitalisation.

Hospital readmissions are both common and costly.¹ A part of the readmissions is considered to be preventable.² They are a promising aspect for quality improvement measures, on the healthcare system and/or hospital level.³

Previous research indicated that it is possible to effectively reduce readmissions with interventions such as patient-centred discharge instructions or scheduling of a follow-up visit prior to discharge.^{4,5}

In Switzerland, it is currently discussed how digitalisation could be used to increase efficiency in the healthcare system. In particular, it is postulated that the use of routine data (such as claims data) does not reflect its potential to increase the efficiency of the healthcare system.⁶

Hospital readmission prediction models (HRPMs) calculate the risk of a patient for a subsequent readmission based on individual characteristics, for example, age or comorbidities. Claims data include detailed information on healthcare service utilisation and costs on the patient level and a systematic review of HRPM showed they are a useful basis for HRPM.^{7,8}

Avoiding hospital readmissions has been given a high priority in Switzerland.⁹ Rates are calculated systematically and published on a hospital level. Thus, developing tools

to identify patients at high risk for a subsequent rehospitalisation as a basis for tailored interventions is relevant for hospitals, quality institutions and cost payers. Given the fact that health insurance claims data are nationwide available, pre-existing and standardised, they might be very useful. They might overcome the barrier of building additional data collections.

The aim of the present study is to develop a proof-of-concept HRPM to evaluate whether patient-sided factors present at the time of a hospital admission and available from Swiss claims data are promising for predicting early hospital readmissions. The potential target audience of such a model are (Swiss) health insurances and their partners. A health insurance running the model on its database could agree specific interventions for high-risk patients identified by the HRPM with the respective hospitals as part of its case management programme.

METHODS

Data source and study design

Basic health insurance is mandatory for all Swiss residents. A large number of private companies offer coverage of a broad spectrum of medical services considered appropriate, medically effective and cost-effective. Persons are free to switch the health insurance once a year. They may to some extent select the amount of their deductible and choose between different health plans.

All health plans allow reimbursement of all covered services from every approved healthcare provider, in the inpatient as well as in the outpatient setting. In return of rebates on the per-capita premiums, some health plans require a specific care provider (general practitioner or telemedicine provider) as first contact for each new health problem. Some of these first contact providers are commissioned by the health insurances to establish a binding treatment procedure and to restrict the reimbursement of other services without consulting the first contact providers again beforehand (eg, specialists visits).¹⁰

This retrospective cohort study used administrative claims data of the Helsana Group, a large health insurance with a market share of about 14% of the Swiss population (ie, app. 1.2 million persons). The claims data provide information about the use of medical services of the insurees, as well as the insurance status and socio-demographic information. The reimbursement of acute inpatient care services is regulated via the nationwide diagnosis-related groups (DRG) tariff SwissDRG.¹¹

Literature search

For identification of potential predictors of readmission, we chose a theory-driven (as opposed to a data-driven) approach. We, therefore, systematically searched for systematic reviews of HRPM and extracted the predictors used in the described models into a pool of candidate variables. Entries in the database PubMed matching the filter 'Systematic Review' and the Medical Subject Headings

(MeSH) terms 'Patient Readmission' and 'Models, Statistical' were selected. As a second search string the filter 'Systematic Review' was used again, this time together with at least one of the following words in the Title or Abstract field: readmission, rehospitalisation, model and predict. No additional restriction on the publication date was applied.

We looked at the abstracts of these publications and applied the following exclusion criteria: no systematic review, did not include the correct outcome (hospital readmissions), paediatric patients only and no acute care intervention as index hospitalisation. The selected publications were retrieved as full text, and two further exclusion criteria were applied: no patient-level characteristics as predictors and no list of predictors reported.

In addition, we retrieved studies evaluating hospitalisations in the Swiss healthcare setting using claims data of Swiss basic mandatory health insurance and extracted patient-level factors associated with hospitalisations. This step ensured the inclusion of country-specific items not considered in international studies.

The listed predictors were regrouped into seven categories: clinical assessment, comorbid conditions, illness severity, overall health and function, perioperative variables, prior use of medical services, sociodemographic variables. The candidate variables were then assessed by means of the three criteria: whether they described patient-level features, if they were available from Swiss health insurance claims data (directly or indirectly, eg, via medication use), and if they were available at admission.

Outcome

The primary outcome was all-cause readmissions to any acute care hospital within 30 days of discharge from the index acute care hospitalisation. This outcome is the most widely used in HRPM¹² so that a comparison of our model performance with previous research is feasible. As per sensitivity analyses, we additionally evaluated readmissions compensated within a single DRG case (ie, within maximum of 18 days and within the same major diagnostic category, MDC) and 60 and 90 days readmissions.

The years 2014, 2015 and 2016 were screened for hospitalisations to any acute care hospital in Switzerland. Exclusion criteria were: patients below age 19 in the year of the hospitalisation, hospitalisations due to accident or delivery, psychiatric hospitalisations, death of patients within 90 days after discharge and cases where the hospitalisation was followed by a stay in any inpatient institution other than an acute care hospital (eg, nursing home, rehabilitation clinic). If a patient was hospitalised several times within the investigated time period, we only included the first one.

Candidate variables and additional measures

All candidate variables on the prior use of health services took into account data from a 1-year period prior to the index hospitalisation. The data contain the purchasing power measured on a zip code level. This variable consists

of centred values with a mean of 100 in the Swiss population. It was used as an approximation of the individuals socioeconomic status. The DRG cost weight reflects the complexity of the hospital stay and consists of centred values with a mean of 1. We could not include this variable in the prediction model since it is only available after discharge. However, this variable allowed us to assess whether our study population differs largely from the Swiss population with respect to the complexity of the hospitalisations.

In order to account for non-linearity in the prediction model, all continuous candidate variables were grouped into categories, without stratification of the data. The age of the patients was grouped into categories of 10 years, with an open lower bound in the category up to 29, and an open upper bound in the category above 90 years. Costs variables and the number of outpatient visits were grouped into no costs/visits, above no costs/visits up to the third quartile, and costs/visits above the third quartile. Purchasing power was grouped into lower than the first quartile, between the first and the third quartile, and above the third quartile. Count variables with low counts (number of emergency visits, previous hospitalisations, comorbidities and relocations) were grouped into none/never, one/once and multiple. The length of stay (LOS) of the index admission was dichotomised into below 3 days and 3 days and above. In a sensitivity analysis, we centred, scaled and transformed (Yeo-Johnson transformation) all continuous variables instead of categorising them. All MDCs accounting for less than 2.5% of the cases were grouped in one category (other). Urbanity discriminates between rural and urban (central city, isolated city and agglomeration).

Statistical analysis

We identified 157 940 patients with an eligible index hospitalisation in the screened period. Patients were excluded if they were asylum seekers, Helsana employees or lived abroad (3876), when they were not insured with Helsana during the full observation period (1 year before and 3 months after the index hospitalisation, 12 626), and patients with medications charged as a lump sum from a nursery home (3221). The final study population, therefore, consisted of 138 222 hospitalised persons, 53.2% of them were female and the mean age was 60.5 years. No missing data had to be imputed in the resulting data set.

First, descriptive analysis of the total study population and subgroups were done. As previous HRPM are often restricted to specific populations⁷ in terms of age, location, emergency admissions or specific conditions, we explored these variables specifically from different perspectives (distribution in the population, raw readmission rates and model performance). We used the following subgroups: the largest age group, the largest geographical region, patients with an emergency admission, surgical patients, medical patients, patients with the most prevalent MDC, the MDC with the highest raw readmission rate and patients with an index hospitalisation of

3 days or more. Apart from the primary outcome (30-day readmissions), we looked at three secondary outcomes with different follow-up periods.

Bivariate analyses between the outcome and available variables were performed using Kruskal-Wallis test, Fisher's exact test, X^2 test and likelihood-ratio-test (LRT). The prediction model was a generalised linear model with a binary dependent variable (logistic regression). In a sensitivity analysis, we used a generalised mixed-effects model with random intercepts per hospital.

Characteristics of the patients and their index hospitalisation in the study population stratified by the primary outcome were calculated. We plotted a Kaplan-Meier curve of the observed readmissions to depict the fraction of patients with and without a readmission by time after discharge. Raw readmission rates of the primary outcome stratified by subgroups of the study population were calculated to show which patients are mainly affected by early readmissions.

The set of predictors in the final model was determined with a heuristic approach in four steps. First, we performed LRT with the primary outcome and each available candidate variable separately. All candidate variables resulting in a $p < 0.001$ together initialised the model in the second step. The second step iteratively removed one variable at the time based on its Akaike information criterion. We stopped this step once no further variable could be removed without worsening the model fit significantly. These first two steps are widely used in the development of HRPM.¹² Since the previous steps tend to overvalue large models with too many predictors which overfit the data.¹³ We added a manual third step. We deleted variables from the model manually when their estimated coefficients were inconsistent (eg, health plan), small with large variation (eg, home care costs), or when the effect was large with the respective characteristic being very scarce (eg, flag for prosthesis, some pharmacy-based cost group (PCG)). In a fourth step, we tested possible effect modification of variables by performing LRT of interaction terms of the selected variables and all available candidate variables. If the estimated coefficients in models split by those variables showing the highest number of significant interaction terms differed relevantly, the interaction terms were added to the model. Estimated coefficients of the prediction model are shown as ORs and their 95% CIs.

In order to assess the model discrimination, we calculated area under the curve (AUC) with k-fold cross-validation ($k=20$). This was done for the total sample, patient subgroups and the different outcomes separately. We show the calibration intercept (calibration-in-the-large) and the calibration slope of the model with the entire study population and the primary outcome as calibration measures.¹⁴ We estimated the intercept and the slope with a weighted linear regression model, with number of observations as weights.

The success of the model prediction was measured with a random split of our data into a training (75%) and a test set (25%). We used a cut point which maximised the

Table 1 Population characteristics for patients with and without readmission within 30 days after discharge

Variable	Total	With	Without	P value
n	138 222	10 300 (7.5%)	127 922 (92.5%)	
Sex (=female)	73 580 (53.2%)	4808 (46.7%)	68 772 (53.8%)	<0.001*
Age	60.53/17.7	63.25/16.9	60.31/17.7	<0.001†
Geographical region				<0.001‡
Zurich	33 861 (24.5%)	2656 (25.8%)	31 205 (24.4%)	
Lake Geneva	17 165 (12.4%)	1057 (10.3%)	16 108 (12.6%)	
Midland	25 790 (18.7%)	1895 (18.4%)	23 895 (18.7%)	
Northwest	21 009 (15.2%)	1565 (15.2%)	19 444 (15.2%)	
East	19 330 (14.0%)	1583 (15.4%)	17 747 (13.9%)	
Ticino	12 562 (9.1%)	988 (9.6%)	11 574 (9.0%)	
Central	8505 (6.2%)	556 (5.4%)	7949 (6.2%)	
Purchasing power	101.78/20.3	101.45/20.1	101.80/20.3	0.209†
Surgical/medical (=medical)	52 657 (38.1%)	5174 (50.2%)	47 483 (37.1%)	<0.001*
Length of stay	4.56/4.8	9.07/8.9	4.20/4.1	<0.001†
Major diagnostic category				<0.001‡
Musculoskeletal system	29 362 (21.2%)	1176 (11.4%)	28 186 (22.0%)	
Digestive system	22 488 (16.3%)	1396 (13.6%)	21 092 (16.5%)	
Circulatory system	17 682 (12.8%)	1588 (15.4%)	16 094 (12.6%)	
Female reproductive system	9263 (6.7%)	489 (4.7%)	8774 (6.9%)	
Kidney and urinary tract	8598 (6.2%)	1403 (13.6%)	7195 (5.6%)	
Ear, nose, mouth and throat	8102 (5.9%)	537 (5.2%)	7565 (5.9%)	
Respiratory system	7213 (5.2%)	649 (6.3%)	6564 (5.1%)	
Skin, subcutaneous tissue and breast	7118 (5.1%)	560 (5.4%)	6558 (5.1%)	
Nervous system	6989 (5.1%)	544 (5.3%)	6445 (5.0%)	
Hepatobiliary system and pancreas	5442 (3.9%)	569 (5.5%)	4873 (3.8%)	
Male reproductive system	4459 (3.2%)	346 (3.4%)	4113 (3.2%)	
Other	11 506 (8.3%)	1043 (10.1%)	10 463 (8.2%)	
Cost weight	0.99/0.7	1.15/1.0	0.97/0.7	<0.001†

*Kruskal-Wallis test.

†Fisher exact test.

‡ χ^2 test.

true positive rate and minimised the false positive rate to determine which patients may experience a readmission.

With the true and the predicted outcome, we then calculated the following measures: the positive prediction value (PPV), the negative prediction value (NPV), sensitivity, specificity and the detection prevalence (share of the patients predicted to have a rehospitalisation). This was done for the total sample with the primary outcome only.

A $p < 0.05$ in statistical tests was considered significant throughout the study if not stated otherwise. We conducted all statistical analysis in R V.3.5.0.¹⁵

Patient and public involvement

Neither patients nor the public were involved in the planning or design of the study, as we worked with secondary

data collected by a health insurer for reimbursement purposes.

RESULTS

Literature search and candidate variables assessment

The literature search resulted in 65 entries with one duplicate. We excluded 55 publications because they did not match the inclusion criteria. This resulted in nine included publications.^{7 16–23} Two publications specifically evaluating the use of Swiss claims data in regression models were additionally included.^{24 25}

The assessment of all variables collected in the identified literature is shown in the online supplementary appendix 1. The extracted candidate variables included:

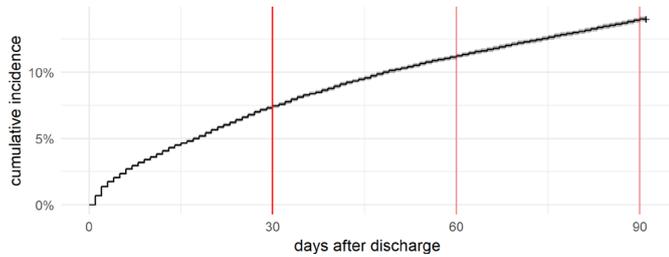


Figure 1 Reverse Kaplan–Meier curve of readmissions.

PCGs, an established measure for detection of a set of 22 comorbidities via medication use in Swiss claims data,²⁶ variables of healthcare utilisation in the year prior to the index hospitalisation (eg, grouped costs, visits and hospitalisations), and socioeconomic and sociodemographic information (eg, sex, age, regional information of residence, purchasing power). Since we wanted to build a model using only information available at admission, the following variables principally available in Swiss health insurance claims data could not be used as predictors: LOS, MDC and DRG cost weight of the index admission, and postoperative health services use.

Population

Table 1 shows characteristics of the patients and their index hospitalisations. The study population was distributed across all geographical regions of Switzerland, had an average purchasing power index, and the index

hospitalisations had an average complexity according to their cost weights.

Readmitted patients were about 3 years older and more often male, the index hospitalisations of readmitted persons were more often medical (as compared with surgical), longer and had a higher complexity as compared with those of non-readmitted persons. The MDC were distributed differently in index admissions followed by a readmission compared with those without a readmission. The most prevalent MDC in index admissions followed by a readmission was circulatory system, while in those without a readmission it was musculoskeletal system.

A complete list of bivariate analysis of all potential predictors available at admission with the primary outcome can be found in the online supplementary appendix 2.

Raw rates

Four different readmission rates were calculated: for readmissions within 30 days of discharge (7.5%, primary outcome), within a single DRG case (ie, within 18 days, 2.9%), and readmissions within 60 (11.2%) and 90 days (14.0%), respectively. Figure 1 shows the cumulative incidence of readmissions by number of days after discharge. The secondary outcome readmission within a single DRG case is not depicted in figure 1 since additional restrictions were applied (only readmissions within the same

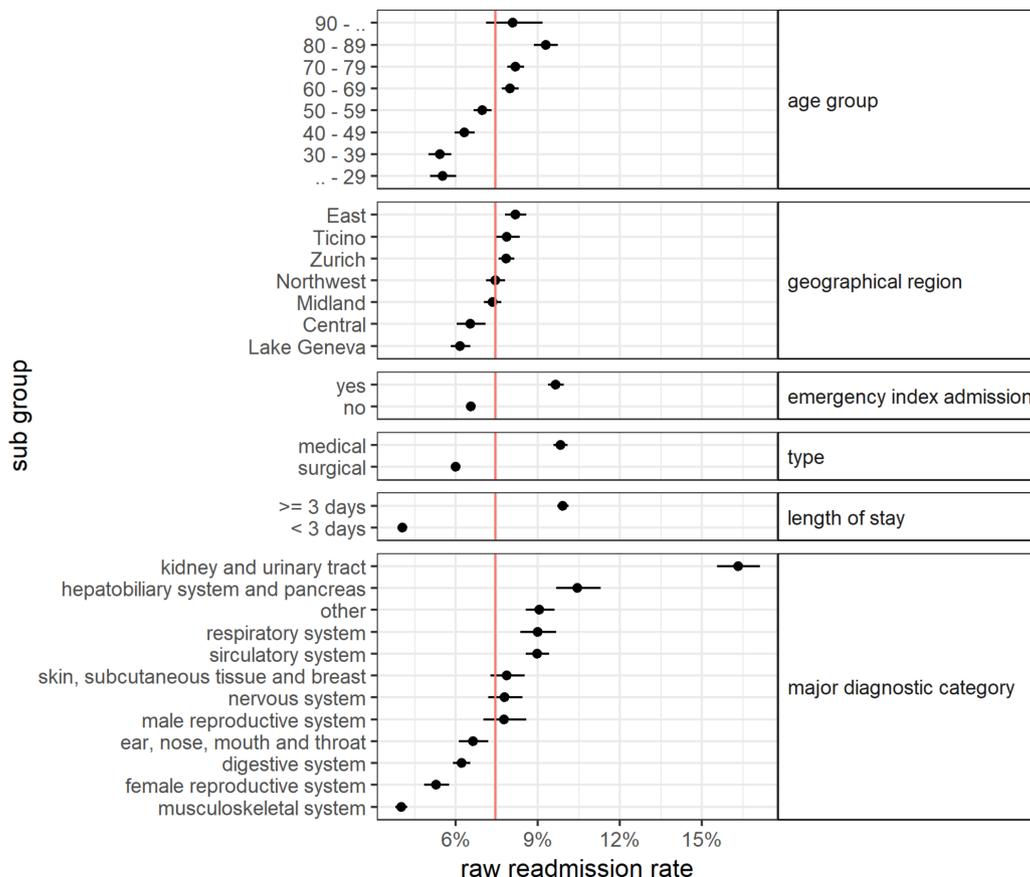


Figure 2 Raw rates of readmission split by subgroups.

MDC as the index admission). The proportion of patients with a readmission increases smoothly and decelerates with increasing number of days after discharge. Based on [figure 1](#), there is no single follow-up period which should be preferred on the others.

[Figure 2](#) shows the raw rates of the primary outcome in the selected subgroups. The red line indicates the readmission rate in the total study population. The readmission rate rose with increasing age, showing a linear correlation between the two except for the lowest and the highest age groups. Readmission rates differed between geographical regions with no apparent pattern. Readmission rates were higher if the index hospitalisation was an emergency admission, for a medical (as opposed to surgical) stay or for stays of 3 days or longer. Subgroups according to the MDC of the index hospitalisation had largely different readmission rates. Patients with the most prevalent MDC, diseases of the musculoskeletal system, had the lowest readmission rate (n=29 362, rate=4.0%). Patients with a hospitalisation because of a disease of the kidney or urinary tract (n=8598, rate=16.3%) had by far the highest readmission rate, which is more than four times the one in patients with a disease of the musculoskeletal system.

Variable selection

The first step of the variable selection process resulted in 33 variables having an LRT with a $p < 0.001$. The automated stepwise variable deletion of step two left 21 variables in the model. In the third step, we deleted seven variables manually (PCG gout, PCG intestinal inflammatory diseases, prosthesis, costs home care, costs medication outpatient, health plan, language region).

Effect modification was explored in the fourth step, but considering interaction terms did not improve the model. Therefore no interaction terms were added to the model. The resulting final model contained the following 14 variables as predictors: PCG cancer, PCG cardiac disease, PCG pain, emergency index admission, number of emergency visits, costs specialists, costs hospital outpatient, costs laboratory, costs therapeutic devices, costs physiotherapy, number of outpatient visits, sex, age group, geographical region.

Model performance

The ORs of each predictor of the final model are depicted in [table 2](#). The predictors with the largest effects were emergency index admission (OR 1.50, 95% CI 1.43 to 1.56), high hospital outpatient costs (1.50 (1.43 to 1.56)) and age group 80–89 years (1.50 (1.43 to 1.56)).

The discriminatory power of our model is limited ([table 3](#)). The AUC using the entire study population was 0.60 (95% CI 0.60 to 0.61). Fitting the model to the specific subgroups, which are likely to be more homogeneous, improved the model discrimination in two subgroups (MDC musculoskeletal system and surgical patients). The best AUC resulted in the group with the MDC musculoskeletal system, where it was 0.66 (95% CI

0.65 to 0.67). The calibration of the model was good. Calibration-in-the-large was 1.02 (95% CI -0.28 to -2.32), the calibration slope was 0.86 (95% CI 0.72 to 1.01).

The PPV using a balanced cut point was low. In the model relating to the entire study population, it made up 9.9%, which means that 9.9% of the patients predicted to have a readmission did in fact have one. NPV was 94.5%, sensitivity was 58.6% and specificity was 56.8%. The detection prevalence was 44.3%.

[Table 4](#) shows the results of the sensitivity analysis with the primary and all three secondary outcomes. The model discrimination increased very slightly with longer follow-up periods.

DISCUSSION

The present study found that a prediction model based on patient-level variables available in claims data and at the time of the index admission alone is not promising. We, therefore, conclude that more sophisticated and resource-intensive attempts comparing a wide range of modelling techniques (eg, logistic regression, random forest, support vector machines, neural networks) are probably not worthwhile.

Results seem to be stable across patient subgroups, that is, with regard to age, geographical region, emergency status or clinical information. The developed model discriminated poorly, though it was well calibrated. Sensitivity analysis using a mixed-effects model with random intercepts per hospital and using pre-processed (centred, scaled and transformed) instead of categorised continuous variables did not improve the model performance meaningfully (results not shown).

Our comparison of the subgroups showed significant differences in their prevalence in the hospitalised and readmitted population and their raw readmission rates. However, we did not find relevant improvements in the model performance if we fitted our model to one of the subgroups. Existing HRPM are often restricted to such subgroups of the general population.⁷ The results of this study do not provide a statistical reason for restricting the target population of an HRPM.

Based on clinical considerations, it is debated controversially whether 30 days readmissions are an appropriate performance indicator for hospitals.²⁷ Our data did not suggest a certain follow-up time in an HRPM, neither in the cumulative incidence rate nor in the performances the models achieved with different outcomes.

Models enabling real-time use by restricting input variables to timely available ones exist,^{28 29} as well as models relying on claims data.^{30 31} However, to our knowledge, the present study is the first combining these two aspects, developing a model with variables available at admission that were based on claims data. The Preadmission Readmission Detection Model (PREADM model)²⁸ included only variables available at admission derived from electronic health records and other administrative sources, including chronic disease registries and data on

Table 2 Estimated coefficients of the final model

Term	OR (95% CI)	P value
(Intercept)	0.05 (0.04 to 0.06)	<0.001
Pharmacy-based cost group (PCG) cancer (=yes)	1.26 (1.12 to 1.42)	<0.001
PCG cardiac disease (=yes)	1.09 (1.03 to 1.14)	0.001
PCG pain (=yes)	1.16 (1.10 to 1.23)	<0.001
Emergency index admission (=yes)	1.50 (1.43 to 1.56)	<0.001
No of emergency visits (=one)	1.06 (1.00 to 1.12)	0.054
No of emergency visits (=multiple)	1.15 (1.07 to 1.25)	<0.001
Costs specialists (=low)	0.93 (0.88 to 0.99)	0.013
Costs specialists (=high)	0.87 (0.81 to 0.94)	<0.001
Costs hospital outpatient (=low)	0.99 (0.94 to 1.04)	0.752
Costs hospital outpatient (=high)	1.28 (1.18 to 1.38)	<0.001
Costs laboratory (=low)	1.13 (1.06 to 1.21)	<0.001
Costs laboratory (=high)	1.31 (1.21 to 1.42)	<0.001
Costs therapeutic devices (=low)	1.07 (1.01 to 1.12)	0.012
Costs therapeutic devices (=high)	1.20 (1.11 to 1.30)	<0.001
Costs physiotherapy (=low)	0.88 (0.83 to 0.93)	<0.001
Costs physiotherapy (=high)	0.87 (0.79 to 0.95)	0.002
No of outpatient visits (=low)	0.81 (0.73 to 0.89)	<0.001
No of outpatient visits (=high)	0.84 (0.74 to 0.94)	0.003
Sex (=male)	1.30 (1.24 to 1.35)	<0.001
Age group (=30–39)	0.98 (0.87 to 1.11)	0.786
Age group (=40–49)	1.16 (1.04 to 1.30)	0.008
Age group (=50–59)	1.25 (1.12 to 1.39)	<0.001
Age group (=60–69)	1.40 (1.26 to 1.56)	<0.001
Age group (=70–79)	1.38 (1.24 to 1.53)	<0.001
Age group (=80–89)	1.50 (1.34 to 1.68)	<0.001
Age group (=90+)	1.26 (1.06 to 1.50)	0.009
Region (=Lake Geneva)	0.75 (0.69 to 0.81)	<0.001
Region (=Midland)	0.92 (0.87 to 0.98)	0.011
Region (=Northwest)	0.94 (0.88 to 1.01)	0.079
Region (=East)	1.05 (0.98 to 1.12)	0.187
Region (=Ticino)	0.96 (0.88 to 1.03)	0.243
Region (=Central)	0.82 (0.74 to 0.90)	<0.001

behavioural risk indicators. It achieved an AUC of 0.69 in the test set, and depending on the chosen cut point the PPV ranged from 17.6% to 34.3%. According to the authors, the model performance of a readmission prediction model is likely to be low, since only a fraction of all readmissions are deemed avoidable (ie, the readmissions might occur due to a random unpreventable health problem). The fact that a model using only claims data can be improved by adding additional information (eg, clinical, psychosocial) was shown in a recent study in the USA.³⁰ However, the reported improvement between the claims-only (AUC=0.63) and the more elaborate model (AUC=0.65) was rather small. Similar findings^{7 31} suggest that a claims-only model benefits little from adding clinical

data, and that readmission risk prediction remains poorly understood.

Our model performance lined up with existing approaches and stayed as such below a practical benefit. To be useful, the model should identify only patients having a high risk for a rehospitalisation. Our model predicts 44.3% of the patients to have a rehospitalisation (detection prevalence), but only 9.9% of them had in fact one (PPV). This marginal improvement of our prediction model over an uninformed model (with 50% detection prevalence and the actual rehospitalisation rate of 7.5% as PPV) does not justify the costs of an implementation of the model into production.³²

Table 3 Model discrimination (area under curve, AUC) in the entire population and in subgroups

Population	AUC	95% CI
Entire study population	0.60	0.60 to 0.61
Age group 70–79	0.59	0.58 to 0.60
Region Zurich	0.60	0.59 to 0.61
Emergency index admission	0.57	0.56 to 0.58
Surgical patients	0.62	0.61 to 0.63
Medical patients	0.58	0.57 to 0.59
Musculoskeletal system	0.66	0.65 to 0.67
Kidney and urinary tract	0.60	0.57 to 0.62
Length of stay ≥ 3 days	0.58	0.57 to 0.59

Nevertheless, the study examined important aspects in the understanding of hospital readmissions. Contrary to existing suggestions,⁷ we did not find meaningful interactions between variables. Patient's age was available in most existing models, but was not considered in most final models due to their small contribution. We found a strong association between age and the readmission risk. This may be explained by the fact that we included a broad range of age groups while other models were solely based on an elderly population.

Patients with an emergency index hospitalisation consistently had an elevated readmission risk across all subgroups (OR 1.50, 95% CI 1.43 to 1.56) in the general population). An emergency visit in the year prior to the index admission was also an important predictor for readmissions, and patients without any outpatient visit had an elevated readmission risk as well. Emergency admissions and having no outpatient visits might be indicators of a delay of seeking care. Since these three variables are associated with an elevated readmission rate, it should be investigated in more depth if delay of healthcare is a relevant risk factor for hospital readmissions. In addition, the readmission rates of patients of the different geographical regions of Switzerland varied widely, even when we controlled for confounding in the regression model. This aspect should be examined in more detail in future research as well.

A main strength of the study is the fact that the data we used are characterised by a high availability and reliability and included the general population as well as all healthcare providers of the whole country. We were able to observe a large number of hospitalisations (n=138 222)

Table 4 Model discrimination (area under curve, AUC) with primary and secondary outcomes

Population	AUC	95% CI
Within 18 days of discharge	0.60	0.59 to 0.61
Within 30 days of discharge	0.60	0.60 to 0.61
Within 60 days of discharge	0.61	0.61 to 0.62
Within 90 days of discharge	0.62	0.62 to 0.62

and covered a large share of inhabitants of Switzerland. The measured readmission rates in the study population were in line with official statistics,³³ reporting 8.7% and 12.3% for the 30 days and 60 days readmission rates for the year 2013.

Several limitations of our study need to be considered. Although the literature search was done systematically, we may have missed important articles and therefore relevant candidate variables which could have been extracted from our data. Second, Swiss claims data were our only data source and our results may thus not be transferable to other countries. Further, each patient was included into the study population only once, since we only used the first hospital admission per person as index hospitalisation.

The aim of the present study was to evaluate the potential of existing evidence (ie, previously published literature) and existing data sources (ie, health insurance claims data) for a promising risk prediction tool. Given that the present study was planned as a proof-of-concept we took some simplifying approaches in the model building process. The categorisation of all continuous variables based on their distribution in the present data helped to allow for non-linear relationships with the outcome, but led to some information loss in these variables. Although, in a sensitivity analysis with centred, scaled and transformed (Yeo-Johnson transformation) continuous variables, we did not find a meaningful improvement of the model discrimination (see online supplementary appendix 3). Furthermore, we used logistic regression, which is the most widely used analysis method in existing HRPM. However, other methods (eg, random forest) could achieve better discriminatory power than logistic regression given the same data.¹²

In addition, we used a theory-based approach for identification of candidate variables. Alternatively, a data mining approach might derive different candidate variables from the available data, which might result in a better model performance. However, acceptance by healthcare professionals is an important prerequisite for an effective implementation of a risk prediction tool in clinical practice. We are convinced that the acceptance is higher when predictors are derived from medical concepts and well established as opposed to technically derived items.

We derive several implications from the present pilot study for further research. First, following the path of using patient-sided factors currently available in Switzerland at the time of hospitalisation alone is not worthwhile for the development of a clinically relevant risk prediction tool. However, and second, health insurance claims data at the patient level seem to be very useful to investigate rehospitalisations and to identify potential predictors. They might provide additional helpful information supplementary to current reporting on the hospital level.⁹ Third, given the fact that patient-sided factors alone seem to have limited predictive power, we hypothesise that a hierarchical model that includes data on the

hospital level could achieve better results. In-hospital process factors and supply of health services, especially of hospital beds, could add important information to explain the differences in regional readmission rates,^{7 34} and thus increase the discriminatory power of the model. Moreover, as readmission rates vary widely between countries,³⁵ it may also be helpful to examine the incentives in place in different healthcare systems.

CONCLUSION

Opposed to our assumption, this attempt of a prediction model for rehospitalisations on patient-sided variables alone derived from Swiss health insurance claims data available at the time of a hospitalisation did not lead to a practically useful risk discrimination. This pilot study may be helpful for the design of future studies and for the allocation of resources for research and quality assurance projects. Based on these results, we recommend that future studies focus on surrogates for patient's health-seeking behaviour, in-hospital process factors, supply parameters or incentives on the healthcare system level. The results are specifically relevant for the ongoing political debate in Switzerland about how to use and implement healthcare information in order to increase the efficiency of the healthcare system. However, since the study is based on an international literature review, the results may also be helpful for researchers from other healthcare systems. Given the fact that a part of the occurring rehospitalisations is preventable, future research on how to identify the population at risk is of high relevance for patients and all involved stakeholders.

Acknowledgements The authors thank Andri Signorell for his helpful support in analysing the data.

Contributors Both authors participated in the elaboration of the concept and design of the study. BB performed the statistical analysis. Both authors contributed to data interpretation. BB drafted the manuscript, EB reviewed and commented. Both authors read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests Both authors were employed at the Helsana Group at the time of writing.

Patient consent for publication Not required.

Ethics approval Prior to the analysis, the pre-existing raw data were anonymised in order to protect the privacy of patients and care providers. Hence, in accordance with the Swiss Federal Law on data protection, neither the consent of the insureds nor an ethics committee approval was required for this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med* 2009;360:1418–28.
- Halfon P, Eggli Y, van Melle G, et al. Measuring potentially avoidable hospital readmissions. *J Clin Epidemiol* 2002;55:573–87.
- Baillie CA, VanZandbergen C, Tait G, et al. The readmission risk flag: using the electronic health record to automatically identify patients at risk for 30-day readmission. *J Hosp Med* 2013;8:689–95.
- Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. *JAMA Intern Med* 2014;174:1095.
- Hansen LO, Young RS, Hinami K, et al. Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Intern Med* 2011;155:520–8.
- Djalali S, Markun S, Rosemann T. Routinedaten – das ungenutzte Potenzial in der Versorgungsforschung. *Praxis* 2017;106:365–72.
- Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;306:1688–98.
- Riley GF. Administrative and claims records as sources of health care cost data. *Med Care* 2009;47:S51–5.
- Burla L, Diebold M, Dutoit L, et al. *Indikatoren Gesundheit2020: Aktualisierung*, 2018.
- Reich O, Rapold R, Flatscher-Thöni M. An empirical investigation of the efficiency effects of integrated care models in Switzerland. *Int J Integr Care* 2012;12:e2.
- SwissDRG. <https://www.swissdrg.org/de/akutsomatik/swissdrg> (03 Dec 2018).
- Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: A systematic review of methods. *Comput Methods Programs Biomed* 2018;164:49–64.
- Kuhn M, Johnson K. *Applied predictive modeling*. New York, NY: Springer New York, 2013.
- Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
- Core Team R. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018.
- Damle RN, Alavi K. Risk factors for 30-d readmission after colorectal surgery: a systematic review. *J Surg Res* 2016;200:200–7.
- Desai MM, Stauffer BD, Feringa HH, et al. Statistical models and patient predictors of readmission for acute myocardial infarction: a systematic review. *Circ Cardiovasc Qual Outcomes* 2009;2:500–7.
- Fisher AV, Fernandes-Taylor S, Campbell-Flohr SA, et al. 30-day readmission after pancreatic resection: a systematic review of the literature and meta-analysis. *Ann Surg* 2017;266:242–50.
- García-Pérez L, Linertová R, Lorenzo-Riera A, et al. Risk factors for hospital readmissions in elderly patients: a systematic review. *QJM* 2011;104:639–51.
- Lichtman JH, Leifheit-Limson EC, Jones SB, et al. Predictors of hospital readmission after stroke: a systematic review. *Stroke* 2010;41:2525–33.
- Ross JS, Mulvey GK, Stauffer B, et al. Statistical models and patient predictors of readmission for heart failure: a systematic review. *Arch Intern Med* 2008;168:1371–86.
- Swain MJ, Kharrazi H. Feasibility of 30-day hospital readmission prediction modeling based on health information exchange data. *Int J Med Inform* 2015;84:1048–56.
- Zhou H, Della PR, Roberts P, et al. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. *BMJ Open* 2016;6:e011060.
- Huber CA, Schneeweiss S, Signorell A, et al. Improved prediction of medical expenditures and health care utilization using an updated chronic disease score and claims data. *J Clin Epidemiol* 2013;66:1118–27.
- Schoepfer A, Vavricka SR, Brüngger B, et al. Systematic analysis of annual health resource utilization and costs in hospitalized patients with inflammatory bowel disease in Switzerland. *Eur J Gastroenterol Hepatol* 2018;30:868–75.
- Huber CA, Szucs TD, Rapold R, et al. Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. *BMC Public Health* 2013;13:1030.
- Joynt KE, Jha AK. Thirty-day readmissions—truth and consequences. *N Engl J Med* 2012;366:1366–9.
- Shadmi E, Flaks-Manov N, Hoshen M, et al. Predicting 30-day readmissions with preadmission electronic health record data. *Med Care* 2015;53:283–9.
- Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Med Care* 2010;48:981–8.



30. McManus DD, Saczynski JS, Lessard D, *et al.* Reliability of predicting early hospital readmission after discharge for an acute coronary syndrome using claims-based data. *Am J Cardiol* 2016;117:501–7.
31. Hammill BG, Curtis LH, Fonarow GC, *et al.* Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circ Cardiovasc Qual Outcomes* 2011;4:60–7.
32. In: Stoto MA, Almario DA, McCormick MC, eds. *Reducing the odds: Preventing perinatal transmission of hiv in the united states*. Washington (DC), 1999.
33. Widmer M, Kohler D. *Behandlungsketten und ihre schnittstellen*, 2015.
34. Oddone EZ, Weinberger M, Horner M, *et al.* Classifying general medicine readmissions. Are they preventable? Veterans Affairs Cooperative Studies in Health Services Group on Primary Care and Hospital Readmissions. *J Gen Intern Med* 1996;11:597–607.
35. Bottle A, Middleton S, Kalkman CJ, *et al.* Global comparators project: international comparison of hospital outcomes using administrative data. *Health Serv Res* 2013;48:2081–100.