The impact of morbidity trajectories on identifying high-cost cases: using Taiwan’s National Health Insurance as an example

Hsien-Yen Chang

Department of Health Policy & Management, Bloomberg School of Public Health, Johns Hopkins University, 624 N. Broadway, Baltimore, MD 21205, USA

Address for correspondence: Hsien-Yen Chang, E-mail: hchang2@jhsph.edu

ABSTRACT

Background Incorporating longitudinal information into risk-adjustment models has been considered important. This study aimed to evaluate how morbidity trajectories impact risk-adjustment models in identifying high-cost cases.

Methods Claims-based risk adjusters, with or without morbidity trajectories derived from 3-year claims from Taiwan’s National Insurance System, were used to predict being a prospective high-cost user. A random sample of Taiwanese National Health Insurance enrollees continuously enrolled from 2002 to 2005 (n = 147,892) was the study sample. A logistic regression model was employed. The performance measures, based on the split analysis, included statistical indicators (c-statistics, sensitivity and predictive positive value), proportions of true cases identified by models and medical utilization of predicted cases.

Results As the comprehensiveness of risk adjustment models increased, the performance of the models generally increased. The effect of adding trajectories on the model performance decreased as the comprehensiveness of the model increased. Such impact was most apparent in statistical indicators and medical utilization of the predicted groups.

Conclusions In identifying high-cost cases, adding morbidity trajectories might be necessary only for less comprehensive risk adjustment models, and its contributions came from higher c-statistics and increasing medical utilization of predicted groups.

Keywords adjusted clinical group (ACG), longitudinal morbidity trajectories, predictive modeling, risk adjustment, Taiwan’s National Health Insurance

Introduction

Research has consistently shown that a large proportion of medical resources are consumed by a small percentage of the total population. For example, the top 20% of the population accounted for ~80% of all healthcare expenditures across different countries. It is hoped that by helping these few people manage and coordinate their medical needs, the expansion of medical expenditures may be controlled; therefore, this group of high users of medical resources has been consistently targeted by different cost-containment programs, such as disease management, care management and utilization review. In addition, with limited resources, healthcare plans cannot possibly enrol all eligible members into these programs; only a small fraction of eligible members could be selected. These programs have been widely applied across different clinical areas, such as heart failure patients, diabetic patients, or pregnant women, and several studies have shown that implementing these programs might help reduce medical utilization and increase the quality of care.

Predictive modeling is a process of applying existing patient data to prospectively identify such persons with high medical needs who are at risk for higher future medical utilization, and is one of the major applications of risk adjustment models. Disease and case management programs often rely on this technique to identify the best targets for intervention. The input for predictive modeling usually comes from information available on claims data, and the performance of these models has been studied comprehensively. However, the major problem with predict
modeling is the non-overlap of the high-risk groups identified by different risk adjustment models; several solutions have been proposed, and incorporating longitudinal elements of morbidity information into risk-adjustment models has always been considered a potentially important next step.\textsuperscript{18,19}

Taiwan launched a government-run, single-payer National Health Insurance (NHI) program in May 1995. All Taiwanese nationals are obligated by law to join this program to ensure adequate risk pooling. The NHI’s benefit packages are comprehensive, including inpatient and outpatient services, pharmacy services and Chinese medicine and dental services. Beneficiaries have complete freedom of choice of providers and therapies, and they do not need to go through ‘gatekeepers’ in order to obtain medical services from specialists. The primary source of funding for the NHI is the payment of premiums shared by the insured, the employers and the government. In terms of reimbursement, the global budget payment system was adopted in order to contain the growth of medical expenditure, and the NHI-contracted healthcare providers have to submit their claims electronically.

Morbidity trajectories, the progress of morbidity burden over time, and its ability to explain prospective costs have been evaluated in a prior study;\textsuperscript{20} nonetheless, the impact of morbidity trajectories on identifying high-cost cases remains unanswered. In this study we created morbidity trajectories based on 3 years of claims data, and evaluated the impact of adding morbidity trajectories on the ability of risk-adjustment models in identifying high-cost cases.

Methods

Data sources
The source of the data was a longitudinal data set prepared by Taiwan’s Bureau of NHI. It contained enrollment and claims files from a 1% random sample of Taiwan’s population (\textasciitilde 200 000 individuals). The enrollment files contained individual subscription information and demographic factors, including sex, date of birth and type of beneficiary (insured or dependent). The claims files contained comprehensive records of utilization of inpatient care, ambulatory care, pharmacy store, dental care and Chinese medicine, including date of service, ICD-9-CM diagnosis codes, claimed medical expenses and amount of co-payment for each encounter. At least 48 months enrollment from year 2002 to year 2005 was required to be included in this analysis, resulting in the final sample size of 147 892.

An annual total cost was aggregated from all inpatient, outpatient and pharmacy store claims for every enrollee, including out-of-pocket claimed reimbursement, medication expenses and co-payments; expenses for dental care and Chinese medicine were excluded from this aggregation. The unit of money is New Taiwan Dollars (NTDs); the exchange rate was 32.17 NTDs: 1 USD in 2005. The individuals’ identifiers were encrypted to protect privacy and confidentiality, and this study was approved by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board.

ACG risk adjustment system
Various risk adjusters derived from the Johns Hopkins Adjusted Clinical Group (ACG) system were used in this study, including ACGs, ADGs (Aggregated Diagnosis Groups), EDCs (Expanded Diagnosis Clusters) and ACG-PM scores. The ACG system assigns all ICD-9-CM codes to one of 32 aggregated diagnostic clusters (ADGs).\textsuperscript{21,22} Each ADG is a grouping of diagnosis codes similar in terms of severity and likelihood of persistence of the health condition treated over a relevant period; ADGs are not mutually exclusive, and individuals can have up to 32 ADGs. Individuals are then placed into 1 of 93 discrete ACG categories according to their assigned ADGs, age and sex. EDCs are binary indicators to show whether an individual has specific diseases/symptoms.\textsuperscript{15}

ACG predictive modeling (ACG-PM) makes use of morbidity metrics available within the ACGs system to prospectively identify persons with high medical needs at risk for higher future medical utilization.\textsuperscript{15} Those included in ACG-PM as risk factors are age groups, sex, selected ACGs, a pregnancy without delivery indicator, hospital dominant markers, a medically fragile indicator and specific EDC disease markers; pharmacy expenses are optional. A standardized predicted resource index expresses anticipated resource use as a relative value, with the mean rescaled to one; higher value represents higher morbidity burden and thus higher medical utilization.

Assignment of morbidity trajectories
Morbidity trajectories were constructed using diagnosis information from 2002 to 2004. An individual was assigned to a ‘low’ (<0.5), ‘medium’ (0.5–1.5) or ‘high’ (>1.5) morbidity level in each year based on his/her standardized predicted resource index. An individual’s change in three-level morbidity burden over 3 years was assigned to one of the following six trajectory groups, using the classification system applied in prior research: constant-high, constant-medium, constant-low, decreasing, increasing and erratic (Supplementary data, Appendix A).\textsuperscript{20}

Risk-adjustment models
Five risk-adjustment models were used, both with and without the measures of the person’s morbidity trajectory. All
diagnosis and cost information in these models was derived from year 2004. These five risk-adjustment models were as follows:

(i) Demographics only
(ii) ACGs with demographics
(iii) ADGs plus selected EDCs with demographics
(iv) Prior total cost with demographics and
(v) Prior total cost, ADGs and selected EDCs with demographics.

Other than Model 4, each model was built upon its previous model (Supplementary data, Appendix B). Model 3 is a broader model not only because it additionally contains selected EDCs, but also because these 32 ADGs are the basis to construct the ACGs and they are not mutually exclusive.21,22 Demographic factors included sex, categorical age, type of beneficiary (the insured or the dependent) and insurance category (based on insured's type of job).23 Selected EDCs were the result of stepwise analyses in explaining prospective expenditures using all EDCs codes; 19 EDCs were chosen (Supplementary data, Appendix C).

Outcomes and measures of model performance
Two sets of thresholds were used to define being a high-cost user: top 1 and 5% users of total costs in 2005. These cut-off points were chosen because these would provide different options for the ‘manageable’ amount of targets (high-cost users) for intervention and are commonly used in different studies.4,17,24 We applied a logistic regression model given the dichotomous outcomes. We conducted all statistical analyses using SAS® software version 9.1. Performance of five risk adjustment models was evaluated from three aspects: statistical indicators, proportions of true cases identified by models and medical utilization of predicted cases. Statistical indicators included c-statistics, sensitivity and predictive positive value (PPV),1 and the thresholds for calculating statistical indicators were set as the corresponding levels of outcomes. The c-statistic represents the area under the receiver operating characteristic curve, and hence provides an overall measure of model performance.

Actual 2005 top users were assigned to one of four mutually exclusive categories (Supplementary data, Appendix D): in the 2004 top user group alone (area a), in predicted top user group identified by risk adjustment models alone (area a), in both groups (area b) or in neither group. The real contribution of risk adjustment models comes from those identified by models alone, because these subjects may not be known without applying risk adjustment models (Supplementary data, area a in Appendix D). The last approach to assess the model performance is to examine the medical utilization of the identified top groups, including 2005 average total costs and the trend ratio of total costs (2005 cost divided by 2004 cost in the top identified group). A better model will have higher total costs, and higher trend ratio of costs (subjects with costs increasing over time are better targets for intervention).

The focus of this study was not so much in the model performance itself, but the change of these performance measures from the model without morbidity trajectories (baseline model) to the same model with morbidity trajectories (advanced model). If adding morbidity trajectories is effective, we would expect to see an increase in all three statistical indicators, the proportion of the predicted top user group identified by risk adjustment models alone (area a) and two medical utilization measures of predicted cases. Split analysis (a randomly selected 70% of study subjects were used for model development; the rest were set aside for model validation) was performed and measures of model performance were obtained from the validation set to avoid overfitting.

Results
Characteristics of subjects
The total sample size was 147,892 while the validation set contained 44,368 subjects. About 48% of subjects were male; the mean age was ~38 years old. About 90% of subjects had non-zero drug or total cost. The average 2005 total/pharmacy cost was 12,599/5110 NTDs. The proportion of the subjects in these six morbidity trajectories were: 25.1% in ‘constant-low’, 14.0% in ‘constant-medium’, 8.7% in ‘constant-high’, 17.3% in ‘decreasing’, 18.6% in ‘increasing’ group and 16.3% in ‘erratic’ group (Table 1).

Statistical indicators
Across both thresholds, all three statistical indicators increased as the comprehensiveness of the model increased, with or without morbidity trajectories; such increases were most apparent in sensitivity and PPV. For example, at the top 1% baseline model, the increase was from 0.78 in Model 1 to 0.92 in Model 5 for the c-statistics, from 0.06 to 0.43 for sensitivity and from 0.05 to 0.44 for PPV (Table 2). The incremental change was only larger in Model 1, at most 0.03 in Model 2, while <0.01 in other models. Across both thresholds, the incremental change decreased as the comprehensiveness of the model increased. This suggested that adding morbidity trajectories only impacted the performance of the demographics model on statistical indicators.
In this section, we present results from three risk adjustment models without prior costs (Table 3 and Supplementary data, Appendix D). Across both thresholds, between 45 and 50% of actual 2005 top users could be identified by prior user status (Supplementary data, area b+area c in Appendix D); in less comprehensive models (Models 1 and 2), the majority of these users could be mostly identified by prior user status alone (area c); however, in the ADGs model, the majority could be identified by both prior user status and ADGs model (area b).

Across both thresholds, the proportion of actual top users that could be identified by risk adjustment models (area a+area b in Appendix D) increased when the comprehensiveness of the model increased; however, the majority could still be identified by prior user status, especially in the comprehensive model (Model 3). The proportion of actual top users identified by risk adjustment models alone (area a in Appendix D) increased or remained similar as the comprehensiveness of the model increased.

Adding morbidity trajectories generally increased the ability of the risk adjustment models to identify actual top users, especially for less comprehensive models at a more relaxed threshold.
threshold; for example, the incremental change in the proportion of actual top users that could be identified by the demographics model \((\text{area } a + \text{area } b)\) increased from 1.5 to 10% while the outcome expanded from 1 to 5%. As for the proportion identified by risk adjustment models alone (Supplementary data, area \(a\) in Appendix D), the incremental change was generally positive; however, the magnitude was small (at most 1.2%); for example, it ranged from 0.2% at top 1% level to 0.6% at top 5% level for the ADGs model (Model 3).

Medical utilization of predicted cases

Across both thresholds, total costs increased as the comprehensiveness of the model increased, with or without morbidity trajectories. For total costs, there was a huge jump between Models 2 and 3 and also between Models 1 and 2. The trend ratios of total costs were generally >1 for models without prior costs (Models 1–3). The trend ratios were the smallest in the prior cost model across three thresholds (0.7–0.9) (Table 4).

Adding trajectories generally increased total costs in 2005 across all thresholds and models; such incremental changes were largest in the least comprehensive model while smallest or even negative in models with prior costs. For example, at the top 1% outcome level the incremental change was 30,000 NTDs for the demographics model, while at the top 5% level the incremental change was −2,000 NTDs for the prior cost model (Model 4). Adding trajectories increased trend ratios at both top 1 and 5% outcome levels for all but the demographics model; for the demographics model the incremental change was negative, about −0.1. The magnitudes of the incremental

---

**Table 3** Proportion (%) of true cases identified by models and prior top users status

<table>
<thead>
<tr>
<th>Model 1: demographics</th>
<th>Model 2: ACGs and demographics</th>
<th>Model 3: ADGs, 19 EDCs and demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Among 2005 top 1% actual user (n = 452)</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline model performance (without morbidity trajectories groups)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In 2004 top</td>
<td>42.04</td>
<td>36.95</td>
</tr>
<tr>
<td>1% actual users only (area (c))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In top 1% predicted group only (area (a))</td>
<td>3.32</td>
<td>3.98</td>
</tr>
<tr>
<td>In both groups (area (b))</td>
<td>3.10</td>
<td>8.19</td>
</tr>
<tr>
<td>By risk adjustment model (area (a + area b))</td>
<td>6.42</td>
<td>12.17</td>
</tr>
<tr>
<td>Incremental change with the addition of morbidity trajectories groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In 2004 top</td>
<td>−2.22</td>
<td>−0.67</td>
</tr>
<tr>
<td>1% actual users only (area (c))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In top 1% predicted group only (area (a))</td>
<td>−0.67</td>
<td>0.44</td>
</tr>
<tr>
<td>In both groups (area (b))</td>
<td>2.21</td>
<td>0.66</td>
</tr>
<tr>
<td>By risk adjustment model (area (a + area b))</td>
<td>1.54</td>
<td>1.10</td>
</tr>
</tbody>
</table>

**Number of study sample (validation only):** 44,368.

<sup>a</sup>Please refer to Supplementary data, Appendix D for areas \(a\), \(b\) and \(c\).

---

**Table 3** Continued

<table>
<thead>
<tr>
<th>Model 1: demographics</th>
<th>Model 2: ACGs and demographics</th>
<th>Model 3: ADGs, 19 EDCs and demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>In top 1% predicted group only (area (a))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In both groups (area (b))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>By risk adjustment model (area (a + area b))</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>In top 1% predicted group only (area (a))</strong></td>
<td>9.67</td>
<td>2.52</td>
</tr>
<tr>
<td><strong>In both groups (area (b))</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>By risk adjustment model (area (a + area b))</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Continued**
changes decreased significantly as the comprehensiveness of the model increased (from 0.1 in Model 1 to 0.01 in Model 5).

Discussion

Main findings
As the comprehensiveness of risk adjustment models increased, the performance of the models generally increased, especially in statistical indicators and the mean costs of the predicted groups; the only exception was the trend ratio. The effect of adding trajectories on the model performance decreased as the comprehensiveness of the model increased. Such impact was most apparent in statistical indicators and the mean costs of the predicted groups, while ignorable or detrimental in trend ratio and proportions of true cases identified by models.

What is already known on this topic
The application of risk adjustment models on identifying high-cost cases has been studied extensively, even using Taiwan’s NHI data. The results have been similar: the more comprehensive the model, the better the identification. However, the performance of incorporating the longitudinal element into risk adjustment models on identifying high-cost cases has never been explored.

What this study adds
Even though the outcome variables and implications were different, the results of this study were similar to what was found in a previous study: adding morbidity trajectories help the performance of the least comprehensive model (demographics model) the most and that of the most comprehensive model the least. In that prior research the outcome was continuous cost and its results were mainly relevant to payment adjustment, while in this current research the outcome was a binary indicator of being a high cost user and it has policy implications for identifying the best targets for disease/care management programs.

Morbidity trajectories were constructed based on predicted resource index, and hence to some extent reflected the cost to be used in the future. Therefore, when the baseline model did not include utilization measures, such as the demographics model, adding morbidity trajectories would help increase the model performance; on the contrary, when the baseline model included utilization measures, such as the prior cost model, the contribution from adding morbidity trajectories might become minimal due to the overlap of both variables. In addition, given that many healthcare expenditures are truly random and cannot be foreseen, an upper limit possibly exists on the risk adjustment model’s ability to identify high-cost users; the most comprehensive cross-sectional model might have reached the limit so that adding additional information may not have any help. Finally, the longitudinal information

<table>
<thead>
<tr>
<th>Outcome: 2005 top 1% actual user</th>
<th>Model 1: demographics</th>
<th>Model 2: ACGs and demographics</th>
<th>Model 3: ADGs, 19 EDCs and demographics</th>
<th>Model 4: 2004 Cost and demographics</th>
<th>Model 5: ADGs, 19 EDCs, 2004 cost and demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline model performance (without morbidity trajectories groups)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 Mean total cost</td>
<td>68 155</td>
<td>138 186</td>
<td>279 528</td>
<td>334 007</td>
<td>335 638</td>
</tr>
<tr>
<td>Trend ratio of total cost</td>
<td>1.18</td>
<td>0.98</td>
<td>1.00</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td>Incremental change with the addition of morbidity trajectories groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 Mean total cost</td>
<td>46 266</td>
<td>5530</td>
<td>8397</td>
<td>3009</td>
<td>2485</td>
</tr>
<tr>
<td>Trend ratio of total cost</td>
<td>~0.10</td>
<td>0.05</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Outcome: 2005 top 5% actual user</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline model performance (without morbidity trajectories groups)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 Mean total cost</td>
<td>63 344</td>
<td>95 416</td>
<td>120 946</td>
<td>138 548</td>
<td>138 237</td>
</tr>
<tr>
<td>Trend ratio of total cost</td>
<td>1.19</td>
<td>0.99</td>
<td>1.01</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>Incremental change with the addition of morbidity trajectories groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 Mean total cost</td>
<td>29 091</td>
<td>5028</td>
<td>1775</td>
<td>2140</td>
<td>660</td>
</tr>
<tr>
<td>Trend ratio of total cost</td>
<td>~0.11</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Number of study sample (validation only): 44 368.
might have been lost during the process of being collapsed into six levels so that adding trajectories did not help the performance of the model.

Models with prior costs had much better performance than diagnosis-based models in identifying high-cost users. This is probably due to the strong correlation between prior and current costs (Pearson’s correlation coefficient between 2004 and 2005 total costs: 0.64); therefore, we could find that ~50% of high-cost users in 2004 remained so in 2005. However, relying on prior costs might identify people with high costs by chance, such as those involved in car accidents; in addition, we also found out that the medical utilization of this group decreased largely over time from 2004 to 2005. Therefore, these are not good targets for intervention.

Taiwan is remarkable in that it is one of a very few countries where a central computerized data repository is available for almost 100% of the population across all medical services, ages and geographic areas, which made this study possible. For countries or healthcare systems with longitudinal data, such as the Medicare system in the USA, National Health Service in the UK or Japan’s NHI, this study suggests that the longitudinal element can be added to the simple demographics factors to identify the best targets for disease/care management programs; such longitudinal element can be a very simple classification of a person’s morbidity over years. However, for other nations without such data, it would be necessary to implement comprehensive cross-sectional risk adjustment.

Limitations
This study had several limitations. First, continuous enrollment was required so that results could not be generalized to those without continuous enrollment. Secondly, diagnosis information from claims data was used so that the completeness of claims data might affect the findings; in addition, pharmacy data was available and could be incorporated for future research. Thirdly, the cutoff points for the morbidity level were arbitrarily chosen; however, in another longitudinal study, the similarity of results from Taiwan’s NHI and Canadian claims data showed that the differences in selection of cut-off points and the time period might not affect the principal findings.

Conclusions
In identifying high-cost cases, adding morbidity trajectories might only be necessary for less comprehensive risk adjustment models, and its contributions came from higher c-statistics (c-statistics, sensitivity and PPV), and increasing medical utilization of predicted groups. The applications and implications of morbidity trajectory groups in other aspects should be further investigated.

Supplementary data
Supplementary data are available at the Journal of Public Health online.

Acknowledgements
The author would like to thank Dr Weng-Foung Huang, a professor at National Yang-Ming University, Taipei, Taiwan, for his collaboration in this project. In addition, this study is based in part on data from the National Health Insurance Research Database provided by the Bureau of National Health Insurance, Department of Health and managed by National Health Research Institutes in Taiwan. The interpretation and conclusions contained herein do not represent those of Bureau of National Health Insurance, Department of Health or National Health Research Institutes in Taiwan.

References


