PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Are Medical Record Front Page Data Suitable for Risk Adjustment in Hospital Performance Measurement: Development and Validation of a Risk Model of In-hospital Mortality after Acute Myocardial Infarction
AUTHORS	Wu, Chaoqun; Zhang, Danwei; Bai, Xueke; Zhou, Tiannan; Wang, Yongfei; Lin, Zhenqiu; He, Guangda; Li, Xi

VERSION 1 – REVIEW

REVIEWER	Nin-Chieh Hsu National Taiwan University Hospital, Taiwan
REVIEW RETURNED	31-Oct-2020

GENERAL COMMENTS	This study is a secondary analysis from a nationwide representative cohort of AMI, named China PEACE-Retrospective AMI study. The most interesting part of this dataset is that it was not a claims-based data. It was abstracted from the original medical records of hospitals by trained personnel. Major comments: 1. The content in the section of "Strength and limitations of this study" is somewhat irrelevant. Readers may expect to see the robust findings of this study (the strength) and the unanswered or potential biases of this study (the limitation) rather than a concise abstract. 2. An important limitation of this secondary analysis is the data abstraction process from the medical record, which warrants mentioning in the limitation section. 3. It seems lucky to have a retrospective dataset of AMI to do model development. However, for a quality monitoring purpose, this data abstraction process is difficult to apply to other diseases because it it labor intensive. Direct data extraction form electronic medical records may be more efficient for risk adjustment. Minor comments: 1. Table 1: Does "Cardiac shock" indicate "Cardiogenic shock?" 2. Figure S1 depicting the flowchart of this study which is essential for readers to understand the framework. Please consider to move to the main text. On the contrary, figure 2 or figure 3 may move to the supplement as these data can be described in text. Presenting the ROC curve and correlation plot is less helpful for readers.

REVIEWER	Thomas Sehested Department of Medicine, Zealand University Hospital, Roskilde, Denmark
REVIEW RETURNED	09-Nov-2020

GENERAL COMMENTS	In this study by Wu et al., the aim was to develop a model of in-hospital mortality using medical record front page data for patients hospitalized with acute myocardial infarction. They used data from 161 Chinese in hospitals in 2011 to develop their model and validated the model in 156 hospitals in 2015. The variables included in the model was demographic characteristics (gender, age, medical insurance status, ethnicity, marital status), admission department, diagnosis at admission (cardiac arrest) and at discharge (STEMI, infarction position, hypertension, diabetes, dyslipidaemia, cardiogenic shock, heart failure, stroke, renal failure). Model performance was good with discrimination (AUC of 0.78). This study appears to be a well-performed and with appropriate methodology. However, I do have some questions for the authors: It is unclear to me what the clinical / or other implications is for this prediction model. The authors say state that it could be suitable for risk adjustment in hospital performance measurement. Please elaborate Is there any application of your model in countries outside of China? In a clinical perspective, I think it would be relevant to compare this model to established models like the GRACE ACS risk score. It is unclear to me when prediction with your model can be used to estimate in-hospital mortality since discharge diagnosis is used in the model. In a way, you are predicting mortality using information from a future state if the idea is to estimate the risk at admission. The validation is done by calendar split of data (2011 vs 2015). It could be relevant to test the model performance random data split approach.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:
Major comments:

1. The content in the section of "Strength and limitations of this study" is somewhat irrelevant. Readers may expect to see the robust findings of this study (the strength) and the unanswered or potential biases of this study (the limitation) rather than a concise abstract.

Response: We appreciate this comment. We rewrote the "Strength and limitations of this study" section in this revision (Page 4).

- The analysis was based on a nationally representative cohort of hospitals in China, from which random samples of patients admitted with AMI was drawn to represent the heterogeneity in outcome of care.
- We used hierarchical generalized linear models that fully consider the patient clustering in hospitals, and is able to distinguish the differences within and between hospitals, which suits the purpose to adjust for case-mix in hospital performance comparison.
- We validated the finding that concise data extracted from medical record front page are good enough to reflect patients' risk profile, using the data from a closer year.
- External validations that include more diverse hospitals and among other diseases will be needed in the future.

2. An important limitation of this secondary analysis is the data abstraction process from the medical record, which warrants mentioning in the limitation section.

Response: We assumed the “limitation” refer to that abstracting complete medical record data is time-consuming and labor-intensive, thus not applicable for nationwide quality measurement. We fully agree with this. It is exactly the major question we want to solve in this study. We leveraged the data abstracted from complete medical records in a nationally representative sample of patients admitted with acute myocardial infarction, and proved that concise data extracted from medical record front page are good enough to reflect patients’ risk profile, which makes it suitable to adjust for case-mix in outcome measurement at hospital-level. This is of particular value,ince complete medical record data are rare, but medical record front page data are widely available in a mandated registry – the Hospital Quality Monitoring System (HQMS) in China, thus it is feasible to establish a nationwide hospital quality monitoring system based on the HQMS. We emphasized that in this revision (Page 5, Line 13).

Assessing quality of care between hospitals needs to take into account patients’ different demographic and clinical characteristics of patients between hospitals, like most of the prior studies have done based on a broad array of information from complete medical record. However, it is still unclear whether the MRFP data collected in HQMS can act as good surrogates for complete medical record model in estimation of risk-standardized mortality.

(Page 14, Line 9)

The feasibility of MRFP model has significant policy implications for China, as the government emphasized the importance of hospital performance monitoring. China needs a nationwide data platform, which supports timely, accurate and sustainable outcome measurement, since the outcomes of care such as mortality provide a global assessment of quality and have the most relevance to patients. However, outcome measurement is challenging, because of variation among hospitals in patients’ risk profile, meanwhile extracting data from electronic medical records is infeasible in most hospitals. T. Our firstly proved that concise medical record front page data that are available in the HQMS can sufficiently reflect patients’ risk profile, which makes it suitable to generating risk-standardized mortality rates at hospital-level. Thus, this existing platform covering 1800 (73%) tertiary hospitals and 2300 (26%) secondary hospitals can serve as a base for national hospital performance measurement, similar to the United States Centers for Medicare & Medicaid Services’ use of administrative claims data.

3. It seems lucky to have a retrospective dataset of AMI to do model development. However, for a quality monitoring purpose, this data abstraction process is difficult to apply to other diseases because it labor intensive. Direct data extraction form electronic medical records may be more efficient for risk adjustment.

Response: We appreciate this comment. As we elaborated above, this is the major purpose of our study. We agree that direct data extraction form electronic medical records are more efficient for risk adjustment in quality monitoring. However, in China, there is nosuch nationwide data system or database now. In this study, data abstracted from the complete medical record of a nationally representative are very valuable – lucky for us as the reviewer commented – because it provided a perfect condition to validate whether concise data from medical record front page only, which is currently available in the national HQMS database, could be used to adjusted risk of patients in hospital-level quality monitoring. We emphasized that in this revision (Page 5, Line 13).

Assessing quality of care between hospitals needs to take into account patients' different demographic and clinical characteristics of patients between hospitals, like most of the prior studies have done based on a broad array of information from complete medical record. However, it is still unclear whether the MRFP data collected in HQMS can act as good surrogates for complete medical record model in estimation of risk-standardized mortality.

(Page 14, Line 9)

The feasibility of MRFP model has significant policy implications for China, as the government emphasized the importance of hospital performance monitoring. China needs a nationwide data platform, which supports timely, accurate and sustainable outcome measurement, since the outcomes of care such as mortality provide a global assessment of quality and have the most relevance to patients. However, outcome measurement is challenging, because of variation among hospitals in patients' risk profile, meanwhile extracting data from electronic medical records is infeasible in most hospitals. T. Our firstly proved that concise medical record front page data that are available in the HQMS can sufficiently reflect patients' risk profile, which makes it suitable to generating risk-standardized mortality rates at hospital-level. Thus, this existing platform covering 1800 (73%) tertiary hospitals and 2300 (26%) secondary hospitals can serve as a base for national hospital performance measurement, similar to the United States Centers for Medicare & Medicaid Services' use of administrative claims data.

Minor comments:

1. Table 1: Does "Cardiac shock" indicate "Cardiogenic shock?"

Response: Yes. We revised the text and tables using 'Cardiogenic shock'.

2. Figure S1 depicting the flowchart of this study which is essential for readers to understand the framework. Please consider to move to the main text. On the contrary, figure 2 or figure 3 may move to the supplement as these data can be described in text. Presenting the ROC curve and correlation plot is less helpful for readers.

Response: We agree with the value of Figure S1. We moved
it to the main manuscript. We also moved Figure 2 to the supplement as recommended. However, we believe Figure 3 could provide readers useful information – the consistence between medical record front page model and complete medical record model in risk adjustment. Thus we prefer to keep them. We defer to the editor's opinion about the Figure 3.

Reviewer #2:

1. It is unclear to me what the clinical / or other implications is for this prediction model. The authors say state that it could be suitable for risk adjustment in hospital performance measurement. Please elaborate.

Response: We appreciate this opportunity to elaborate the implications of this model.
In hospital quality monitoring, patient outcomes are key measures, as they provide a more global assessment of healthcare quality and have the most relevance to patients. However, outcome measurement is challenging, because of variation among hospitals in the risk profile of their patients. Thus risk adjustment is essential for the outcome measurement.
In this study, we aim to prove that concise data extracted from medical record front page are good enough to reflect patients' risk profile, which makes it suitable to adjust for case-mix in outcome (i.e. in-hospital mortality) measurement at hospital-level. This is of particular value, since complete medical

record data are rare, but medical record front page data are widely available in a mandated registry – the Hospital Quality Monitoring System (HQMS) in China, thus it is feasible to establish a nationwide hospital quality monitoring system based on the HQMS, which could compare risk-standardized mortality rates among thousands of hospitals.
Thus, we leveraged the data abstracted from complete medical records in a nationally representative sample of patients admitted with acute myocardial infarction, and established two models – one was based on only medical record front page data, and the other was based on complete medical record data. The results demonstrated that risk-standardized mortality rates from the two models are very close and highly correlated, indicating that medical record front page data could be suitable for risk adjustment in hospital performance measurement in China. We emphasized that in this revision (Page 5, Line 13).

Assessing quality of care between hospitals needs to take into account patients' different demographic and clinical characteristics of patients between hospitals, like most of the prior studies have done based on a broad array of information from complete medical record. However, it is still unclear whether the MRFP data collected in HQMS can act as good surrogates for complete medical record model in estimation of risk-standardized mortality.

(Page 14, Line 9)

The feasibility of MRFP model has significant policy implications for China, as the government emphasized the importance of hospital performance monitoring. China needs a nationwide data platform, which supports timely, accurate and sustainable outcome measurement, since the outcomes of care such as mortality provide a global assessment of quality and have the most relevance to patients. However, outcome measurement is challenging, because of variation among hospitals in patients' risk profile, meanwhile extracting data from electronic medical records is infeasible in most hospitals. T. Our firstly proved that concise medical record front page data that are available in the HQMS can sufficiently reflect patients' risk profile, which makes it suitable to generating risk-standardized mortality rates at hospital-level. Thus, this existing platform covering 1800 (73%) tertiary hospitals and 2300 (26%) secondary hospitals can serve as a base for national hospital performance measurement, similar to the United States Centers for Medicare & Medicaid Services' use of administrative claims data.

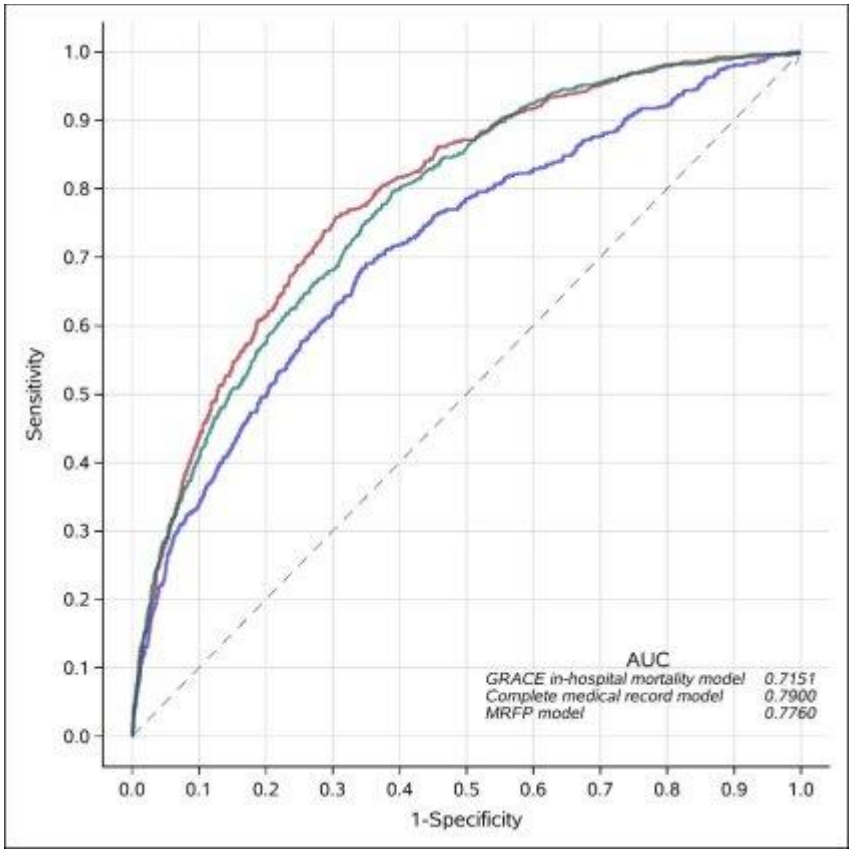
2. Is there any application of your model in countries outside of China?

Response: Our model was newly developed and validated in the current study, and has not been applied in other countries. However, similar researches had been conducted in other countries, such as the United States (Krumholz HM, et al. Circulation 2006;113(13):1683-92.). In the United States, the Centers for Medicare & Medicaid Services found it was challenging to measure outcome in hospital quality monitoring, because of variation among institutions in the risk profile of their patients. However, the only nationally available data on hospitalizations in the United States are Medicare claims, which do not fully reflect the information in the medical record, but only a concise database mainly extracted from the medical record front page. Thus, researchers did a similar study with ours, and proved that an span style="font-family:'Times New Roman"'>administrative claims-based model produces estimates of risk-standardized mortality that are good surrogates for estimates from a medical record model. Now the hospital performance measurement based on administrative claims-based model is routinely conducted in the United States.

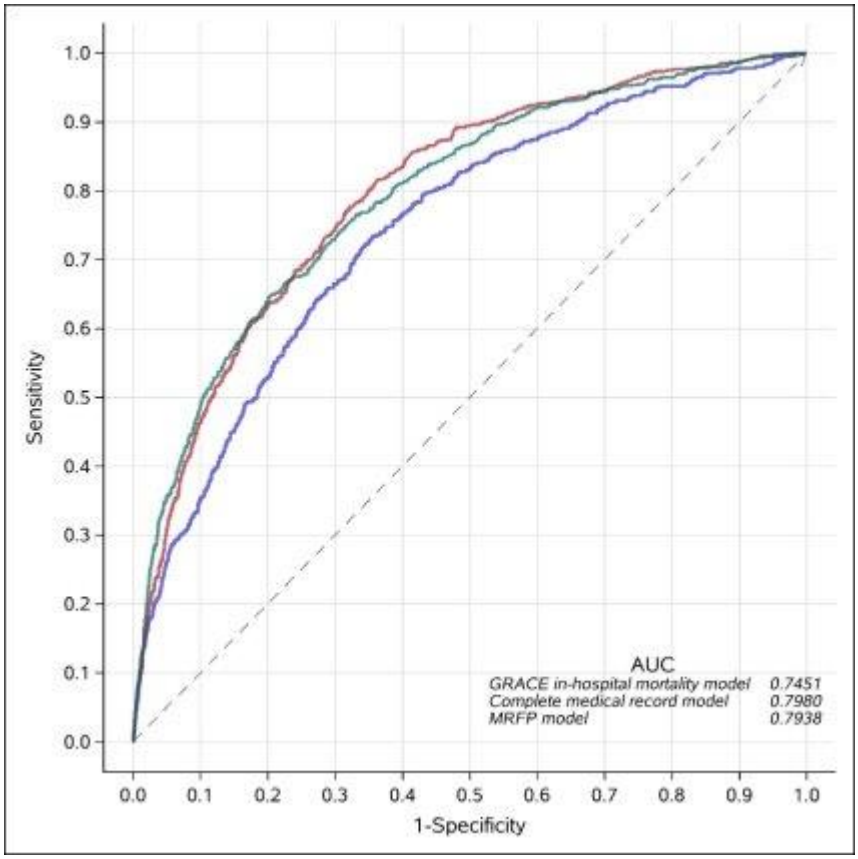
3. In a clinical perspective, I think it would be relevant to compare this model to established models like the GRACE ACS risk score.

Response: Thanks for this suggestion. We compared our patient prediction model with GRACE ACS in-hospital mortality model in this revision, and added the results in the appendix. The results further showed that our models were suitable for predicting in-hospital deaths in our population.

We calculated the predicted probability of in-hospital death among our development and validation cohorts by 3 models (GRACE ACS model, our complete medical record model and medical record front page model), separately. Then we calculated the AUC of each model and test the statistical difference between GRACE and our models. Integrated discrimination improvement(IDI) was also calculated to evaluated the overall improvement of out models compared with GRACE. Results showed that compared with GRACE ACS model, both our two patient risk prediction model had better AUC (all p value<0.001, see bellowing Figure1) and positive IDI among development and validation cohorts. In detail, the IDI of medical record front page model compared with GRACE model was 0.010 (0.003,0.017) in development cohort, and 0.028 (0.021,0.036) in validation cohort.



(a)



(b)

Figure 1. ROC of 3 models among development(a) and validation(b) cohort.

4. It is unclear to me when prediction with your model can be used to estimate in-hospital mortality since discharge diagnosis is used in the model. In a way, you are predicting mortality using information from a future state if the idea is to estimate the risk at admission.

Response: Thanks for this comment. We would like to clarify that the purpose of our model was not to predict patients' mortality, but to adjust for case-mix for hospital-level risk standardized mortality rates. Thus, we chose discharge diagnosis, instead of admission diagnosis for two reasons. First, discharge diagnosis could provide more complete and accurate information about the risk profile like medical history and comorbidities. Second, admission diagnosis is not available in the HQMS data platform, which is the expected application scenario of our findings.

5. The validation is done by calendar split of data (2011 vs 2015). It could be relevant to test the model performance random data split approach.

Response: We appreciate this comment, and agree that random data split is a routine approach for model validation. However considering the study aim that was to prove concise data extracted from medical record front page are good enough for hospital-level risk adjustment, we believe in the validation of this study it is more relevant to demonstrate that building a model using medical record front page data to reflect patients' risk is feasible, rather than show the robustness of specific coefficients in the model, Thus, we had chosen to, and would rather leave the data split approach unchanged – using data from year 2015 as a prospective validation, because it is closer to current status and future application scenarios.
In response to the reviewer's suggestion, we also applied the random data split approach – we pooled the patients from 2011 and 2015, then randomly chose half of them

to build hospital level models, and used the other half for validation. We provided the relevant results below, and defer to the reviewer opinion whether to add them in the appendix.
A total of 9037 cases (4233 from year 2011 cohort and 4804 from year 2015 cohort) were chosen as development cohort, and the others were as validation cohort. Gender, medical insurance status, admission department, age, MI position, dyslipidemia, cardiogenic shock, heart failure, stroke and renal failure were included in the patient-level and hospital-level HGLM model based on medical record front page data. The AUC of patient-level model was 0.8, and the slope of calibration curve was 0.928(intercept:0.005) (see bellowing Figure2).

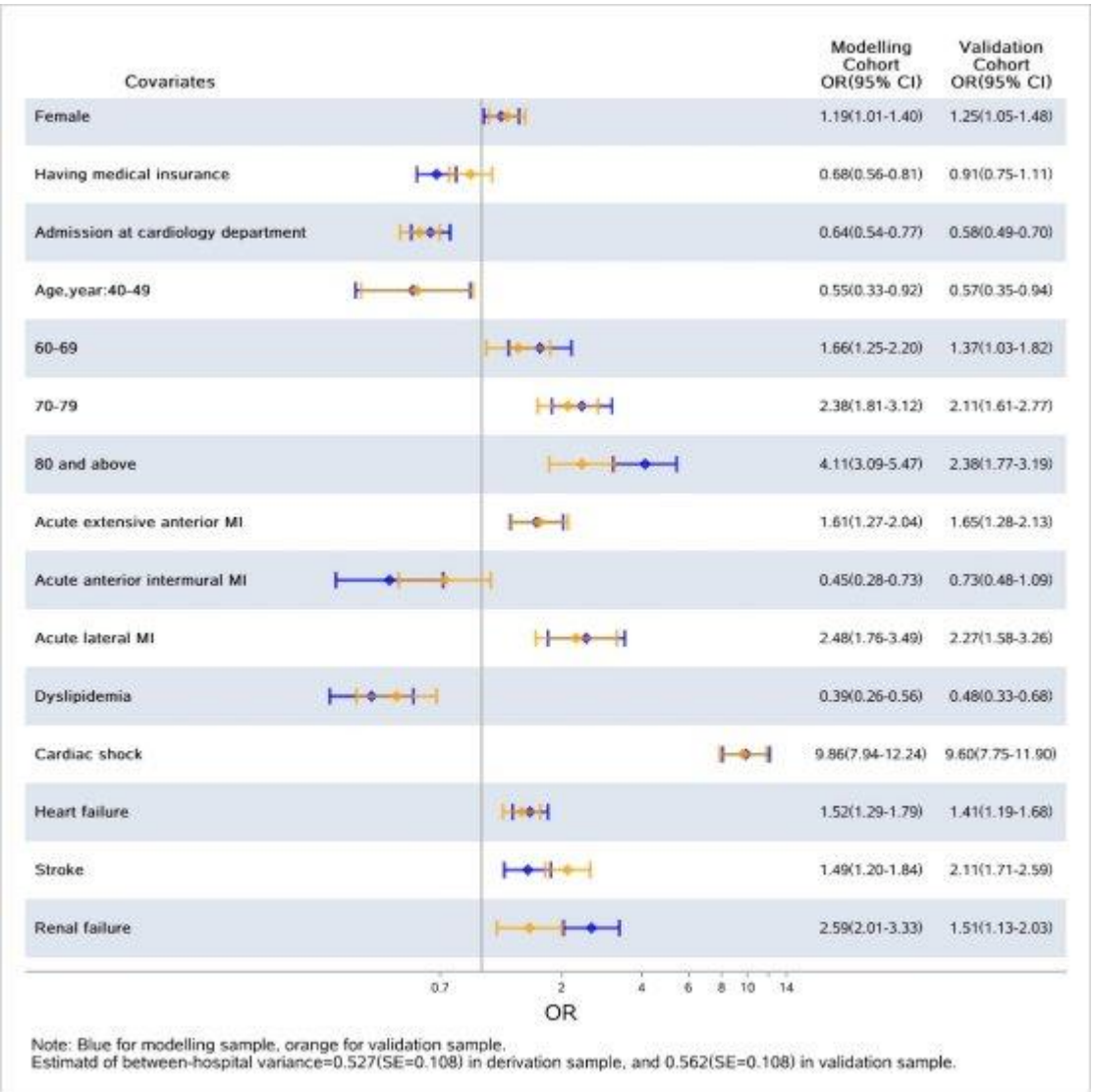
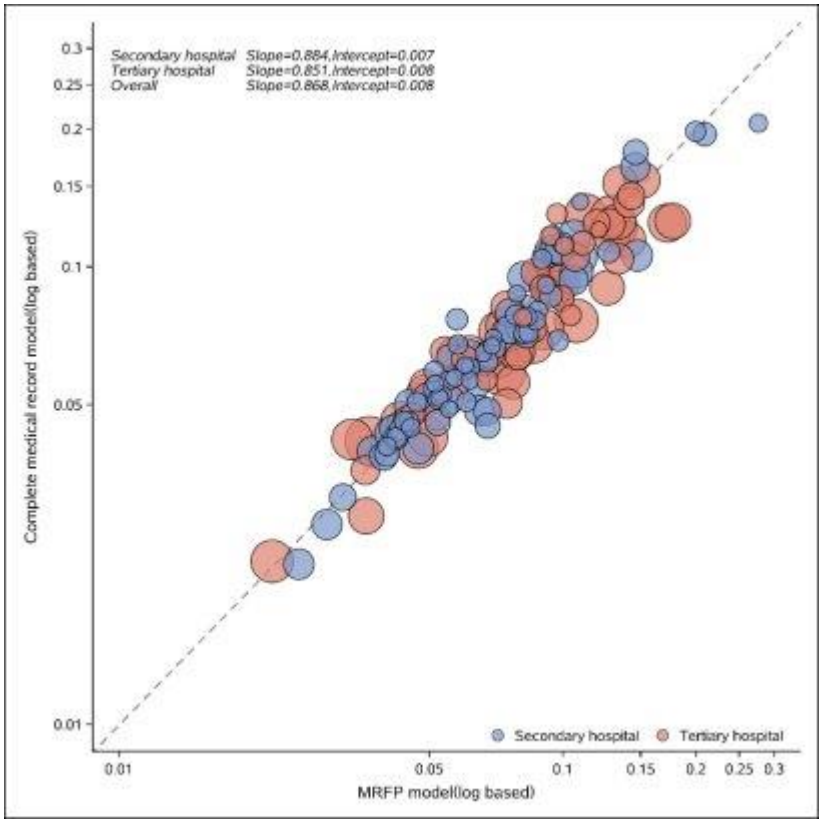


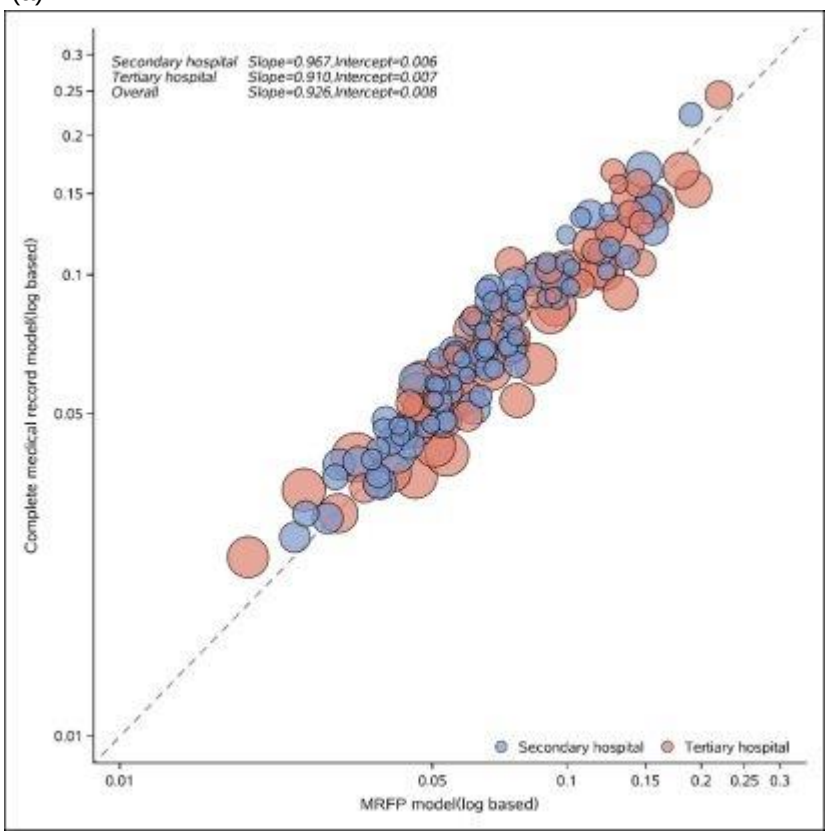
Figure 2. Odds ratios of medical record front page model among development and validation cohort by random data split approach.

For the hospital-level models, the median of absolute difference between the RSMR predicted by the complete medical record data and medical record front page data was less than 0.1% (IQR: -0.76% - 0.29%), and the 10th and 90th percentiles were -1.8% and 1.1%. For the regression model comparing the RSMR between the medical record front page data and complete medical record data, the slope (intercept) was 0.87 (0.008). In validation cohort, the median of absolute difference between the RSMR predicted by the complete medical record data and medical record front page data was less than 0.3% (IQR: -0.41% - 0.87%), and the 10th and 90th percentiles were -1.3% and 1.7%.

For the regression model comparing the RSMR between the medical record front page data and complete medical record data, the slope (intercept) was 0.93(0.008) (see bellowing Figure3).



(a)



(b)

Figure 3. Correlation of risk standardized mortality rate estimated by medical record front page model and complete medical record model among development(a) and validation(b) cohort by random data split approach.

Similar results using random data split approach proved the stability of our research.