Methodological paper

Hospital Readmission Ratio

Methodological report of 2015 model

2017

Jan van der Laan
Corine Penning
Agnes de Bruin

Index

1. Introduction

1.1 Indicators of quality of hospital care

Overall quality of hospital care can be estimated using several types of quality indicators based on hospital admission data. Such indicators for identifying potentially suboptimal quality of hospital care might focus for example on unexpected in-hospital or post-discharge mortality, potentially preventable hospital readmissions or unexpected long duration of admissions. In the Netherlands, hospital admission and discharge data is registered in the LBZ, a national hospital discharge register covering all general, university and three specialised hospitals. Other specialised clinics, independent treatment centres and private clinics are not included. Inpatients as well as day cases and prolonged observations without overnight stay are registered. For each hospital discharge administrative data of the admission are registered, as well as diagnoses and procedures.

In the Netherlands, hospitals participating in the LBZ registration are annually provided by Dutch Hospital Data with a set of indicators based on their performance in the previous year. Since 2014 this set includes the (unadjusted) hospital readmission rate, which is the ratio of the number of observed readmissions to the total number of hospital admissions. However, since this ratio does not correct for case mix differences, it might be less indicative of differences in the true number of potentially preventable readmissions. Therefore, Dutch Hospital Data has asked Statistics Netherlands to develop a model to estimate the expected readmission risks adjusted for relevant covariates, in a fashion similar to the estimation of the hospital standardized mortality rates (HSMR).

1.2 Predictive value of other hospital readmission models

Internationally, models for estimating hospital readmission rates are used for the purpose of risk stratification but also as a quality indicator. From previous studies it is known that several different patient characteristics can contribute to the risk to be readmitted to the hospital. In a recent systematic review by Kansagara *et al.* (2011), an overview is presented of the various validated models that have been used internationally, the covariates included in those models and their overall predictive value. Common covariates include comorbidity indexes, age, sex and/or prior use of medical services (hospitalizations). Regardless of the number of included covariates, the results of only a small fraction of the models are moderately discriminative (AUC/C-statistic>0.70).

1.3 Output

Statistics Netherlands has only calculated the model for the hospital readmission risks, not the outcomes for the individual hospitals. Statistics Netherlands has calculated the model on the basis of 2015 data. Dutch Hospital Data uses this model to estimate the expected readmission risk, adjusted for relevant covariates, for each individual primary (index) hospital admission in 2016. For each hospital the standardized (adjusted) readmission ratio can then be calculated as the observed number of readmissions (x 100) divided by the sum of the expected readmission risks of the index admissions of that hospital.

2. Methods

The (hospital) readmission ratio is calculated using the expected (hospital) readmission risk as the denominator and observed readmission as the numerator. The expected readmission risk within the same hospital is predicted for each individual admission within a given period, adjusted for patient and admission characteristics of that admission as covariates. Readmission risk was predicted for all (index) admissions that potentially could be followed by a readmission, excluding admissions for diagnoses with complex care paths where planned readmissions are often involved. *Readmissions are defined as those admissions that occurred within 30 days of the discharge date of the preceding index admission*. Detailed information on the characteristics and criteria of index and readmissions is given in paragraphs 2.1.4 and 2.1.5.

Expected readmission risk is determined for each of the included diagnosis groups, which are based on the CCS *(Clinical Classifications Software*[1]), which clusters ICD codes of the main diagnoses of the admissions into 259 clinically meaningful categories. In accordance with the HSMR, we further clustered these groups into 157 diagnosis groups, which are partly the same clusters used for the SHMI (Summary Hospital-level Mortality Indicator) in the UK (HSCIC, 2016). To determine readmission risk we used logistic regression models, with an observed readmission as the target (dependent) variable and various variables available in the LBZ as covariates.

The methodology for estimating the expected readmission risk is very similar to that used for estimating expected mortality rates applied for calculating the HSMR rates for 2015, described in detail elsewhere (Van der Laan *et al.* 2016). In the following section we therefore briefly describe the applied methods, while deviations from the HSMR methodology or other methods specific to the current project are described in more detail.

2.1 Target population and data set

2.1.1 Admissions – general criteria

We consider both the population of hospitals and the population of admissions. Our population of (re)admissions consists of "all hospital stays (inpatient admissions) of Dutch residents in Dutch short-stay hospitals within the study period". Only completely registered admissions with a registered main diagnosis were included. In the LBZ, the date of discharge, and not the day of admission, determines the year a record is assigned to. Therefore, the registered hospital stays of year *t* comprise all inpatient admissions that ended in year *t*. Day admissions and prolonged observations were excluded, since subsequent readmissions might be elective, for example, for prolonged treatment.
Lastly, admissions of foreigners were excluded from the model, since readmissions might have also taken place in a hospital in their residential country. The number of admissions of foreigners is relatively small.

[1] *See http://www.hcup-us.ahrq.gov/toolssoftware/icd_10/ccs_icd_10.jsp*

2.1.2 Study period

To calculate the predicted readmission risks for 2015 we selected a population of index admissions with discharge dates in a study period of 12 consecutive months. We selected this 12-month period in such a way that (nearly) all readmissions – that by definition occur within 30 days of the discharge date of the index admissions – end in the year 2015. This is necessary as the LBZ is organised according to year of discharge, and 2015 was the latest available LBZ dataset. After investigation (see section 4.1) it was decided to select all index admissions with a discharge date from November 1st 2014 up to October 31st 2015 (study period) for the readmission model of 2015.

Hospitals with partially incomplete data within the study period were only included for the months in which data was completely registered, taking into account that index admissions could be identified up to two months before the final month of completely registered data, so that readmissions after the last month of discharge of the index admissions could be sufficiently identified.

2.1.3 Hospitals

"Hospital" is the primary observation unit. Hospitals report admission data (hospital stay data) in the LBZ. However, not all hospitals participate in the LBZ. In principle, the hospital readmission risk model includes all general hospitals, all university hospitals and short-stay specialised hospitals with inpatient admissions participating in the LBZ in the study period. The readmission ratio is calculated using LBZ data on admissions within the same hospital, using the hospital-specific patient identification number as the unique key for identifying (re)admissions. In case of merging locations, the merged hospital was considered as a single unit during the entire study period.

2.1.4 Additional criteria for index admissions

Expected readmission risk was only calculated for those inpatient admissions for which readmission was possible, and excluding some specific diagnosis groups (index admissions). Thus, index admissions had to meet the following criteria:

- The patient did not die during the admission
- The main diagnosis of the admission was not related to oncology (CCS groups 11-45) or psychiatry (CCS groups 65-75) since hospital care for these diagnoses is usually complex and follow-up care might be required. In addition the main diagnosis was not related to obstetrics (CCS groups 176-196; 218), since most deliveries do not take place during inpatient admissions, so it cannot be determined whether an admission for this purpose is the "true" index admission.
- The date of discharge was from November 1st of year *t*-1 up to October 31st of year *t*.

2.1.5 Additional criteria for potential readmissions

Inpatient admissions only qualified as potential readmissions if the following criteria were matched:

- The main diagnosis of the admission was not related to oncology (CCS groups 11-45) or psychiatry (CCS groups 65-75) since hospital care for these diagnoses is usually complex and follow-up care might be required. In addition the main diagnosis was not related to obstetrics (CCS groups 176-196; 218), since most deliveries do not take place during inpatient admissions, so it cannot be determined whether an admission for this purpose is a "true" readmission.
- The main diagnosis of the admission was not related to social, socio-economic or psychosocial circumstances or administrative purposes (ICD10: Z55-Z65), other

circumstances (ICD10: Z70-Z76) or screening, follow-up care or rehabilitation (CCS groups 254-258), since admissions for these purposes are usually planned.
- The discharge date of the admission was before or on December 31st of year *t*.
- The maximal time lapse between the admission date of the readmission and the discharge date of the index admission is 30 days (29 days at maximum). For example, when an index admission has a discharge date of January 1st 2015, a subsequent admission on January 30th 2015 is classified as a readmission, while a subsequent admission on January 31st is not.
- If the readmission started on the same day as the discharge date of the index admission, the minimal time lapse between both admissions is one hour. If the hour of discharge of the index admission or the hour of admission of the readmission is unknown in this specific situation, the subsequent admission is not identified as a readmission.

Note that the main diagnosis of the readmission does not have to be related to the main diagnosis of the index admission. Furthermore, only readmissions within the same hospital are considered.

2.2 Target variable

The target variable for the regression analysis is the occurrence of a readmission (in the same hospital) within 30 days of the discharge date of the preceding index admission.

We used the hospital-specific patient identification number (patient ID), registered in the LBZ, as the unique key for identifying admissions in the same patient in a single hospital. The validity of this patient ID for this purpose was investigated for each hospital. The methods and results of this analysis are presented in section 4.2. There were a few hospitals with patient ID registration issues, but these hospitals were not excluded from the final regression analysis as the effects on the outcomes were small.

The dataset is composed based on the criteria presented in section 2.1. According to the additional criteria for index admissions and readmissions, two variables are added to the dataset to mark both types of admissions. Readmissions can also count as index admissions in case they are followed by another readmission. Unique patients are identified through the hospital-specific patient identification number and within the selection of admissions per patient, for each index admission the presence of a readmission within 30 days is determined. Each index admission can only be followed by a single subsequent readmission, and a single readmission can also be only allocated to a single index admission. If an index admission is followed by multiple potential readmissions within 30 days, only the first occurring readmission is marked as such. Based on this algorithm for each index admission the presence of a readmission is marked. Subsequently, all index admissions and the corresponding covariates are selected, plus the target variable (whether the primary admission was followed by a readmission or not) and these were entered into the model.

2.3 Stratification

Instead of performing one logistic regression for all admissions, we performed a separate logistic regression for each main diagnosis group. These sub-populations of index admissions are more homogeneous than the entire population. Hence, this stratification may improve the precision of the estimated readmission probabilities. As a result of the stratification, covariates are allowed to have different regression coefficients across diagnosis groups. Due to the

exclusions of specific CCS groups for the index admissions, 35 of the 157 diagnosis groups (as used for the HSMR) are fully excluded. Therefore, the model included 122 separate logistic regressions, one for each diagnosis group selected (see Appendix II for the diagnosis groups included).

2.4 Covariates (explanatory variables or predictors of readmission risk)

By including covariates of patient and admission characteristics of the index admissions in the model, the hospital readmission risk is adjusted for these characteristics. For this purpose we selected the same covariates that are also regularly used in the (H)SMR model estimations, which are variables (available in the LBZ) known to be associated with in-hospital mortality.

The LBZ variables that are included in the model as covariates are age, sex, socio-economic status, severity of main diagnosis (based on mortality risk categories), urgency of admission, Charlson comorbidities, source of admission and month of admission. These variables are described below. Detailed information on these variables and their content is available in the HSMR methodology report (Van der Laan *et al.* 2016). For the variables socio-economic status, severity of main diagnosis and source of admission the detailed classifications are presented in the file 'Classification of variables', published together with the methodology report of the HSMR (Van der Laan *et al.* 2016).

For the regressions, all categorical covariates are transformed into dummy variables (indicator variables), having scores 0 and 1. A patient scores 1 on a dummy variable if he/she belongs to the corresponding category, and 0 otherwise. As the dummy variables for a covariate are linearly dependent, one dummy variable is left out for each categorical covariate. The corresponding category is the so-called reference category. We used the first category of each covariate as the reference category.

Covariates:

- **Age at admission** (in years): *0, 1-4, 5-9, 10-14, …, 90-94, 95+*.
- **Sex** of the patient: *male, female.*
- **SES (socio-economic status)** of the postal area of patient's address: *lowest, below average, average, above average, highest, unknown.*
- **Severity of main diagnosis** groups: *[0-0.01), [0.01-0.02), [0.02-0.05), [0.05-0.1), [0.1-0.2), [0.2-0.3), [0.3-0.4), [0.4-1], Other*.
- **Urgency** of the admission: *elective, acute.*
- **Comorbidity_1 – Comorbidity_17**. All these 17 covariates are dummy variables, having categories: *0* (*no*) and *1* (*yes*).
- **Source of admission**: *home, nursing home or other institution, hospital.*
- **Month of admission.** Six 2-month periods: *January/February, …, November/December.*

2.5 Estimation of the model

Logistic regression models were estimated for each of the 122 diagnosis groups using the variables of the index admissions mentioned in the previous paragraph and the dichotomous variable indicating whether an admission was followed by a readmission as the target variable. Computations were performed using the glm function in R (R Core Team, 2015). After investigation (see chapter 4) it was decided to use the same methodology for estimating the models as for the HSMR (Van der Laan, 2016). Categories, including the reference category, are

collapsed if the number of index admissions is smaller than 50 or when there are no readmissions in the category. For more information on this see the aforementioned methodology report for the HSMR.

3. Description of the final model

3.1 Dataset

Table 3.1.1 shows the number of hospitals that were included in the readmission model. All general and university hospitals could be included. Two short stay specialised hospitals were excluded: one hospital had not participated in the LBZ while the other had registered only six months of complete data in 2015.

3.1.1 Number of hospitals in the readmission model (2015).

	General hospitals[a)]	University hospitals	Selected specialised hospitals[b)]	Total hospitals
Total number	75	8	3	86
Used in model	75	8	2	85

a) Excluding military hospital
b) One clinic for lung diseases, one cancer hospital and one eye hospital

The number of index admissions included in the readmission model, the total number of identified readmissions and the unadjusted readmission rate are listed in Table 3.1.2.

3.1.2 Admissions in readmission model (2015).

Total number of index admissions included in model	1.329.158
Number of identified readmissions	128.769
Unadjusted readmission rate	9.7%

3.2 Impact of the covariates on readmission rate

Appendix I shows which covariates have a statistically significant (95 percent confidence) impact on readmission rate for each of the 122 regression models (one for each diagnosis group). Tables 3.2.1 and 3.2.2 show the total number of significant covariates and the total Wald statistics for the 122 regression models. The tables are sorted in descending order (most important variables at the top). The first table shows in how many diagnosis groups a variable is significant in the model. The effect of variables on the predicted probabilities, and, therefore, the importance of the variables for the case mix correction performed by the models, is better measured with the Wald-statistics (shown in the second table).

For the readmission model, sex is more important than for the HSMR 2015 model (see Van der Laan *et al. 2016)*, while for the HSMR model the Charlson comorbidities 9 and 17 (liver disease/severe liver disease) and 16 (metastatic cancer) are more important.

3.2.1 Statistical significance of the covariates for the 122 logistic regressions (summary), readmission model 2015.

Covariate	No. of significant results	Covariate	No. of significant results
Age	95	Charlson_11	28
Urgency	72	Source of admission	20
Severity	65	Charlson_5	18
Charlson_13	59	Charlson_4	14
Sex	50	Charlson_7	14
Charlson_6	47	Month of admission	13
Charlson_3	46	Charlson_16	11
Charlson_14	43	SES	11
Charlson_2	39	Charlson_12	3
Charlson_10	38	Charlson_17	2
Charlson_1	35	Charlson_15	1
Charlson_9	30	Charlson_8	0

3.2.2 Wald chi-square statistics for the 122 logistic regressions, readmission model 2015.

Covariate	Sum of Wald statistics	Sum of df	Covariate	Sum of Wald statistics	Sum of df
Age	9620	1856	Charlson_1	539	104
Urgency	5426	121	Charlson_9	398	76
Severity	2896	285	Charlson_4	394	75
Sex	1204	120	Charlson_11	388	72
Charlson_13	952	102	Charlson_10	361	116
Source of admission	798	162	Charlson_5	210	72
Month of admission	788	604	Charlson_7	188	76
Charlson_14	686	98	Charlson_16	145	80
Charlson_6	642	112	Charlson_12	46	37
SES	641	522	Charlson_17	29	9
Charlson_3	616	95	Charlson_15	5	1
Charlson_2	566	88	Charlson_8	3	5

3.3 Model evaluation for the 122 regression analyses

Appendix II shows the Areas Under the Curve (AUCs; also known as C-statistics) for each of the 122 regression models. Two types of AUCs are presented: the regular AUCs (determined using the same method as with the HSMR) and the AUCs determined using cross-validation (see section 4.3). As discussed in section 4.3, the AUCs determined using cross-validation are considered to be more accurate measures for the predictive power of the models. From these AUCs it can be concluded that most models have weak predictive power. Of the 122 diagnosis groups, there are only 6 with an AUC of 0.70 or above:

- Other upper respiratory disease (diagnosis nr. 89): AUC = 0.73
- Other connective tissue disease (nr.126): AUC = 0.70

- Joint disorders and dislocations; trauma-related; sprains and strains (nr. 132): AUC = 0.70
- Fracture of upper limb (nr.135): AUC = 0.70
- Intracranial injury (nr.138): AUC = 0.70
- Superficial injury; contusion (nr. 144): AUC = 0.70

Although the predictive power of the models is low, the case mix correction performed by the models does remove some of the differences between the hospitals caused by population differences. However, because of the poor fit of the models, it is possible that there are still population differences remaining for which the models do not correct.

Further research was done into the validity of the current method for the readmission model. The results of this evaluation are presented in chapter 4.

3.4 Regression coefficients

The file "coefficients readmission index 2015.xslx" contains the estimated regression coefficients (columns 'Estimate') for each of the 122 logistic regressions as well as their standard errors (columns 'Std. Err.'). For the sake of clarity, the reference categories are given in the first row of the corresponding covariates, and by definition have zero coefficient for each regression. In many cases categories are collapsed (see section 2.5). This results in equal coefficients for the collapsed categories. If all categories were collapsed into one category for a certain variable and for a certain diagnosis group (i.e. if there was only one category with ≥50 admissions and ≥1 readmission), the variable was dropped from the model and all associated coefficients were set to zero. The significance of each of the coefficients is shown in Appendix I.

3.5 Limitations

The readmission indicator has largely the same limitations as the HSMR. Below we will address some issues that are specific to the readmission indicator.

- In principle all readmissions are included in the model: planned and unplanned; related and not related to the index admission. Ideally only unplanned readmissions should be included. However, these are not registered as such in the LBZ. The LBZ contains the variable urgency (acute versus not acute). An admission is registered 'acute' if care is needed within 24 hours and therefore does not seem to reflect the difference between planned and unplanned readmissions. To avoid the inclusion of planned readmissions, some diagnosis groups where planned readmissions are likely (for example the various groups concerning cancer) are excluded as index and readmissions. Also diagnoses that are likely planned readmissions (for example follow-up care and rehabilitation) are excluded as potential readmissions. However, there will still be planned readmissions remaining in the dataset.
- Unlike with the HSMR, Statistics Netherlands does not provide readmission ratios for 2015, based on the model of 2015. Dutch Hospital Data (DHD) will use the estimated models to calculate the ratios using hospital data from 2016. This means that the models are applied to a different year than that on which they were estimated. As is shown in section 4.4 this results in a bias and extra variance. Fortunately, the bias can be estimated and the overall average of the ratio can be presented to the hospitals.
- It is difficult to predict readmissions using the variables present in the models: the models explain only a very small part of the observed variation (see sections 4.3 and 4.4). This makes it more likely that there are unobserved population differences that

are not corrected for, that influence the readmission probability. This means that some of the differences in the current readmission ratio can be caused by unobserved population differences.

- Readmissions are currently only observed when they occur within the same hospital as identification is based on the hospital-specific patient identification numbers. Readmissions that occur in another hospital are therefore not identified. As a result, for hospitals where patients are often readmitted in another hospital, the indicator could underestimate the readmission ratio and vice versa. In the future, readmissions in other hospitals can possibly also be taken into account by linking data on national identification numbers.

4. Additional investigations for the model of 2015

This is the first year that the models for the readmission ratio have been estimated. Although in principle the same method is used as for the HSMR, the situation is not completely the same. For example, the amount of data used for estimating the models is much smaller: for the HSMR four years of data are used while for the readmission ratio only one year of data is used. Therefore, it is necessary to evaluate whether the methods used for the HSMR can also be used for the hospital readmissions ratio. Some of the potential issues that have been identified in the preparation phase or during modelling have been investigated. This includes selecting the most optimal study period, evaluating the influence of administrative errors, whether or not 'severity' of the main diagnosis should be included in the model, model selection and validation and the influence of case mix correction. The results are described in this chapter. Furthermore, sections 4.4 and 4.5 also look at the effect and quality of the case mix correction.

4.1 Selection of period

As was indicated in section 2.1, for the readmission model we selected all index admissions during a study period of 12 consecutive months. As the LBZ datasets are registered according to year of discharge, we investigated what period of index admissions should be taken in order to capture all readmissions with a discharge date before 31st December of year t.

For each index admission the first readmission that starts within 30 days after the index admission has ended, is linked to the readmission, if such a readmission can be found. Therefore we also need data on potential readmissions for a certain period after the study period of the index admissions. The latest LBZ dataset available for this study was the LBZ 2015, including all admissions ending in 2015. Therefore admissions that have started in 2015 but ended in 2016 are not present in this data set. In order to avoid the issue that readmissions belonging to an index admission end in 2016, the study period of the index admissions needs to end sometime before the end of 2015. We investigated this using LBZ data from 2014 and 2015 to see how many readmissions end in 2015 given a certain period of index admissions in 2014.

Table 4.1.1 shows the percentage of readmissions that occur in 2015 for different selection periods for the index admissions. It shows that for index admissions with a discharge date from October 15th up to October 31st 2014 less than 1% of the readmissions occur in 2015. This means that for selecting a most recent 12-month study period of index admissions, it is feasible to take index admissions with discharge dates from November 1st of year t-1 up to October 31st of year t, as then only a negligible number of readmissions will be missed because they are not present in the dataset of year t.

The table shows the overall fractions. However, it is possible that for certain diagnosis groups or hospitals these fractions are higher (e.g. because the lengths of stay are longer). Therefore, these fractions were also calculated for each diagnosis group and hospital. However, no diagnosis groups and hospitals were found where the fractions were significantly higher.

Therefore, for selecting the index admissions for the readmission model 2015, we decided to use a study period from November 1st 2014 up to October 31st 2015.

4.1.1 The number and percentage of readmissions ending in 2015 for different selection periods for index admissions in 2014.

Period of discharge for index admissions (2014)	Number of index admissions	Number of readmissions	Number of readmissions ending in 2015	Percentage of readmissions ending in 2015
1 jan – 31 may	703 764	61 320	0	0%
1 jun -14 jun	62 656	5 625	0	0%
15 jun – 30 jun	71 430	6 018	2	0%
1 jul - 14 jul	63 988	5 374	1	0%
15 jul – 31 jul	73 781	6 467	2	0%
1 aug - 14 aug	55 579	4 933	2	0%
15 aug – 31 aug	69 714	5 805	2	0%
1 sep - 13 sep	63 274	5 370	1	0%
15 sep – 30 sep	72 828	6 162	4	0%
1 oct - 14 oct	64 110	5 516	12	0%
15 oct – 31 oct	78 828	6 863	22	0%
1 nov - 14 nov	65 420	5 757	50	1%
15 nov - 30 nov	70 946	6 058	225	4%
1 dec - 14 dec	65 273	5 201	1 144	22%
15 dec - 31 dec	73 891	5 783	4 173	72%

4.2 Selection of hospitals on which to estimate the models

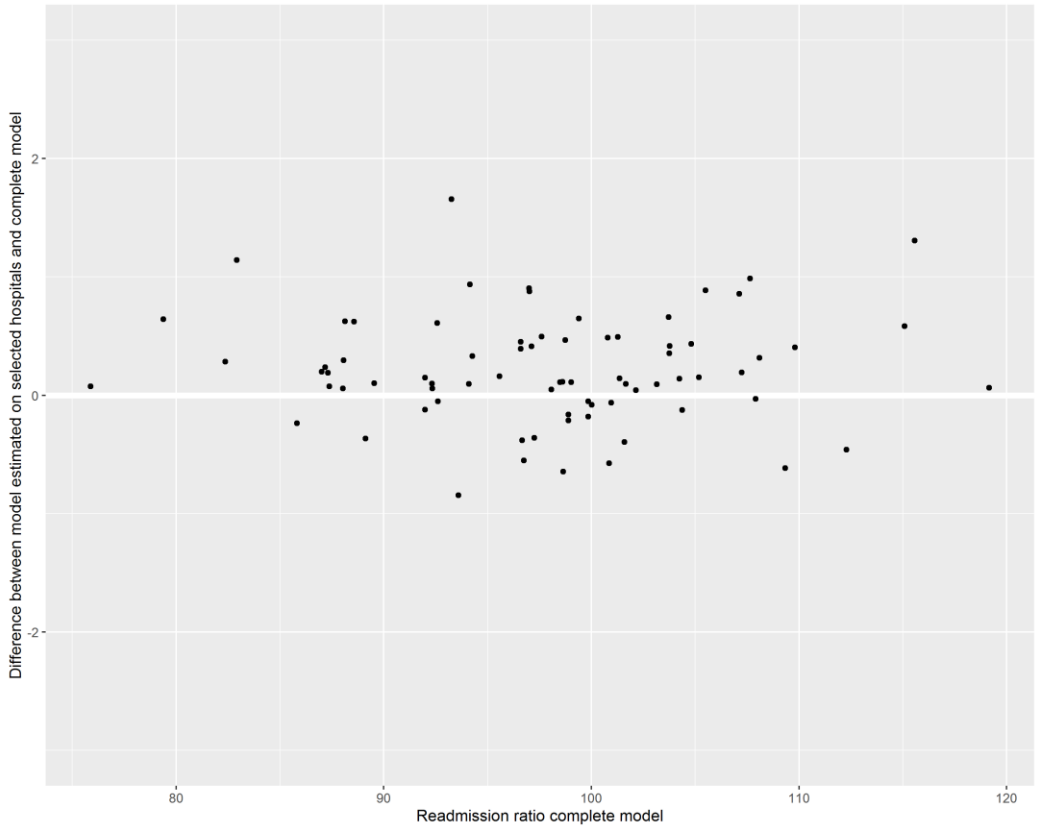
It is known that in the LBZ hospitals sometimes register the same hospital-specific patient ID number for different patients or different patient ID numbers for the same patient. The first issue usually indicates errors; the second is slightly more common and is often caused by a change in the hospital administrative recording system within a hospital and/or by a merger of hospitals having different patient registration systems, where in the year of the merge unique patients with admissions in both hospitals had been registered in the LBZ with different patient IDs. Readmissions are determined using the patient ID numbers that hospitals have registered in the LBZ. If the patient ID is not related to unique patients, this could lead to an incorrect estimate of the number of readmissions for these hospitals. It could also affect the models.

To investigate this effect on the models we first identified hospitals with patient ID registration issues. Statistics Netherlands has linked the LBZ data to the national population register, on the basis of which an unique (pseudonomyzed) national personal identifier could be added to the LBZ dataset. Using this linked dataset it can be derived for each hospital how often the same patient ID number is used for multiple persons and vice versa (the same person has different patient ID numbers). Using this information six hospitals were identified where these issues occurred more than occasionally. In order to investigate the effect of this on the outcomes of the remaining hospitals, the models and readmission ratios were recalculated excluding the data from the six hospitals with patient ID registration issues.

Figure 4.2.1 shows the effect of removing the hospitals mentioned above from the dataset on which the models are estimated. It shows that the effects on the ratios of the other hospitals were generally small. Therefore, it was not considered necessary to remove these hospitals from the dataset when estimating the final models (see Chapter 3). However, for the special

investigations described in this chapter (Chapter 4), these six hospitals together with the three smallest hospitals (defined as having less than 60 readmissions or less than 1000 index admissions) are left out of the presented results as their outcomes are considered non-representative.

4.2.1 Effect on the readmission ratio of removing hospitals with coding issues from the data set.



4.3 Selection of variables included in the model

As was mentioned earlier, in principle the same variables were included in the readmission models as in the HSMR models. In the HSMR models all variables are always included; no model selection is performed. Variables are only dropped from the models when only one category (all variables are categorical) has enough observations. However, the HSMR models are estimated using four years of data while the models for the readmission ratio are estimated using only one year of data. Therefore, the risk of overfit is larger for the models for the readmission ratio and it might be preferable to perform model selection.

Overfit is the situation in which a model fits the dataset on which it was estimated very well. However, when evaluating the model using a new data set (e.g. comparing the predicted values to the real values in a new data set), the model performs worse. This is usually caused by having too many parameters in the model relative to the size of the dataset. Because of this large number of parameters, the model is able to describe the peculiarities of the specific dataset on which it was estimated. The next sample of data has its own set of peculiarities and a very specific model will not describe a new dataset very well. Because one wants a model that

generalizes to new data, models are often evaluated using a dataset different than the one on which it was estimated. This is called cross-validation.

Another possible issue might be the use of the variable 'severity' which is derived from the HSMR models. This variable measures the severity of a diagnosis within a diagnosis group. However, this severity is determined using historical mortality. It therefore measures the severity in terms of mortality risks. It is possible that this does not have the same relevance for the likelihood of a readmission. Therefore it was investigated if this variable contributes to better prediction of the readmission probability or not.

The three methods investigated are:

1. The original model including all covariates (as long as more than one category has enough observations). This is the model as is also used in the HSMR. For this model the area under curve (AUC, also known as C-statistic) was determined using two methods.
    a. First the AUC was determined on the same dataset on which the model was estimated.
    b. Second, the AUC was determined using cross validation. As was mentioned before estimating the AUC using the same dataset on which the model was estimated generally overestimates the true AUC. Therefore, the AUC was also determined using 10-fold cross validation[2]. This generally gives a better estimate of the true AUC.
2. A model that includes all variables except severity (AUC determined using cross validation).
3. A model using model selection (AUC determined using cross validation).
   Model selection was performed using LASSO regression (see Tibshirani 1997). In LASSO regression all variables are included in the model, however, the effect they have on the prediction is restricted. The degree to which each variable is restricted depends on the effect each variable has on the target variable.

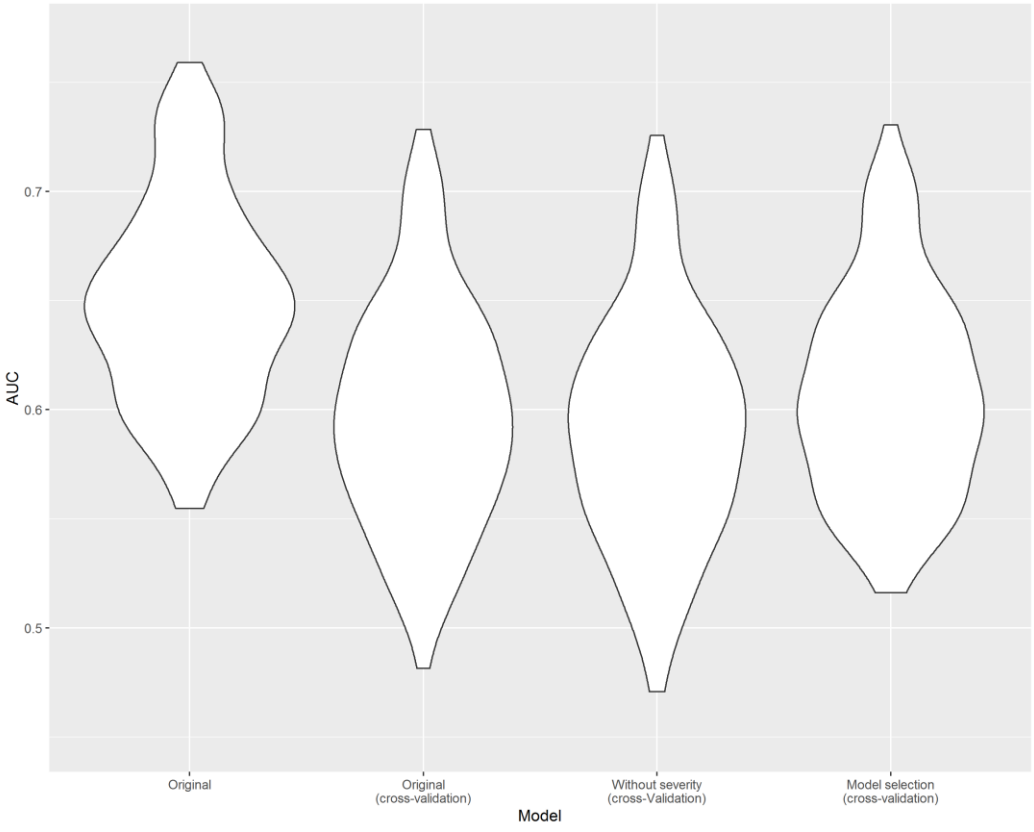
Figure 4.3.1 shows density plots (also called violin plots) of the AUC of the 122 diagnosis groups for the three methods investigated including the two variants for the first method resulting in four 'violins'. For each method (shown on the x-axis) a violin is drawn. Each violin shows the distribution of the AUC values of the 122 models. The violins are symmetric around their vertical axes and each violin has the same area. They show which AUC values occur more often (the violin is wider) and which less often (the violin' is narrower). So, for the first method ('Original') most AUC values are around 0.65. Values below 0.55 and above 0.77 have not been observed in the original model.

Table 4.3.2 shows the overall AUC value for the three different methods. This is the AUC of all diagnosis groups combined. The AUCs presented in the figure are for the models of each of the 122 diagnosis groups. For each diagnosis group a separate model is estimated. However, it is also possible to estimate an AUC combining the predictions of all diagnosis groups. As can be seen from the table and figure, the overall AUC is higher than when taking the average of all the individual diagnosis group AUCs. This is because the diagnosis group itself is an important predictor of readmission. In the models for each diagnosis group, the diagnosis itself is not

[2] The dataset is divided into 10 parts. One of the parts is removed from the dataset. The remaining 9 parts are used to estimate the model. Predictions are then calculated for the part removed from the dataset. This is repeated for each of the 10 parts. This results in predictions for the complete dataset. These predictions are used to calculate the AUC (for more information see James *et al.* 2013).

present in the model. Therefore, the AUC for that diagnosis group measures how well the model predicts differences in readmission rates within that diagnosis group (these are shown in figure 4.3.1) which might be difficult to predict. However, when there are large differences in the readmission probabilities between diagnosis groups, the overall AUC can still be high as a large part of the differences in readmission rates can be explained by the diagnosis groups. Therefore, because diagnosis (group) is an important predictor of readmission, the overall AUC is higher than the average AUC of each of the diagnosis groups separately.

4.3.1 Area under the curve for the original model, the original model after cross-validation, the model without severity (cross-validation), and the model after model selection (cross-validation).



4.3.2 Overall AUC for the original model, the original model after cross-validation, the model without severity (cross-validation) and the model after model selection (cross-validation).

Model	Overall AUC
Original	0,691
Original (cross-validation)	0,674
No severity (cross-validation)	0,671
Model selection (cross-validation)	0,687

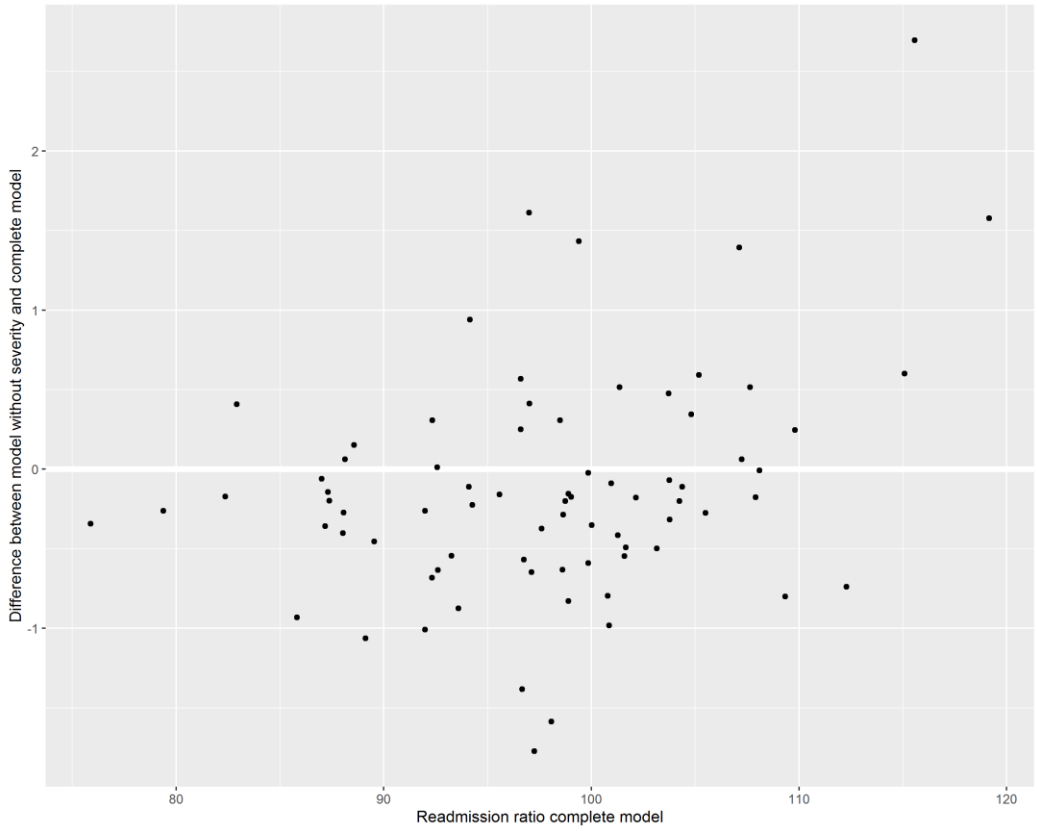
When comparing the original AUC and the cross-validated AUC of the complete models, we see that the cross-validated one is considerably lower. This is caused by the fact that the AUC of the

original one is estimated using the same data as was used for estimating the model. As mentioned earlier, in case of overfit, the model fits the data on which it is estimated very well, but fits new data less well. In case of cross-validation the AUC is calculated using different data than that was used to estimate the model, resulting in a lower but more realistic AUC.

Severity

Removing severity from the model has little effect on both the overall AUC and the individual AUC's for the diagnosis groups, compared to the original AUC's as determined using cross-validation. Figure 4.3.3 shows the effect on the readmission ratios of removing severity from the model. In general the effect is quite small (<1.5). However, there does seem to be a small (0.3) but significant (99% confidence) correlation between the original ratio and the change in the ratio when removing severity from the models. For hospitals with a high ratio the ratio increases even further when removing severity and for hospital with a low ratio the index decreases. This means the spread in the ratios increases when severity is removed from the models. Because of this and because the variable is an important predictor for readmission in a large number of the models as can be seen in table 3.2.2, it was decided to not remove severity from the models.

4.3.3 Difference between the readmission ratio without severity and with severity included in the models.



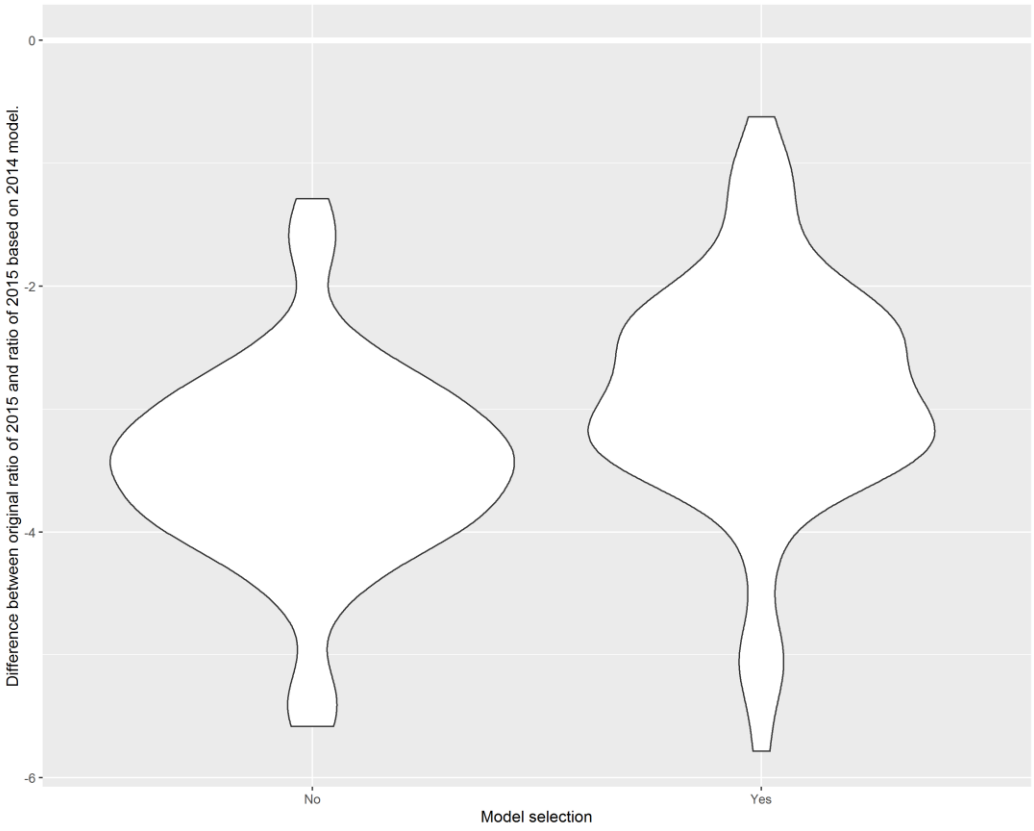
Model selection

Model selection does improve the AUC slightly compared to the original AUCs as determined using cross-validation, see figure 4.3.1 en table 4.3.2. As is discussed in the next section, the

model estimated on 2015 data is going to be used to calculate readmission probabilities for 2016 data. Model selection could help when applying the model to a new year as it could reduce overfitting. In order to see if this is indeed the case, the models were also estimated using 2014 data (details of the data selection are described in the next section) after which they were used to determine readmission ratios for 2015. The difference between these ratios and the original ratios (estimated using 2015 data) is an indication for how well the models generalize to a new year.

Figure 4.3.4 shows a density (violin) plot of these differences. In both cases the ratios estimated using the 2014 data underestimate the ratio for 2015. Both are biased: the total readmission ratio should be 100. In this case this ratio is approximately 3 points smaller than 100. The unweighted average bias of the models without model selection is larger than that of the models with model selection (−3.5 versus −2.9). However, as this bias can be measured and corrected for, the spread is more important. With model selection the standard deviation is higher than without model selection (0.96 versus 0.81). Overall the differences between the two methods are small. As it is in principle the goal to keep the methods for the readmission ratio and the HSMR the same wherever possible, it was decided to not apply model selection.

4.3.4 The difference between the readmission ratios of 2015 calculated using a model estimated on 2014 data and the original model based on 2015 data, with and without model selection.



4.4 Model validation

Unlike with the HSMR where Statistics Netherlands calculates the models and the HSMR, for the readmission ratio, Statistics Netherlands only calculates the model. The model is then used by

DHD to calculate the readmission ratios using more recent data. This means that the present models are calculated using data from November 1st 2014 up to October 31st 2015, while the ratios are calculated using data from 2016. Therefore, it is important to see how applicable the models are to more recent data. In order to investigate that, the models (including all variables; no model selection) were estimated on data from 2014 (from January 1st 2014 up to October 31st 2014) and ratios were calculated for 2015. These were compared to the ratios calculated using the models estimated using 2015 data. The 2014 model was estimated on fewer months than the 2015 model, since the final two months of 2013 could not be added. In 2013 the LBZ data were registered using a different data model and as a result different patient ID numbers were used when compared to 2014. Therefore, data from 2013 and 2014 could not be used in the same model.

4.4.1 Difference between the readmission ratio for 2015 calculated using models estimated on 2014 data and the ratio for 2015 calculated using 2015 data.

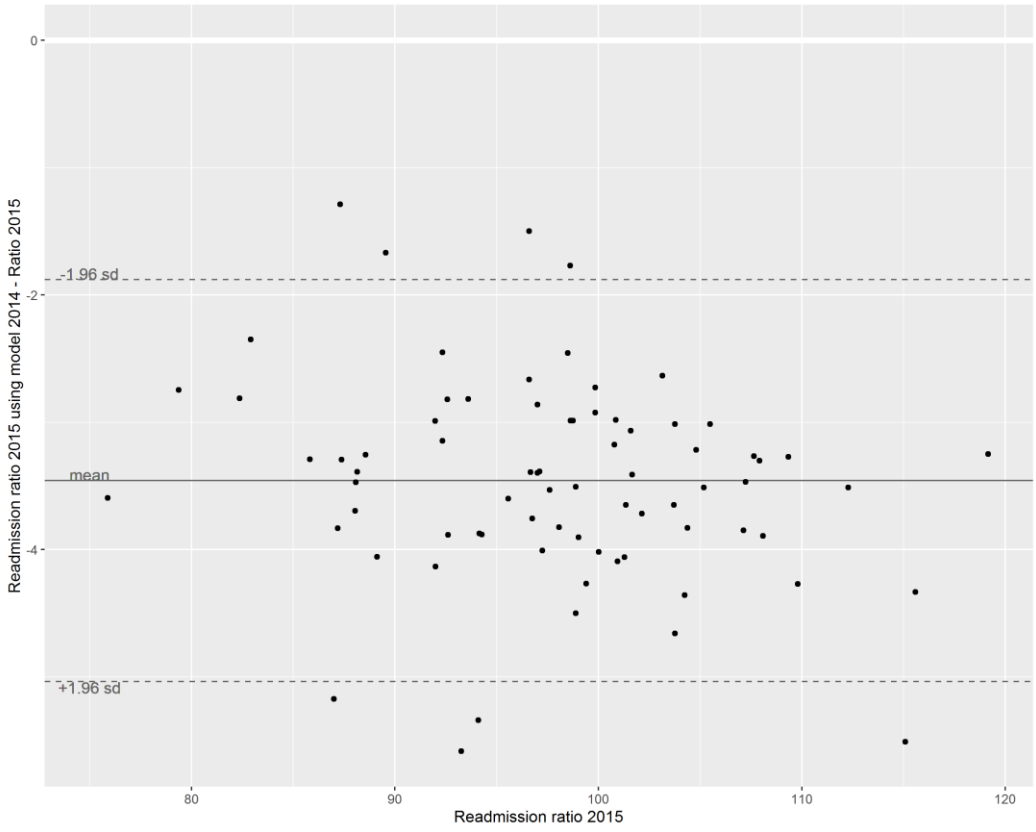


Figure 4.4.1 shows the difference between the two ratios. As can be seen from the figure the models estimated using 2014 data underestimate the ratio for 2015. The weighted average ratio is 96.4 while it should be 100. This is because, given the severity distribution of the patients' condition, the probability of readmission has increased in 2015. The average unadjusted readmission rate has remained approximately constant, which means that the probability of readmission for 2015 is overestimated using the 2014 model. Fortunately the average bias in the ratio can be measured as the weighted average should be 100. Therefore, it is possible to show the bias (presenting the weighted average) or correct for the bias when presenting the results. The length of the confidence intervals do not change when results are corrected for the bias.

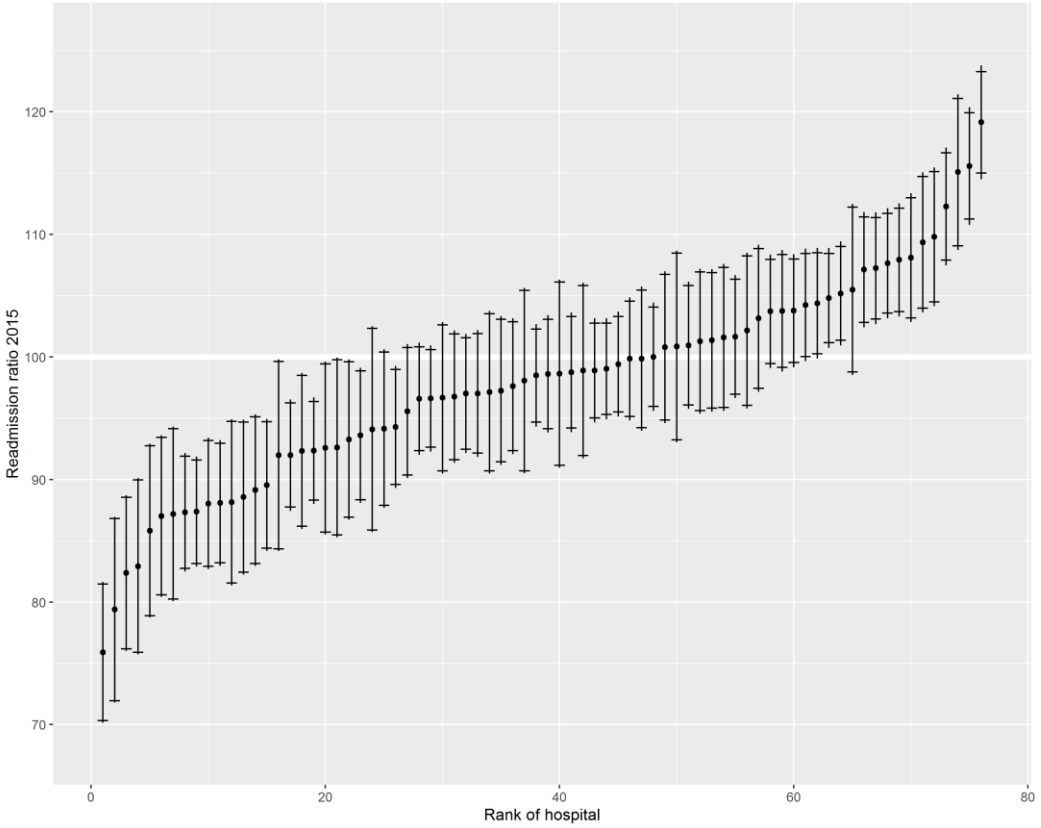
The spread in the difference between the two ratios is a measure of the uncertainty in the ratios caused by uncertainty in the models. The standard deviation is 0.81, which is relatively small compared to uncertainty caused by the statistical variation in the number of readmissions (a hospital with 100 admissions and a 10% readmission rate might have had 10 readmissions, but that could easily have been 11 or 9 readmissions resulting in a different ratio).

Usually the uncertainty introduced by the model is ignored when calculating the confidence intervals for the ratios. However, it is in principle possible to include the uncertainty in the model. The standard deviation in the readmission ratio I_h for hospital *h* is given by

$$\sigma_{I,h} = I_h \sqrt{\frac{1}{O_h} + \sigma_{model}^2}$$

where O_h is the number of observed readmissions and σ_{model}^2 the relative variance in the predicted number of readmissions. This variance can also be estimated by comparing the predicted numbers of admissions from both models and is estimated to be 0.000082. Figure 4.4.2 shows the difference between the two intervals. The differences are very small, and it is therefore not necessary to take the model uncertainty into account when estimating the confidence intervals.

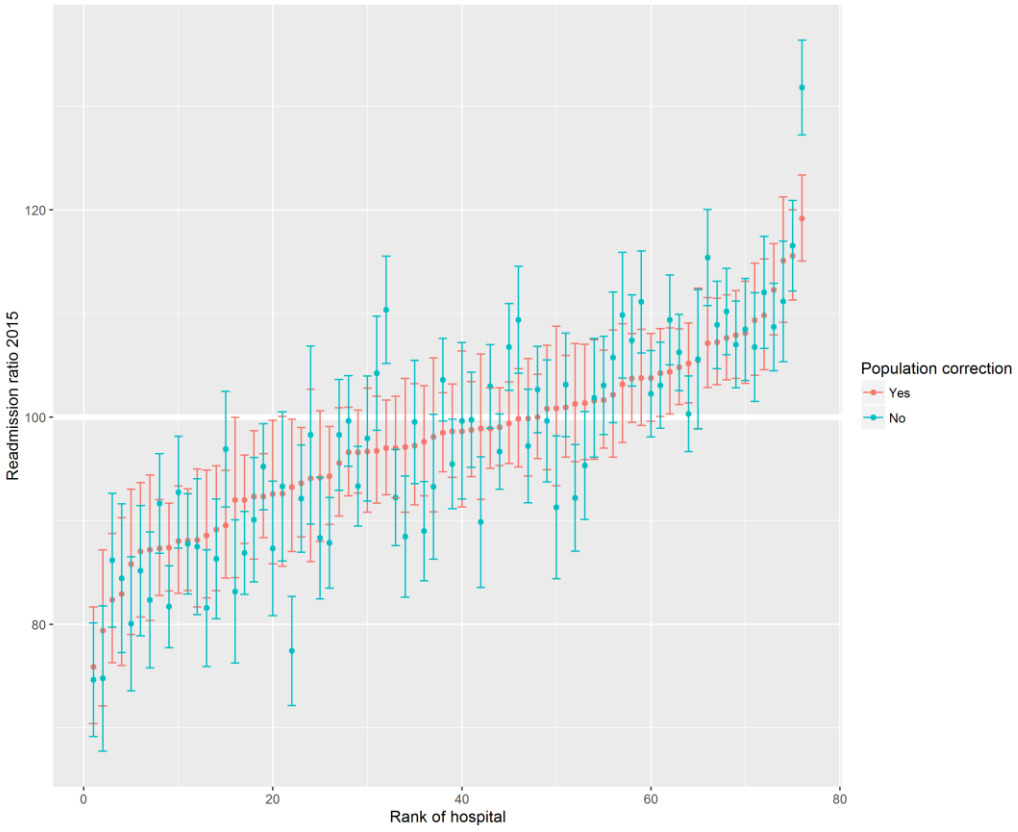
4.4.2 Comparison between the 95% confidence intervals including and excluding model uncertainty (the intervals excluding model uncertainty are indicated by the horizontal lines).



4.5 Effect of case mix correction

As could be seen in section 4.3 the AUCs of the models are generally low. Most of them are between 0.5 and 0.7, while in general values below 0.7 indicate poor fit. The individual AUCs are also shown in the appendix. The overall AUC as determined using cross-validation is 0.67. One of the goals of modelling readmission is to correct for differences in the patient population between the hospitals to make the ratios more comparable than with unadjusted readmission ratios. When the variables included in the model explain the relevant population differences, the models correct for the population differences even when the fit is poor. In case the remaining unexplained variance is either caused by hospital policies (for which we do not want to correct as that is what we want to measure) or is explained by background properties for which no differences between the hospitals are expected, we have corrected for all relevant population differences. The problem with a poor fit is however that it is not unlikely that there are still variables present that affect the probability of readmission and for which hospitals differ in their patient populations.

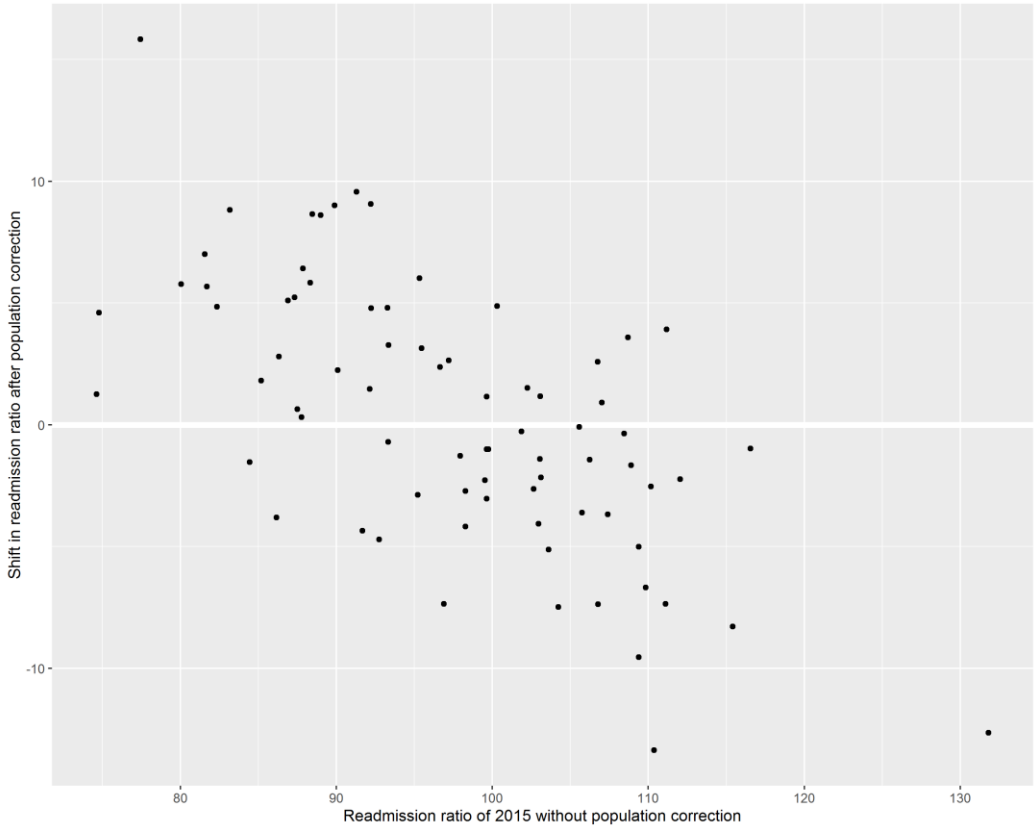
4.5.1 Comparison of readmission ratios corrected and not corrected for patient population.



The goal of the case mix correction by using relevant covariates in the model is to remove differences between hospitals caused by population differences between the hospitals. Therefore, one would expect that the differences between the hospitals decrease because of this correction. Figure 4.5.1 shows the adjusted (corrected) and unadjusted (uncorrected) readmission ratios. It shows that for a number of hospitals the correction does influence the ratio. For some hospitals the ratio increases because of the correction. These are probably hospitals with a relatively 'light' patient population (with a less severe condition), where relatively few readmissions are expected. For some other hospitals the ratio decreases. These

are hospitals with a relatively 'heavy' patient populations (with a more severe condition), where the unadjusted ratio overestimates the readmissions. Overall, the spread in the ratios decreases from 10.8 to 8.3 because of the correction. This indicates that the correction does have an effect and although the correction might not be complete it does make the figures more comparable. This is also confirmed by figure 4.5.2 which shows the shift in the readmission ratio as a result of the case mix correction. It shows that hospitals with a low unadjusted readmission ratio are shifted upwards on average while hospitals with a high unadjusted readmission ratio are shifted downwards (the correlation is -0.66; significant with 95% certainty).

4.5.2 Shift in readmission ratio after case mix correction.



4.6 Conclusion

We have investigated whether it is necessary to modify the method used for the calculation of the Hospital Mortality Ratio for the Hospital Readmission Ratio and we have evaluated the quality of the estimated models. The conclusion is that the method of the HSMR does not need to be adapted. We can use the same method as used for the HSMR. However, the estimated models have weak predictive power. It is therefore not unlikely that the used case mix correction is incomplete. When looking at the effect of the case mix correction we see that the correction does reduce some of the differences between the hospitals. Therefore, although the correction is probably incomplete, the correction does help in making the readmission rates more comparable.

References

HSCIC (2016). Indicator Specification: Summary Hospital-level Mortality Indicator. Version 1.22, 24 February 2016. Health & Social Care Information Centre. https://www.content.digital.nhs.uk/shmi

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani (2013). *An Introduction to statistical learning*, Springer, New York.

Kansagara D., H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman and S. Kripalani (2011). "Risk prediction models for hospital readmission: a systematic review", *The Journal of the American Medical Association* 306(15), pp. 1688-98. DOI: 10.1001/jama.2011.1515.

R Core Team (2015). *R: A language and environment for statistical computing,* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Tibshirani, Robert (1997). "Regression shrinkage and selection via the LASSO", *Journal of the Royal Society Series B* 58 (1), pp. 267-288.

Van der Laan, Jan, Agnes de Bruin, Corine Penning, Marjolein Peters and Frank Pijpers (2016). HSMR *2015: Methodological report*, Statistics Netherlands, The Hague. URL: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende%20onderzoeksbeschrijvingen/hsmr-2015-methodological-report

Appendix I: Results of the logistic regressions

Statistical significance (95% confidence) of the covariates for the 122 logistic regressions (1=significant; 0=non-significant; "-"=variable dropped because all categories are collapsed, due to < 50 admissions or no readmissions in all but one category).

Diagnosis group	Sex	Age	Urgency	Severity	Comorbidity_1	Comorbidity_2	Comorbidity_3	Comorbidity_4	Comorbidity_5	Comorbidity_6	Comorbidity_7	Comorbidity_8	Comorbidity_9	Comorbidity_10	Comorbidity_11	Comorbidity_12	Comorbidity_13	Comorbidity_14	Comorbidity_15	Comorbidity_16	Comorbidity_17	SES	Month admission	Source admission
1	0	1	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1
2	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	1	1	-	0	-	0	0	0
3	0	1	0	0	0	0	0	-	0	0	-	-	0	0	0	-	0	1	-	0	-	0	0	0
4	0	0	0	1	-	-	-	-	-	0	-	-	-	0	-	-	0	1	-	-	-	0	0	-
5	0	0	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	0	0	-
6	0	1	1	1	0	1	1	0	0	0	0	-	1	1	-	-	1	0	-	0	-	0	1	0
37	1	1	1	1	0	0	1	0	-	1	0	-	1	1	0	0	0	0	-	0	-	0	0	0
38	1	1	1	0	0	0	1	0	0	0	0	-	0	0	0	-	1	0	-	0	-	0	0	0
39	1	1	1	0	0	1	1	0	1	0	0	-	0	0	0	-	1	1	-	0	-	0	0	1
40	1	1	0	1	0	1	1	0	0	0	0	-	0	0	1	-	1	0	-	-	-	0	0	0
41	1	1	1	1	0	0	0	0	1	0	0	-	0	0	0	-	0	0	-	0	-	0	0	0
42	1	1	1	0	1	0	1	0	0	1	0	-	0	0	0	1	1	1	-	0	-	0	1	0
43	0	0	0	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	0	0	-
44	0	1	1	0	0	0	0	0	-	0	-	-	0	0	-	-	0	1	-	0	-	0	0	0
45	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	0	1	1	-	0	1	0	0	0
46	0	0	0	-	0	-	-	-	-	0	-	-	-	0	-	-	0	0	-	0	-	1	0	0
51	1	0	1	1	-	-	-	-	-	0	-	-	-	0	-	1	-	0	-	-	-	0	0	0
52	0	1	0	-	-	-	-	-	1	0	-	-	-	0	-	-	-	-	-	-	-	0	0	-
53	1	1	1	1	0	-	-	0	0	1	-	-	-	0	-	-	-	-	-	-	-	0	0	0
54	0	0	0	0	0	-	-	0	-	0	-	-	-	1	-	0	-	0	-	0	-	0	0	0
55	0	1	0	1	0	0	1	0	0	0	0	-	0	1	0	0	0	0	-	0	-	0	0	0
56	0	1	0	-	-	-	-	0	0	0	-	-	-	0	-	-	0	0	-	-	-	0	0	0
57	1	1	1	1	0	1	0	1	0	1	0	-	1	1	0	0	1	1	-	1	-	0	0	1
58	1	1	1	1	1	1	1	1	0	1	0	-	1	0	0	0	0	0	-	0	-	1	1	0
59	0	1	1	0	0	0	1	0	0	0	0	-	0	0	0	-	0	0	-	-	-	0	0	1
60	0	1	1	1	1	1	0	0	-	0	0	-	0	1	-	-	1	0	-	0	-	0	0	0
61	0	0	0	1	1	1	1	0	-	0	-	-	-	0	1	-	1	-	-	-	-	0	0	0
62	1	1	1	1	1	1	1	0	0	1	1	-	1	1	1	0	1	1	-	0	-	0	0	1
63	0	1	1	0	1	1	1	1	0	1	0	-	1	1	1	0	1	0	-	0	-	0	0	1
64	1	1	0	-	1	1	1	1	0	1	0	-	1	1	1	-	1	1	-	0	-	0	0	0
65	0	1	1	0	1	1	0	0	0	1	1	-	0	0	-	-	1	1	-	1	-	0	1	0
66	0	0	1	0	0	0	0	-	-	0	-	-	-	0	-	-	0	-	-	-	-	0	0	0
67	0	1	1	0	1	1	1	0	0	1	0	-	-	0	0	-	0	0	-	-	-	1	0	0
68	1	1	1	1	1	1	1	1	0	1	1	-	0	1	1	-	1	1	-	0	-	0	0	0
69	0	0	0	0	1	0	0	-	-	0	-	-	-	0	-	-	0	-	-	-	-	0	0	0
70	1	1	0	-	1	1	1	0	1	1	0	-	0	1	1	0	1	0	-	0	0	0	0	1
71	1	1	0	1	1	0	1	1	0	1	1	-	0	0	1	0	1	0	-	0	-	0	0	0
72	1	1	0	1	0	0	1	1	0	0	0	-	0	0	1	0	1	0	-	0	-	0	0	0
73	0	1	1	1	0	0	1	0	0	0	0	-	0	0	1	-	1	0	-	0	-	1	0	0
74	0	0	1	1	0	0	0	0	-	0	0	-	0	0	-	-	0	0	-	-	-	0	0	0
75	0	1	1	1	0	1	1	0	1	0	1	-	-	0	1	-	1	0	-	0	-	0	0	0
76	0	0	1	1	1	0	0	0	0	1	0	-	0	0	1	0	1	0	-	0	-	0	0	0
77	0	1	1	1	0	0	1	-	-	0	0	-	1	1	-	-	1	1	-	0	0	0	0	0
78	1	1	1	0	0	1	1	1	1	1	1	-	0	1	0	0	1	1	-	0	0	0	1	1
79	0	1	1	0	0	1	0	0	0	1	1	-	-	0	0	-	0	0	-	0	-	0	0	0
80	1	1	1	0	0	1	0	0	0	1	0	-	0	0	0	-	0	1	-	1	-	0	0	0
81	1	0	0	0	0	0	0	-	0	0	-	-	-	0	-	-	0	0	-	-	-	0	0	1
82	1	1	1	1	1	1	1	0	0	0	0	-	0	1	1	0	1	1	-	0	-	1	0	0
83	0	1	0	1	1	1	0	-	-	1	1	-	-	0	-	-	0	-	-	-	-	0	0	0
84	1	1	0	-	0	0	0	0	0	1	-	-	-	0	-	0	0	0	-	-	-	0	0	0
85	0	1	1	0	0	0	0	0	-	0	0	-	1	0	0	-	1	0	-	0	-	0	0	0
86	0	1	0	-	0	0	0	-	-	1	-	-	-	0	-	-	0	0	-	-	-	0	0	1
87	0	0	0	1	-	-	-	-	-	0	-	-	-	0	-	-	-	0	-	0	-	0	0	-

Diagnosis group	Sex	Age	Urgency	Severity	Comorbidity_1	Comorbidity_2	Comorbidity_3	Comorbidity_4	Comorbidity_5	Comorbidity_6	Comorbidity_7	Comorbidity_8	Comorbidity_9	Comorbidity_10	Comorbidity_11	Comorbidity_12	Comorbidity_13	Comorbidity_14	Comorbidity_15	Comorbidity_16	Comorbidity_17	SES	Month admission	Source admission
88	0	1	0	1	0	1	0	0	0	1	1	-	0	1	0	-	1	0	-	0	-	0	0	0
89	0	1	1	1	1	1	0	0	0	1	0	-	1	1	1	-	1	1	-	0	-	0	0	1
90	0	1	1	1	0	1	1	0	0	1	1	-	0	0	0	-	1	1	-	0	-	0	0	0
91	0	1	1	1	0	-	-	-	-	1	-	-	-	0	-	-	-	1	-	1	-	0	0	-
92	0	1	0	1	0	0	0	-	-	0	-	0	1	0	-	-	1	1	-	0	-	1	0	0
93	0	0	0	0	-	-	-	-	-	0	-	-	-	0	-	-	0	-	-	-	-	1	0	-
94	1	0	0	1	0	-	0	-	-	0	-	0	1	1	0	-	0	1	-	0	-	0	0	-
95	0	1	0	0	0	-	0	-	-	1	-	-	-	0	-	-	1	0	-	-	-	0	0	0
96	0	1	0	0	0	-	-	-	-	0	-	-	1	1	-	-	1	0	-	0	-	0	0	-
97	0	1	1	1	0	0	1	0	0	1	0	-	1	1	0	-	1	0	-	0	-	0	0	0
98	0	0	1	-	0	-	-	-	-	0	-	-	0	0	-	-	0	-	-	-	-	0	0	0
99	0	1	0	1	0	0	0	0	0	0	0	-	0	0	0	0	0	1	-	0	-	0	0	0
100	0	1	1	0	1	0	0	1	1	1	0	-	0	0	0	-	1	0	-	0	-	0	0	1
101	0	0	1	1	0	-	0	-	-	0	-	-	-	1	-	-	0	1	-	0	-	0	0	-
102	1	1	1	1	1	0	1	0	0	0	0	0	1	1	0	0	0	1	-	1	0	0	0	0
103	1	1	1	1	0	-	-	-	-	0	-	-	0	0	-	-	0	-	-	-	0	0	1	0
104	1	1	0	1	0	0	1	-	-	0	0	-	1	1	1	-	0	0	-	1	0	0	0	0
105	0	1	0	1	1	0	0	0	-	0	0	-	0	0	0	-	1	0	-	0	-	1	0	0
106	1	1	0	0	1	1	0	1	0	1	0	0	1	1	0	-	1	1	-	0	1	0	0	0
107	1	1	0	1	0	0	1	0	0	0	0	-	0	0	1	-	1	1	-	0	-	0	1	0
108	1	1	1	0	0	0	0	0	1	0	0	-	1	1	0	1	1	1	-	0	-	0	0	0
109	1	1	0	0	0	-	0	-	-	0	0	-	1	0	0	-	1	0	-	0	-	1	0	0
110	0	0	0	0	1	1	1	0	0	0	0	-	0	0	0	-	1	0	-	0	-	0	0	0
111	0	1	1	0	0	0	0	0	-	0	0	-	0	0	0	-	1	0	-	1	-	0	0	0
112	1	1	1	1	0	1	1	0	1	1	1	-	0	1	0	0	1	1	-	1	0	0	0	1
113	0	0	1	1	0	0	0	0	0	1	0	-	0	0	0	0	1	1	-	0	-	0	0	0
114	1	1	1	1	0	0	0	1	1	0	0	-	0	0	0	-	0	1	-	0	-	0	0	0
115	-	1	1	0	1	0	1	-	-	1	-	-	-	0	1	-	0	0	-	-	-	0	0	-
116	0	1	1	-	-	-	-	-	-	0	-	-	-	1	-	-	-	0	-	-	-	0	0	-
117	-	1	1	1	1	-	0	-	-	0	0	-	-	1	-	-	1	1	-	0	-	0	0	0
119	0	1	1	0	0	1	1	0	0	1	0	-	0	1	1	0	1	1	-	0	-	0	0	0
120	1	1	0	1	0	0	1	1	0	0	0	-	1	1	1	0	0	0	-	0	-	1	1	0
121	1	1	0	1	1	0	0	-	-	1	0	-	0	1	1	-	1	0	-	-	-	0	0	0
122	1	1	1	1	1	1	1	0	0	1	1	-	1	1	0	0	0	0	-	1	-	0	1	0
123	0	1	1	0	0	-	0	0	0	0	0	-	-	0	-	-	0	1	-	0	-	0	0	-
124	0	1	1	0	0	0	0	0	1	0	1	-	0	1	0	0	1	1	-	0	-	0	0	0
125	1	0	1	0	0	0	0	-	0	0	0	-	-	1	-	-	1	0	-	0	-	0	0	0
126	1	1	1	1	0	1	0	0	0	1	1	-	1	0	0	0	0	1	-	0	-	0	1	0
127	0	1	0	0	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0
128	0	1	1	1	-	-	-	-	-	1	-	-	-	0	-	-	1	-	-	-	-	0	0	0
129	1	-	0	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1
130	1	-	0	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1
131	0	0	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1
132	0	1	1	0	0	-	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	0	0	-
133	1	1	0	1	1	1	1	0	1	1	0	-	0	0	0	0	1	0	-	0	-	0	0	0
134	0	1	1	1	0	-	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	0	0	0
135	0	1	1	0	0	0	0	0	0	0	0	-	0	0	1	-	0	0	-	-	-	0	0	0
136	1	1	1	1	1	0	0	0	0	0	0	-	-	0	0	0	0	0	-	-	-	0	0	0
137	0	1	1	0	0	1	1	0	0	1	0	-	0	0	1	0	1	0	-	0	-	0	0	0
138	0	1	0	1	0	1	0	0	1	0	0	-	1	0	0	0	0	0	-	0	-	0	0	1
139	0	1	1	1	1	-	0	-	-	1	-	-	-	0	-	-	1	-	-	-	-	0	0	0
140	0	1	1	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-	-	-	-	0	0	-
141	1	1	1	0	-	-	1	-	-	0	-	-	-	0	-	-	-	-	-	-	-	0	0	-
142	1	1	1	1	1	1	1	0	0	1	0	-	1	1	1	0	1	1	-	0	-	0	1	0
143	1	1	1	1	1	0	1	0	0	1	0	-	1	1	1	0	1	0	-	1	-	0	0	0
144	1	1	0	0	1	1	1	1	0	0	0	-	0	1	0	0	0	0	-	0	-	0	0	0
145	0	0	0	1	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	0	0	0
146	1	1	0	0	1	0	0	-	-	0	-	-	0	0	-	-	1	0	-	0	-	0	0	1
147	0	1	0	1	0	0	1	0	0	1	-	-	-	0	-	-	0	1	-	-	-	0	0	0
148	1	1	1	-	1	1	0	0	1	1	0	-	1	0	1	0	1	1	-	0	-	0	0	1
149	0	1	0	-	0	0	1	1	0	0	0	-	1	0	0	-	1	1	-	0	-	0	0	0
150	0	1	0	1	0	0	0	-	-	0	-	-	-	0	0	-	0	0	-	-	-	0	0	0
151	0	0	-	-	0	0	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	0	0	-

Diagnosis group	Sex	Age	Urgency	Severity	Comorbidity_1	Comorbidity_2	Comorbidity_3	Comorbidity_4	Comorbidity_5	Comorbidity_6	Comorbidity_7	Comorbidity_8	Comorbidity_9	Comorbidity_10	Comorbidity_11	Comorbidity_12	Comorbidity_13	Comorbidity_14	Comorbidity_15	Comorbidity_16	Comorbidity_17	SES	Month admission	Source admission
152	0	0	0	-	0	0	1	-	1	0	0	-	0	0	0	-	0	1	-	1	-	0	0	0
153	1	1	1	-	0	0	1	0	0	1	0	-	0	0	1	-	1	1	-	0	-	0	0	0
154	0	1	1	-	0	1	0	0	1	0	0	-	0	1	0	-	1	1	-	0	-	0	0	-
155	0	1	1	1	0	-	-	-	-	1	-	-	-	0	-	-	0	0	-	0	-	0	1	-
156	1	1	1	1	0	1	1	0	0	1	0	-	1	0	1	0	1	1	-	0	-	0	0	1
157	1	1	0	1	0	1	1	0	1	0	0	-	1	0	0	-	0	0	-	0	-	0	0	0
	50	**95**	**72**	**65**	**35**	**39**	**46**	**14**	**18**	**47**	**14**	**0**	**30**	**38**	**28**	**3**	**59**	**43**	**1**	**11**	**2**	**11**	**13**	**20**

The numbers of the comorbidity groups in the header of the table above are the following comorbidities:

Comorbidity_1 - Acute myocardial infarction
Comorbidity_2 - Congestive heart failure
Comorbidity_3 - Peripheral vascular disease
Comorbidity_4 - Cerebral vascular accident
Comorbidity_5 - Dementia
Comorbidity_6 - Pulmonary disease
Comorbidity_7 - Connective tissue disorder
Comorbidity_8 - Peptic ulcer
Comorbidity_9 - Liver disease / Severe liver disease
Comorbidity_9 - Liver disease / Severe liver disease
Comorbidity_10 - Diabetes / Diabetes complications
Comorbidity_11 - Diabetes complications
Comorbidity_12 - Paraplegia
Comorbidity_13 - Renal disease
Comorbidity_14 - Cancer
Comorbidity_15 - HIV
Comorbidity_16 - Metastatic cancer
Comorbidity_17 - Severe liver disease

Appendix II: AUC

The area under the curve (AUC) or C-Statistic for the logistic regressions of the 122 main diagnosis groups.

Diagnosis group*)		Number of index admissions	Number of readmissions	AUC original	AUC cross-validated
1	Tuberculosis	440	69	0.76	0.66
2	Septicemia (except in labor)	4 324	664	0.59	0.53
3	Bacterial infection; unspecified site	1 624	245	0.64	0.54
4	Mycoses	544	99	0.70	0.56
5	HIV infection	335	76	0.69	0.58
6	Hepatitis, viral and other infections	6 549	596	0.66	0.62
37	Other and unspecified benign neoplasm	18 697	1 144	0.66	0.64
38	Thyroid and other endocrine disorders	6 263	492	0.70	0.65
39	Diabetes mellitus without complication	4 727	367	0.69	0.64
40	Diabetes mellitus with complications	5 943	1 024	0.66	0.64
41	Nutritional deficiencies and other nutritional, endocrine, and metabolic disorders	14 166	1 240	0.68	0.65
42	Fluid and electrolyte disorders	8 561	1 067	0.67	0.64
43	Cystic fibrosis	710	125	0.63	0.51
44	Immunity and coagulation disorders, hemorrhagic disorders	3 136	555	0.62	0.54
45	Deficiency and other anemia	11 284	1 784	0.61	0.59
46	Diseases of white blood cells	1 979	349	0.60	0.51
51	Meningitis, encephalitis, and other central nervous system infections	2 094	217	0.65	0.56
52	Parkinson`s disease	1 692	165	0.64	0.56
53	Multiple sclerosis and other degenerative nervous system conditions	3 838	404	0.67	0.63
54	Paralysis and late effects of cerebrovascular disease	1 226	115	0.67	0.50
55	Epilepsy and convulsions	11 417	1 015	0.60	0.56
56	Coma, stupor, and brain damage	785	97	0.69	0.55
57	Headache and other disorders of the sense organs	21 196	1 014	0.64	0.61
58	Other nervous system disorders	30 707	1 621	0.66	0.64
59	Heart valve disorders	9 721	1 316	0.57	0.54
60	Peri-, endo-, myocarditis, and cardiomyopathy	4 803	573	0.64	0.58
61	Essential hypertension, hypertension with compl., and secondary hypertension	3 436	287	0.67	0.62
62	Acute myocardial infarction	31 074	3 569	0.60	0.59
63	Coronary atherosclerosis and other heart disease	40 102	4 079	0.59	0.58
64	Nonspecific chest pain	24 771	1 615	0.64	0.63
65	Pulmonary heart disease	8 416	669	0.65	0.61
66	Other and ill-defined heart disease	1 146	145	0.64	0.51
67	Conduction disorders (heart disease)	5 396	428	0.65	0.59
68	Cardiac dysrhythmias	35 213	3 522	0.65	0.64
69	Cardiac arrest and ventricular fibrillation	1 841	174	0.65	0.56
70	Congestive heart failure, nonhypertensive	26 168	4 392	0.59	0.58
71	Acute cerebrovascular disease	30 071	2 381	0.61	0.60
72	Transient cerebral ischemia, and other cerebrovascular disease	12 466	1 106	0.63	0.60

Diagnosis group[*)]		Number of index admissions	Number of readmissions	AUC original	AUC cross-validated
73	Peripheral and visceral atherosclerosis	8 144	1 480	0.63	0.61
74	Aortic and other artery aneurysms	6 750	882	0.56	0.51
75	Aortic and arterial embolism or thrombosis	4 846	805	0.62	0.59
76	Other circulatory disease	8 835	1 283	0.60	0.57
77	Phlebitis, varicose veins, and hemorrhoids	4 023	401	0.68	0.63
78	Pneumonia	33 521	3 745	0.61	0.60
79	Influenza	2 392	222	0.68	0.59
80	Tonsillitis and upper respiratory infections	24 209	1 570	0.68	0.67
81	Acute bronchitis	6 065	670	0.58	0.53
82	Chronic obstructive pulmonary disease and bronchiectasis	34 525	6 443	0.56	0.55
83	Asthma	7 841	860	0.60	0.55
84	Aspiration pneumonitis; food/vomitus	1 421	240	0.63	0.54
85	Pleurisy; pneumothorax; pulmonary collapse	6 432	987	0.61	0.58
86	Respiratory failure; insufficiency; arrest	1 424	240	0.66	0.56
87	Lung disease due to external agents	476	74	0.67	0.52
88	Other lower respiratory disease	7 373	965	0.60	0.57
89	Other upper respiratory disease	21 795	1 645	0.74	0.73
90	Intestinal infection	10 538	1 006	0.65	0.62
91	Disorders of mouth, teeth, and jaw	5 283	141	0.73	0.60
92	Esophageal disorders	3 899	467	0.65	0.60
93	Gastroduodenal ulcer	1 197	110	0.67	0.53
94	Gastritis, duodenitis, and other disorders of stomach and duodenum	2 210	300	0.67	0.61
95	Appendicitis and other appendiceal conditions	15 999	1 125	0.58	0.55
96	Peritonitis and intestinal abscess	1 104	233	0.66	0.57
97	Abdominal hernia	13 151	1 029	0.65	0.62
98	Regional enteritis and ulcerative colitis	4 793	735	0.57	0.51
99	Intestinal obstruction without hernia	8 134	1 243	0.57	0.53
100	Diverticulosis and diverticulitis	9 895	1 067	0.60	0.56
101	Anal and rectal conditions	6 110	586	0.59	0.53
102	Biliary tract disease	36 233	4 789	0.65	0.64
103	Liver disease; alcohol-related	1 584	429	0.64	0.58
104	Other liver diseases	4 418	1 062	0.66	0.63
105	Pancreatic disorders (not diabetes)	7 633	1 547	0.59	0.56
106	Gastrointestinal hemorrhage	9 059	1 270	0.62	0.59
107	Noninfectious gastroenteritis	8 937	965	0.63	0.60
108	Other gastrointestinal disorders	11 422	1 562	0.60	0.57
109	Nephritis; nephrosis; renal sclerosis	3 851	470	0.65	0.60
110	Acute and unspecified renal failure	4 194	757	0.60	0.55
111	Chronic kidney disease	3 956	815	0.60	0.54
112	Urinary tract infections	21 340	2 681	0.61	0.59
113	Calculus and other diseases of urinary tract	21 717	3 202	0.63	0.61
114	Genitourinary symptoms and ill-defined conditions	8 192	1 037	0.69	0.67
115	Hyperplasia of prostate and other male genital disorders	11 641	1 008	0.63	0.60
116	Nonmalignant breast conditions	4 411	146	0.71	0.64
117	Prolapse and other female genital disorders	19 251	877	0.67	0.64

Diagnosis group[*)]		Number of index admissions	Number of readmissions	AUC original	AUC cross-validated
119	Skin and subcutaneous tissue infections	13 307	1 139	0.66	0.64
120	Other skin disorders, chronic ulcer of skin	6 080	858	0.70	0.67
121	Infective arthritis and osteomyelitis	3 393	439	0.64	0.58
122	Osteoarthritis, rheumatoid arthritis, and other musculoskeletal deformities	63 740	2 911	0.65	0.64
123	Other non-traumatic joint disorders	4 919	249	0.74	0.68
124	Spondylosis, back problems, and osteoporosis	25 114	1 488	0.67	0.65
125	Pathological fracture	1 972	226	0.65	0.55
126	Other connective tissue disease	14 509	545	0.73	0.70
127	Cardiac and circulatory congenital anomalies	2 440	320	0.65	0.59
128	Noncardiac congenital anomalies	8 301	721	0.70	0.68
129	Short gestation; low birth weight; and fetal growth retardation	15 586	1 599	0.63	0.62
130	Intrauterine hypoxia, perinatal asphyxia, and jaundice	13 403	772	0.60	0.58
131	Other perinatal conditions	51 410	2 651	0.55	0.54
132	Joint disorders and dislocations; trauma-related; sprains and strains	9 909	259	0.74	0.70
133	Fracture of neck of femur (hip)	19 088	1 327	0.60	0.58
134	Skull and face fractures, spinal cord injury	3 013	158	0.69	0.59
135	Fracture of upper limb	12 890	726	0.73	0.70
136	Fracture of lower limb	14 103	1 173	0.67	0.65
137	Other fractures	11 807	715	0.63	0.59
138	Intracranial injury	12 426	583	0.73	0.70
139	Crushing injury or internal injury	5 679	286	0.70	0.64
140	Open wounds of head; neck; and trunk	1 720	89	0.74	0.57
141	Open wounds of extremities	1 488	107	0.73	0.63
142	Complication of device, implant or graft	23 850	3 563	0.63	0.62
143	Complications of surgical procedures or medical care	24 520	3 395	0.59	0.58
144	Superficial injury; contusion	10 074	465	0.74	0.70
145	Burns	1 121	53	0.73	0.54
146	Poisoning by psychotropic agents, drugs, or other medications	8 527	619	0.65	0.61
147	Other injuries and conditions due to external causes	2 687	215	0.70	0.61
148	Syncope	13 265	951	0.63	0.60
149	Fever of unknown origin	7 231	1 046	0.64	0.61
150	Lymphadenitis and gangrene	1 979	281	0.71	0.65
151	Shock	293	34	0.65	0.48
152	Nausea and vomiting	3 907	616	0.62	0.56
153	Abdominal pain	15 198	1 670	0.59	0.57
154	Malaise and fatigue	3 727	403	0.66	0.61
155	Allergic reactions	2 781	264	0.74	0.69
156	Rehabilitation and other aftercare, medical examination/evaluation/screening	38 097	2 609	0.62	0.60
157	Residual codes; unclassified	18 519	1 273	0.71	0.70

[*)] The diagnosis group numbers refer to the file 'Classification of variables' published together with the HSMR 2015 methodological report (see Van der Laan *et al*., 2016). In this file, the CCS-groups and corresponding ICD10-codes of the 157 diagnosis groups used for the HSMR are given. For the readmission ratio only 122 of these groups are used, but the numbering was kept the same.

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2016–2017	2016 to 2017 inclusive
2016/2017	Average for 2016 to 2017 inclusive
2016/'17	Crop year, financial year, school year, etc., beginning in 2016 and ending in 2017
2014/'15–2016/'17	Crop year, financial year, etc., 2014/'15 to 2016/'17 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
CCN Creatie, The Hague

Design
Edenspiekermann

Information
Telephone +31 88 570 7070
Via contact form: www.cbs.nl/information