USING FINANCIAL MODELLING FOR INTEGRATED HEALTHCARE DATABASES

Alexei Botchkarev
GS Research & Consulting
Toronto, Ontario, Canada
e-mail: alex.bot@gsrc.ca

Peter Andru
Ministry of Health and Long-Term Care
Toronto, Ontario, Canada
e-mail: peter.andru@ontario.ca

ABSTRACT

Ontario (Canada) Health System stakeholders support the idea and necessity of the integrated source of data that would include both clinical (e.g. diagnosis, intervention, length of stay, case mix group) and financial (e.g. cost per weighted case, cost per diem) characteristics of the Ontario healthcare system activities at the patient-specific level. At present, the actual patient-level case costs in the explicit form are not available in the financial databases for all hospitals. The goal of this research effort is to develop financial models that will assign each clinical case in the patient-specific data warehouse a dollar value, representing the cost incurred by the Ontario health care facility which treated the patient. Several mathematical models have been developed and verified using real datasets showing feasibility of the selected approach and correctness of the models.

Index Terms — Databases, Healthcare, Modelling, Financial Data Processing, Case Cost

1. INTRODUCTION

I. BUSINESS RATIONALE

Ontario (Canada) Health System stakeholders are expecting high quality financial and clinical information for their planning and analysis purposes. All stakeholders support the idea and necessity of the consolidated, integrated and authoritative source of data and information that would include both clinical and financial characteristics of the Ontario healthcare system activities at the patient-specific level.

At the same time, generally speaking, the actual patient-level case costs in the explicit form are not available in the financial databases. Exception is for the cases submitted by the hospitals participating in the Ontario Case Costing Initiative (OCCI). However, only approximately 10% of the Ontario hospitals have been involved in the OCCI.

Financial/mathematical models should be developed in order to calculate estimates of case costs for all clinical cases information of which is stored in the Provincial Health Planning Database (PHPDB). Initial research performed by the Ontario Ministry of Health and Long-Term Care (MOHLTC) confirmed interest of the stakeholders and identified general approaches to the problem [1].

In plain language, the ultimate goal of this research effort is to assign each clinical case in the patient-specific PHPDB data warehouse a dollar value, representing the cost incurred by the Ontario health care facility which treated the patient. Several mathematical models have been developed and verified using real PHPDB and OCCI datasets.

Business scenarios were developed to describe the business purposes and requests to be addressed by the analysts with the help of the data based on the financial models. Analysts are working for the Ontario Ministry of Health and Long-Term Care (MOHLTC) or Local Health Integrated Networks (LHINs), and have a province-wide or LHIN-wide perspective. Their lowest level of data analysis is the patient discharge for the Discharge Abstract Database (DAD). They may wish to roll up or analyze information using the following dimensions: age, gender, CMG, diagnosis, procedure, outcome, facility, LHIN of facility, LHIN of residence, or other. They may wish to compare LHIN totals, facility totals, LHIN to LHIN, or facility to facility. They want to know reliable totals and significant differences in discharge costs.

Sample enquiry scenarios include:

An analyst in a LHIN wants to know the costs for hip replacement cases across all the facilities performing the procedure in that LHIN, or for all the facilities providing the service to residents of that LHIN.

An analyst in a LHIN wants to know the range of costs for cardiac cases across all the facilities performing the procedure within the province.

A ministry analyst wants to know the total costs for kidney transplant, failure, or rejection across facilities and LHINs over the last three years.

III. TERMS AND DEFINITIONS

Case. An instant of a disease that led to the individual's inpatient stay, which has been registered by the health service organization, reported to the Canadian Institute for Health Information (CIHI) DAD, database and eventually has a corresponding record with all appropriate attributes in the PHPDB.

Case Cost. Expenditures (direct and indirect) incurred by the health service facility relating to the treatment of a specific (patient-level) case. There's no way of "precise measuring" of the dollar value of each and all specific cases. Irrespective of the methodology and calculation techniques employed to determine the Case Cost, dollar value of a specific case is always an approximation of the "real-life" hospital expenses. In some sense, Case Cost is a conceptual phenomenon reflecting hospital business functioning as a clinical and financial entity.

Case Costing. A process of allocating expenditures of various hospital departments to each individual Case with an objective to determine the Case Cost. The process is performed according to the Ontario Cost Distribution Methodology (OCDM) and OCCI methodology. This project is not involved in the Case Costing process, its formal description, implementation or modification. Although, the project uses Case Costing results (Case Costs), and parameters employed by the Case Costing methodology. In general, case costing can be accomplished in various ways. Within the framework of this project, the term Case Costing is used only with the implication that the process is based on the OCDM/OCCI methodology.

Actual Case Cost (ACC). Case Cost that is considered to be the most trustworthy and most closely reflecting "real-life" hospital expenditures on a specific case is referred to as Actual Case Cost. Actual Case Costs are used as a benchmark for comparing various case costs.

Case Cost Model (CCM). A formal representation of the Case Cost, as a conceptual phenomenon, which involves mathematics, logical expressions, well-defined procedures, computer software, that is constructed with the purpose of producing output – Case Cost Estimate, as a function of one or more clinical and financial parameters. CCM can constitute a single formulae/equation or a set of complex algorithms implemented in a software package.

Case Cost Estimate (CCE). Approximate dollar value of the Case Cost, determined/calculated by employing the Cost Case Model. CCE may constitute a single value or a range of values.

Aggregate Actual Case Cost (AACC). The sum of two or more actual costs of cases (ACCs) usually produced as a result of a query with the underlying question such as: What was the actual cost of all cases of the XYZ hospital? What was the actual cost of all cases with Case Mix Group (CMG) code xxx?

Aggregate Case Cost Estimate (ACCE). The sum of two or more cost estimates of cases (CCEs) usually produced as a result of a query with the underlying question such as: What was the cost estimate of all cases of the XYZ hospital? What was the cost estimate of all cases with CMG code xxx?

Cost Modelling. 1. A process of applying structured methodology to create and validate Case Cost Models (CCM). 2. A process of employing CCM to produce Case Cost Estimate (CCE).

Financial Modelling. A complex process which involves people, data, equipment, software, methods, and includes a set of interrelated activities of: acquiring clinical and financial data; creating and validating Case Cost Models and employing them to produce Case Cost Estimates; extracting, transforming and loading data; performed to integrate clinical and financial data in the PHPDB and make it available to the healthcare analysts.

IV. DESCRIPTION OF THE DATA EXPERIMENT

A. *Objective*

Design and use case cost models to calculate case cost estimates (CCE) based on the DAD database patient-specific clinical data appended with the case cost parameters published by OCCI/OCDM. Compare CCE with the case costs (benchmark) submitted to the OCCI by participating hospitals. Analyze estimating errors to verify and fine-tune the models.

B. *Steps of the Data Experiment Process*

1. The initial source of the patient-specific clinical data was – Inpatient Discharge (DAD) FY 2004-2005 file [2]. Dataset pertaining to a large complex teaching hospital (LCTH) 2004/05 was used in the experiment.

Layout of the file was changed to preserve only fields/columns containing clinical information which could be considered most pertinent to determining case cost estimates (e.g. CMG, diagnoses, interventions, length of stay, etc). The file has 130 fields with clinical data. It contains 7782 patient-level cases grouped in 163 CMGs (the set of data was limited to 163 CMG codes with matching number of cases between DAD and OCCI CAT (Cost Analysis Tool – www.occp.com – Costing Reports).

2. Clinical information in the file was appended with the following OCDM cost data for the LCTH hospital [3]: Acute expenses; Acute Weighted Cases; Cost per Weighted Case; Acute Direct Cost Per Diem; Acute Overhead Cost Per Diem; Acute Total Cost Per Diem; Acute Total Patient Days.

3. Several (total of 12) financial models were developed and used one at a time to calculate CCE for each patient-level case.

4. Using CCEs and simulating the query: "What's the total cost of the LCTH CMG xyz cases in 2004-05?", aggregate cost estimates (and several other parameters) were calculated for each of the 163 CMG codes:

– Estimate of the Total cost of the cases in the CMG group
– Estimate of the Average cost of the case in the CMG group
– Estimate of the Standard deviation of the case costs in the CMG group
– Estimate of the Minimum case cost in the CMG group
– Estimate of the Maximum case cost in the CMG group

The reason for selecting and calculating estimates of the above parameters is that the same parameters are available from the OCCI CAT database [4].

5. OCCI CAT database was queried: "What are the case cost parameters of the LCTH CMG xyz cases in 2004-05?" for each of the 163 CMG codes. OCCI CAT returns the same case cost parameters as outlined in Step 4. The difference is that this is the official OCCI data, and it was used later as the Actual (correct, verified) cost.

6. Estimated (step 4) and actual (step 5) cost parameters were compared and estimating errors characterizing the efficiency of the financial models were calculated for each of the 163 CMG codes:

– Error estimating the Total cost of the cases in the CMG group

$$E\,(ACCE_{cmg}) = ACCE_{cmg} - AACC_{cmg}$$

– Error estimating the Average cost of the case in the CMG group

$$E\,(avgCCEcm) = avgCCEcmg - avgACCcmg$$

– Error estimating the Standard deviation of the case costs in the CMG group

$$E\,(stdevCCEcmg\,) = stdevCCEcmg - stdevACCcmg$$

– Error estimating the Minimum case cost in the CMG group

$$E\,(\,minCCEcmg\,) = minCCEcmg - minACCcmg$$

- Error estimating the Maximum case cost in the CMG group

$$E\,(\,maxCCEcmg\,) = maxCCEcmg - maxACCcmg$$

7. Estimating error analysis was conducted to compare various financial models, determine their accuracy, identify factors impacting results, and fine-tune models. Analysis of all types of errors calculated on step 6 would require handling matrixes with dimensions 5 x 163. Being feasible generally, it wouldn't provide results which could be easily understood and interpreted. That led to the formulation of a limited number of performance measures.

The first performance measure is P_{ab} - the percentage of the CMG groups for which the absolute value of the relative error estimating total cost of cases per CMG group lies within certain limits.

$$P_{ab} = n_{ab} / n,$$

where

n – number of the CMG groups in the experiment (in this case – 163);

n_{ab} – number of the CMG groups for which the ratio of the estimating error of the total cost of cases for the CMG group to the actual total cost for this CMG group (in percent) falls within the interval from a to b:

$$\Pi_{ab} \subset \Pi \left\{ a < \left| \frac{E(ACCEcmg)}{AACCcmg} \right| \leq b \right\}$$

Three other measures of performance characterize absolute error of estimating average, minimum and maximum case costs per CMG code averaged over all (163) CMG codes:

$$\bar{E}(avgCCEcmg) = \left\{ \sum_n \left| E(avgCCEcmg) \right| \right\} / n$$

$$\bar{E}(minCCEcmg) = \left\{ \sum_n \left| E(minCCEcmg) \right| \right\} / n$$

$$\bar{E}(maxCCEcmg) = \left\{ \sum_n \left| E(maxCCEcmg) \right| \right\} / n$$

C. Sample Financial Model

Although several financial models with gradual performance improvement were developed and verified [5], the best results were obtained using the model which is formulated in the equation (1):

The following parameters are used in the model:

The second line of the formula represents the normalization multiplier;

i-th case belongs to the n-th CMG group, i.e. search of minimum and maximum case costs in Model 8 estimates and replacing them with minimum and maximum case costs from Model 7 is done individually for each CMG group.

$CCE(Mod7)_i$ – CCE estimate calculated for the i-th case with Model 7 (non-normalized) – equation (2);

$CCE(Mod8)_i$ – CCE estimate calculated for the i-th case with Model 8 (non-normalized) – equation (3);

μ = RAND * 10exp(-9), RAND – randomly generated number in the interval from zero to one;

$minLOS_i$ and $maxLOS_i$ – minimum and maximum values of the LOS_i for the n-th CMG group. LOS in this case means all types of stay, i.e. acute, alternate and special care;

I_i – number of interventions in the i-th case. Parameter is calculated from the DAD database. I_i is set to zero, if $I_i - I_{avg} < 5I_{avg}$. I_{avg} for the dataset is 2;

D_i – number of diagnoses in the i-th case. Parameter is calculated from the DAD database. D_i is set to zero, if $D_i - D_{avg} < 3D_{avg}$. D_{avg} for the dataset is 5;

K6 and K7 – empirical coefficients used to assign weight to the number of diagnoses and interventions respectively.

PAC_RIW_i – relative intensity of the i-th case. Parameter is available in the DAD database – field PAC_RIW_WT;

RIW_val_i – relative intensity of the i-th case calculated by the CIHI. Parameter is available in the DAD database – field RIW_{val};

$minRIW\_val_i$ and $maxRIW\_val_i$ – minimum and maximum values of the RIW_val_i for the n-th CMG group;

$minPAC\_RIW_i$ and $maxPAC\_RIW_i$ – minimum and maximum values of the PAC_RIW_i for the n-th CMG group;

CPWC – cost per weighted case. Parameter is calculated annually by the OCDM for each hospital. For the LCTH in 2004-05 CPWC~$6,000;

$avgACC_{cmg}$ - average actual cost of the case in the CMG group. Parameter is available in the OCCI database;

$LOSac_i$ – Length of Stay in acute care for the i-th case. This parameter is available in the DAD database – field AcuteL;

$LOSalc_i$ – Length of Stay in alternate care for the i-th case. This parameter is available in the DAD database – field ALClen;

$LOSsc_i$ – Length of Stay in a special care unit for the i-th case. This parameter is available in the DAD database – field TotHRS.

CpD – cost per diem. Parameter is calculated annually by OCDM for each hospital. For the LCTH in 2004-05 CpD ~$1,600. Note: Total cost per diem is used without breakdown into Direct and Indirect costs which are also available in the OCDM report.

CpD * K1 – cost per diem in acute care. K1 > 1.

CpD * K2 – cost per diem in alternate care. K2 < K1.

CpD * K3 – cost per diem in a special care unit. K3 > K1.

$avgACC_{cmg}$ - average actual cost of the case in the CMG group. Parameter is available in the OCCI database.

K8 and K9 – empirical coefficients used to assign weight to the number of diagnoses and interventions respectively.

K12 and K13 – empirical coefficients used to assign weights to the maximum case cost estimates per CMG calculated with Model 7 and Model 8 respectively.

The following set of coefficients optimized model's performance: K1=0.62; K2=0.58; K3=3.04; K6=0.64; K7=0.04; K8=0.28; K9=0.22; K12=0.38; K13=1.8.

D. Data Experiment Results and Conclusions

Performance measures for the model with optimal coefficients are presented in the table below.

TABLE I
Financial Model Performance Measures

$\bar{E}(avgCCE_{cmg})$, $	$\bar{E}(minCCE_{cmg})$, $	$\bar{E}(maxCCE_{cmg})$, $
1,852.88	1,097.29	17,013.22
P_{ab}, %	Interval of relative error, %	
	a	b
25%	0	5
15%	5	10
13%	10	15
13%	15	20
21%	20	30
10%	30	50
2%	50	50+

Results of the data experiment show feasibility of the selected approach and correctness of the developed models. For 66% of the CMG groups, estimated aggregate CMG costs were calculated with a relative error under 20%. Such accuracy could be considered acceptable for practical planning and forecasting purposes. Only for 2% of the CMG groups relative error exceeded 50%.

Next phases of the research will focus on the understanding of the facility-to-facility and year–to-year data variability.

$$
\text{CCE(Mod11)}_i = \begin{cases}
\text{CCE(Mod8)}_i, \text{ if } \text{CCE(Mod7)}_i \neq \text{minCCE(Mod7)}_{cmg_n} \text{ and } \text{CCE(Mod7)}_i \neq \text{maxCCE(Mod7)}_{cmg_n} \\
\qquad \text{ and } \text{CCE(Mod8)}_i < \text{maxCCE(Mod7)}_{cmg_n} \\
\text{CCE(Mod8)}_i, \text{ if } \text{CCE(Mod7)}_i = \text{minCCE(Mod7)}_{cmg_n} \text{ and } \text{CCE(Mod8)}_i < \text{minCCE(Mod7)}_{cmg_n} \\
\text{minCCE(Mod7)}_{cmg_n}, \text{ if } \text{CCE(Mod7)}_i = \text{minCCE(Mod7)}_{cmg_n} \text{ and } \text{CCE(Mod8)}_i >= \text{minCCE(Mod7)}_{cmg_n} \\
\text{maxCCE(Mod7)}_{cmg_n}, \text{ if } \text{CCE(Mod7)}_i = \text{maxCCE(Mod7)}_{cmg_n} \\
[\text{K12*CCE(Mod7)}_i + \text{K13*CCE(Mod8)}_i] / 2, \text{ if } \text{CCE(Mod8)}_i >= \text{maxCCE(Mod7)}_{cmg_n}
\end{cases} \qquad (1)
$$

$$
\left\{ \sum_{cmg_n} (\text{avgACC}_{cmg} * N_{c,\,cmg}) \sum_{cmg_n} \sum_{i=1,\dots N_{c,\,cmg}} \begin{cases}
\text{CCE(Mod8)}_i, \text{ if } \text{CCE(Mod7)}_i \neq \text{minCCE(Mod7)}_{cmg_n} \text{ and } \text{CCE(Mod7)}_i \neq \text{maxCCE(Mod7)}_{cmg_n} \\
\qquad \text{ and } \text{CCE(Mod8)}_i < \text{maxCCE(Mod7)}_{cmg_n} \\
\text{CCE(Mod8)}_i, \text{ if } \text{CCE(Mod7)}_i = \text{minCCE(Mod7)}_{cmg_n} \text{ and } \text{CCE(Mod8)}_i < \text{minCCE(Mod7)}_{cmg_n} \\
\text{minCCE(Mod7)}_{cmg_n}, \text{ if } \text{CCE(Mod7)}_i = \text{minCCE(Mod7)}_{cmg_n} \text{ and } \text{CCE(Mod8)}_i >= \text{minCCE(Mod7)}_{cmg_n} \\
\text{maxCCE(Mod7)}_{cmg_n}, \text{ if } \text{CCE(Mod7)}_i = \text{maxCCE(Mod7)}_{cmg_n} \\
[\text{K12*CCE(Mod7)}_i + \text{K13*CCE(Mod8)}_i] / 2, \text{ if } \text{CCE(Mod8)}_i >= \text{maxCCE(Mod7)}_{cmg_n}
\end{cases} \right\}
$$

where,

$$
\text{CCE(Mod7)}_i = [(\text{LOSac}_i - \text{LOSsc}_i / 24) * \text{K1} + \text{LOSalc}_i * \text{K2} + (\text{LOSsc}_i / 24) * \text{K3}] * \text{CpD} + \text{CPWC} * [D_i * \text{K8} + I_i * \text{K9}] + \mu \qquad (2)
$$

$$
\text{CCE(Mod8)}_i = \text{CPWC} * [\text{PAC\_MOD}_i + D_i * \text{K6} + I_i * \text{K7}] \qquad (3)
$$

$$
\text{PAC\_MOD}_i = \begin{cases}
\text{PAC\_RIW}_i, \text{ if minPAC\_RIW}_i \neq \text{maxPAC\_RIW}_i \text{ for n-th CMG group} \\
\qquad \text{or minPAC\_RIW}_i = \text{maxPAC\_RIW}_i \text{ and minRIW\_val}_i = \text{maxRIW\_val}_i \\
\qquad \text{and minLOS}_i = \text{maxLOS}_i \text{ for n-th CMG group} \\
\text{RIW\_val}_i * \left(\sum_{cmg_n} \text{PAC\_RIW}_i / \sum_{cmg_n} \text{RIW\_val}_i \right), \text{ if minPAC\_RIW}_i = \text{maxPAC\_RIW}_i \\
\qquad \text{and minRIW\_val}_i \neq \text{maxRIW\_val}_i \text{ for n-th CMG group} \\
\text{CCE(Mod7)}_i * \left(\sum_{cmg_n} \text{PAC\_RIW}_i / \sum_{cmg_n} \text{CCE(Mod7)}_i \right), \text{ if minPAC\_RIW}_i = \text{maxPAC\_RIW}_i \text{ and minRIW\_val}_i = \text{maxRIW\_val}_i \\
\qquad \text{and minLOS}_i \neq \text{maxLOS}_i \text{ for n-th CMG group}
\end{cases}
$$

ACKNOWLEDGMENT

The work was prepared in connection with both authors' official duties at the Ontario Ministry of Health and Long-Term Care. The views, opinions and conclusions expressed in this document are those of the authors alone, and do not necessarily represent the views of the Ontario Ministry of Health and Long-Term Care or any of its individual departments.

REFERENCES

1. Ontario Ministry of Health and Long-Term Care (MOHLTC). Financial Modelling Project: HDDSU & IMU. Discovery Document. Version 05. Aug. 2006, Toronto, Ontario (Unpublished report).

2. Ontario Ministry of Health and Long-Term Care. Provincial health planning database. Online. 2006. Available from URL: http://portal.healthdss.moh.gov.on.ca

3. Ontario Ministry of Health and Long-Term Care. Finance and Information Management Branch. Online. 2006. Available from URL: http://www.mohltcfim.com

4. Ontario Case Costing Initiative (OCCI). OCCI Costing Analysis Tool. Online. 2006. Available from URL: http://www.occp.com

5. P. Andru, A. Botchkarev Hospital Case Cost Estimates Modelling – Algorithm Comparison.

   http://arxiv.org/ftp/arxiv/papers/0802/0802.4126.pdf, 2008.