Reduction of operative mortality after implementation of Surgical Outcomes Monitoring and Improvement Programme by Hong Kong Hospital Authority

WC Yuen, K Wong, YS Cheung, Paul BS Lai *

ABSTRACT

Introduction: Since 2008, the Hong Kong Hospital Authority has implemented a Surgical Outcomes Monitoring and Improvement Programme (SOMIP) at 17 public hospitals with surgical departments. This study aimed to assess the change in operative mortality rate after implementation of SOMIP.

Methods: The SOMIP included all Hospital Authority patients undergoing major/ultra-major procedures in general surgery, urology, plastic surgery, and paediatric surgery. Patients undergoing liver or renal transplantation or who had multiple trauma or massive bowel ischaemia were excluded. In SOMIP, data retrieval from the Hospital Authority patient database was performed by six full-time nurse reviewers following a set of precise data definitions. A total of 230 variables were collected for each patient, on demographics, preoperative and operative variables, laboratory test results, and postoperative complications up to 30 days after surgery. In this study, we used SOMIP cumulative 5-year data to generate risk-adjusted 30-day mortality models by hierarchical logistic regression for both emergency and elective operations. The models expressed overall performance as an annual observed-to-expected mortality ratio.

Results: From 2009/2010 to 2015/2016, the overall crude mortality rate decreased from 10.8% to 5.6% for emergency procedures and from 0.9% to 0.4% for elective procedures. From 2011/2012 to 2015/2016, the risk-adjusted observed-to-expected mortality ratios showed a significant downward trend for both emergency and elective operations: from 1.126 to 0.796 and from 1.150 to 0.859, respectively (Mann-Kendall statistic = –0.8; P<0.05 for both).

Conclusion: The Hospital Authority’s overall crude mortality rates and risk-adjusted observed-to-expected mortality ratios for emergency and elective operations significantly declined after SOMIP was implemented.

New knowledge added by this study

• A Surgical Outcomes Monitoring and Improvement Programme allows monitoring of performance and fair comparison of individual Hospital Authority hospitals against the overall Hospital Authority average. It enhances the understanding of surgical performance and helps identify areas for improvement.

• The Hospital Authority’s overall crude mortality rates and risk-adjusted observed-to-expected mortality ratios for emergency and elective operations significantly declined after SOMIP was implemented.

Implications for clinical practice or policy

• A properly organised, risk-adjusted clinical audit can accurately measure surgical outcomes and provide information for surgeons to deliver quality improvement.

Introduction

Audits of surgical mortality are used worldwide to monitor surgical outcome and achieve quality assurance.1 By measuring and comparing properly collected, risk-adjusted surgical outcome data, quality of surgical care could be enhanced in participating institutions.2 It has been demonstrated in several countries that adoption of a national surgical audit programme can reduce mortality.2,3

The Hong Kong Hospital Authority (HA) was established in 1991. It is a government statutory body responsible for the management of 42 public hospitals and institutions, 47 specialist out-patient clinics, and 73 general out-patient clinics in Hong Kong. Seventeen HA hospitals have surgical departments; all of them provide an elective surgery
Methods

Between July 2008 and June 2016, SOMIP captured data of all elective and emergency major/ultra-major operations (except those in children younger than 1 year) that were performed by general surgery, urology, plastic and reconstructive surgery, and paediatric surgery teams at all 17 HA hospitals. A total of 230 variables were collected from each patient: 10 patient demographic variables, 83 preoperative and operative variables, 31 laboratory test results, and 40 postoperative events and 66 postoperative adverse events in the first 30 days after surgery. Demographic data and laboratory test results were mostly automatically retrieved from various HA clinical information systems. For data that required manual retrieval, six full-time SOMIP nurse reviewers were employed by the HA head office for this purpose. Preoperative and operative variables, as well as postoperative complications occurring up to 30 days after the index operation, were retrieved from patient records by the SOMIP nurse reviewers. Mortality at 30 days, 60 days, and 90 days were also retrieved from the HA electronic database. These data were endorsed and submitted by each surgical department's surgical supervisor within 60 days of surgery.

Both the manually captured and automatically captured data were entered into a tailor-made SOMIP electronic database. Data variable definitions were listed in the operation manual of the programme. To ensure data validity and consistency, all nurse reviewers completed comprehensive training on data definition and criteria, and regular nurse reviewer meetings were held to clarify any queries. All data were endorsed by the surgical supervisor of the respective surgical department. When necessary, data definitions were modified.

An inter-rater reliability test was performed each May and completed within a month so as to ensure consistency among nurse reviewers as well as data accuracy. Fifty cases were sampled for evaluation using a stratified systematic sampling method. For those selected cases, an independent nurse from the SOMIP team repeated the data collection using a designated data template, without prior knowledge of the information recorded by the original nurse reviewer. Data quality was measured by comparing two sets of data, and inferred by a score defined by each surgical department's surgical supervisor.
as the percentage of agreement between nurse reviewer and the SOMIP working team for each data item. The mean score of all data items was used to assess overall performance. The overall result was satisfactory and the mean score of all data items was 99.3% (range of individual item scores, 95.2%-100%).

Among the variables collected, preoperative risk factors including demographic data; general health and lifestyle variables; and major respiratory, cardiovascular, hepatobiliary, renal, vascular, central nervous system, and immune co-morbidities were deemed particularly important. These risk factors were modified from those in the American College of Surgeons National Surgical Quality and Improvement Program (NSQIP) to suit the local context.7-9

Operative variables included intra-operative blood loss, American Society of Anesthesiologists physical status classification, procedure complexity score, surgical subspecialty, wound classification, operative magnitude, and operative time. Before analysis and reporting, SOMIP data were cleaned and verified by four surgeons and the SOMIP surgical supervisors of each hospital. Questionable cases were reviewed accordingly. Cases of liver transplantation, renal transplantation, multiple trauma, and major bowel ischaemia (Table 1)10-17 were excluded from the risk-adjusted model analysis.

Comparative risk-adjusted models (different models for emergency operations and elective operations) were generated using hierarchical logistic regression. The 30-day risk-adjusted mortality models expressed hospital performance as expected odds ratios. A risk-adjusted observed-to-expected mortality ratio (O/E ratio) was then calculated for each hospital. The O/E ratio is a quotient between the observed number of deaths and the expected number of deaths; the latter was calculated by a logistic regression method based on significant independent risk factors. Together with the 90% confidence intervals, O/E ratios were depicted by caterpillar plots and boxplots. Hospitals with the lower limit of the 90% confidence interval of O/E ratios greater than 1 were defined as ‘high outliers’; hospitals with the upper limit of the 90% confidence interval of O/E ratios lower than 1 were defined as ‘low outliers’. The risk-adjusted outcome of a ‘high-outlier’ hospital was probably worse than the average outcome, and that of a ‘low-outlier’ hospital was probably better than the average outcome.

In addition to risk-adjusted postoperative mortality, various general medical and surgical complications, as well as specific complications (anastomotic leakage, surgical site infection, acute myocardial infarction, pneumonia), were recorded. The list of complications recorded and the method to derive the SOMIP risk-adjustment model have been described in detail in the annual SOMIP Report.10-17 Different levels of confidence were used for different outcome variables—90% confidence interval for mortality rates, 95% confidence interval for major complications, and 99% confidence interval for morbidity rates. Using SOMIP data together with other useful information extracted from the HA Executive Information System (eg, bed occupancy, nursing manpower, intensive care unit support, and surgeon workload), root-cause analyses were performed using multilevel logistic regression, as described in the annual SOMIP Report.10-17

The discriminative power of the risk-adjusted models was measured by the C-index, area under the receiver-operating characteristic curve (AUC). The closer the C-index is to 1, the better the discriminative power of the model is; a C-index of ≥0.8 indicates excellent discriminative power.18 The calibration accuracy of the models was assessed by the Hosmer-Lemeshow goodness-of-fit test (HL test). The calibration of the model was rejected if P<0.05. The Mann-Kendall non-parametric trend test was used to identify trends (positive or negative) in the annual data series for both crude mortality rates (2009/2010 to 2015/2016) and risk-adjusted O/E ratios (2011/2012 to 2015/2016). A very high positive value of the Mann-Kendall statistic (S) indicated an increasing trend; a very low negative value indicated a decreasing trend. The test statistic Z-score was used as a measure of trend significance.

Results

Descriptive data

Age distribution trends are summarised in Table 2.10-17 From 2008/2009 to 2015/2016, the proportion of people aged 61-70 years increased by 7 percentage points (from 19% to 26%), whereas the proportions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver transplantation</td>
<td>100</td>
<td>76</td>
<td>80</td>
<td>68</td>
<td>79</td>
<td>58</td>
<td>62</td>
<td>58</td>
</tr>
<tr>
<td>Renal transplantation</td>
<td>82</td>
<td>65</td>
<td>77</td>
<td>64</td>
<td>103</td>
<td>74</td>
<td>88</td>
<td>66</td>
</tr>
<tr>
<td>Multiple trauma</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>19</td>
<td>24</td>
<td>98</td>
<td>21</td>
</tr>
<tr>
<td>Major bowel ischaemia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>23</td>
<td>20</td>
<td>35</td>
<td>18</td>
</tr>
</tbody>
</table>

TABLE 1. No. of cases excluded and reasons, from 2008/2009 to 2015/201610-17
of people aged 41-50 years and 71-80 years decreased by 2 and 5 percentage points, respectively.

Table 3 shows proportions of patients taking regular medication for diabetes mellitus and hypertension before surgery. From 2009/2010 to 2015/2016, as well as those currently smoking (within 1 year) and drinking more than 2 units of alcohol per day in the previous 2 weeks. Over 70% of patients had at least one of these four conditions, whereas about 40% had a history of regular use of hypertension medication before surgery.

Annual numbers of elective operations by specialty are summarised in Table 4. For the 10 listed elective operations, the most frequently performed were in urology, consistently constituting 28% of the caseload from 2008/09 to 2015/16. The least frequently performed procedures were parotid surgery (1%) and paediatric surgery (1%).

Overall crude mortality rates and risk-adjusted observed-to-expected mortality ratios

From July 2008 to June 2016, eight SOMIP reports were published. They showed that the HA overall crude mortality rate approximately halved over this time. The crude 30-day mortality rate for emergency operations dropped gradually from 10.8% in the year 2009/2010 to 5.6% in 2015/2016 (Fig 1). Similarly, the crude 30-day mortality rate for elective operations more than halved: from 0.9% in 2009/2010 to 0.4% in 2015/2016 (Fig 2).

In the 5-year cumulative comparison analysis (2011/2012 to 2015/2016), both models had excellent discriminative power and good calibration accuracy. For emergency operations, the AUC was >0.9 and the HL test statistic was >0.1; for elective operations, the AUC was >0.89 and the HL test statistic was

### Table 2. Distribution of patients by age-group, from 2008/2009 to 2015/2016

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>585 (3%)</td>
<td>595 (3%)</td>
<td>452 (2%)</td>
<td>538 (2%)</td>
<td>522 (2%)</td>
<td>595 (2%)</td>
<td>565 (2%)</td>
<td>569 (2%)</td>
<td>4421 (2%)</td>
</tr>
<tr>
<td>11-20</td>
<td>399 (2%)</td>
<td>396 (2%)</td>
<td>366 (2%)</td>
<td>340 (1%)</td>
<td>390 (2%)</td>
<td>343 (1%)</td>
<td>357 (1%)</td>
<td>359 (1%)</td>
<td>2950 (2%)</td>
</tr>
<tr>
<td>21-30</td>
<td>503 (2%)</td>
<td>605 (3%)</td>
<td>549 (2%)</td>
<td>590 (2%)</td>
<td>571 (2%)</td>
<td>592 (2%)</td>
<td>660 (3%)</td>
<td>635 (2%)</td>
<td>4705 (2%)</td>
</tr>
<tr>
<td>31-40</td>
<td>1181 (5%)</td>
<td>1214 (5%)</td>
<td>1187 (5%)</td>
<td>1285 (5%)</td>
<td>1250 (5%)</td>
<td>1286 (5%)</td>
<td>1276 (5%)</td>
<td>1278 (5%)</td>
<td>9957 (5%)</td>
</tr>
<tr>
<td>41-50</td>
<td>2672 (12%)</td>
<td>3057 (13%)</td>
<td>2786 (12%)</td>
<td>2886 (12%)</td>
<td>2783 (12%)</td>
<td>2696 (11%)</td>
<td>2745 (11%)</td>
<td>2671 (10%)</td>
<td>2296 (12%)</td>
</tr>
<tr>
<td>51-60</td>
<td>4251 (19%)</td>
<td>4680 (20%)</td>
<td>4599 (20%)</td>
<td>5095 (21%)</td>
<td>5012 (21%)</td>
<td>5128 (21%)</td>
<td>5300 (21%)</td>
<td>5433 (21%)</td>
<td>39498 (21%)</td>
</tr>
<tr>
<td>61-70</td>
<td>4253 (19%)</td>
<td>4771 (20%)</td>
<td>4829 (21%)</td>
<td>5272 (22%)</td>
<td>5317 (22%)</td>
<td>5695 (23%)</td>
<td>6198 (25%)</td>
<td>6604 (26%)</td>
<td>42939 (22%)</td>
</tr>
<tr>
<td>71-80</td>
<td>5486 (25%)</td>
<td>5752 (24%)</td>
<td>5486 (23%)</td>
<td>5341 (22%)</td>
<td>5104 (22%)</td>
<td>5064 (21%)</td>
<td>5005 (20%)</td>
<td>5007 (20%)</td>
<td>42225 (22%)</td>
</tr>
<tr>
<td>81-90</td>
<td>2295 (11%)</td>
<td>2427 (10%)</td>
<td>2833 (12%)</td>
<td>2664 (11%)</td>
<td>2544 (11%)</td>
<td>2661 (11%)</td>
<td>2762 (11%)</td>
<td>2745 (11%)</td>
<td>20931 (11%)</td>
</tr>
<tr>
<td>91-100</td>
<td>232 (1%)</td>
<td>243 (1%)</td>
<td>293 (1%)</td>
<td>226 (1%)</td>
<td>227 (1%)</td>
<td>258 (1%)</td>
<td>225 (1%)</td>
<td>226 (1%)</td>
<td>1930 (1%)</td>
</tr>
<tr>
<td>&gt;100</td>
<td>-</td>
<td>2 (0%)</td>
<td>7 (0%)</td>
<td>3 (0%)</td>
<td>3 (0%)</td>
<td>6 (0%)</td>
<td>4 (0%)</td>
<td>27 (0%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>21839 (100%)</td>
<td>23740 (100%)</td>
<td>23382 (100%)</td>
<td>24244 (100%)</td>
<td>23723 (100%)</td>
<td>24321 (100%)</td>
<td>25099 (100%)</td>
<td>25531 (100%)</td>
<td>191879 (100%)</td>
</tr>
</tbody>
</table>

* Data are shown as No. (%) of patients

### Table 3. Distribution of patients by habit before surgery/admission, from 2009/2010 to 2015/2016

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular diabetes mellitus medication before operation</td>
<td>3581 (15%)</td>
<td>3452 (15%)</td>
<td>3646 (15%)</td>
<td>3789 (16%)</td>
<td>3925 (16%)</td>
<td>3981 (16%)</td>
<td>4324 (17%)</td>
<td>26698 (14%)</td>
</tr>
<tr>
<td>Regular hypertension medication before operation</td>
<td>10569 (45%)</td>
<td>9560 (41%)</td>
<td>9819 (41%)</td>
<td>9870 (42%)</td>
<td>10311 (42%)</td>
<td>10647 (42%)</td>
<td>11071 (43%)</td>
<td>71847 (37%)</td>
</tr>
<tr>
<td>Current smoker within 1 year</td>
<td>3428 (14%)</td>
<td>3593 (15%)</td>
<td>3903 (16%)</td>
<td>3823 (16%)</td>
<td>3892 (16%)</td>
<td>3982 (16%)</td>
<td>4226 (17%)</td>
<td>26847 (14%)</td>
</tr>
<tr>
<td>Alcohol &gt;2 units per day in the 2 weeks prior to admission</td>
<td>777 (3%)</td>
<td>709 (3%)</td>
<td>707 (3%)</td>
<td>663 (3%)</td>
<td>535 (2%)</td>
<td>499 (2%)</td>
<td>520 (2%)</td>
<td>4410 (2%)</td>
</tr>
<tr>
<td>Total</td>
<td>18355 (77%)</td>
<td>17314 (74%)</td>
<td>18075 (75%)</td>
<td>18145 (76%)</td>
<td>18683 (77%)</td>
<td>19109 (76%)</td>
<td>20141 (79%)</td>
<td>129802 (68%)</td>
</tr>
</tbody>
</table>

* Data are shown as No. (%) of patients
>0.2. The risk-adjusted observed-to-expected 30-day mortality ratio for both types of surgery showed a similar downward trend to the crude mortality rates. For emergency operations, the risk-adjusted O/E ratios were 1.126, 1.022, 1.113, 0.921, and 0.796 across the 5 years (Fig 3). These values show a statistically significant downward (negative) trend ($S = -0.8; P<0.05$). The results of the Mann-Kendall analyses are summarised in Table 4. For elective operations, risk-adjusted O/E ratios were 1.150, 1.229, 0.881, 0.862, and 0.859 across the 5 years (Fig 4). These values show a statistically significant downward (negative) trend ($S = -0.8; P<0.05$) [Table 5].

### Discussion

Before the turn of the century, most hospital records in HA hospitals were handwritten and retained by individual hospitals. There was no convenient means by which to share patient details among hospitals. Around 2001, the HA implemented a number of clinical management electronic systems, such as the electronic patient record, Operation Theatre Record System, and Clinical Data Analysis and Reporting System, at all HA hospitals. By virtue of this infrastructure, patient records and information about diagnoses and operations could be accessed at a central level. Based on this central clinical database, a Quality Assurance Subcommittee under the Coordinating Committee of Surgery commenced small-scale comparative clinical audits for ultra-major operations in 2002, focusing on one to two ultra-major operations per year. The audits provided basic information about hospital performance for the selected operation, such as number of procedures, age distribution of patients, and mortality rate.

Subsequently, the Coordinating Committee of Surgery developed a more robust system to monitor more major operations at the same time. It was decided to follow the framework of NSQIP, which was developed in 1994 by the Veterans Affairs Hospitals.
in the United States to monitor risk-adjusted surgical operation outcomes. Like HA hospitals, Veterans Affairs Hospitals are managed by a central governing body and equipped with a comprehensive electronic medical records system. Studies showed a significant improvement in both mortality and morbidity over time and thus, in 2004, NSQIP was extended to private hospitals and endorsed by the American College of Surgeons. Using NSQIP as the blueprint, the HA launched SOMIP in 2008. The SOMIP adopted similar risk-adjustment variables, use of nurse reviewers to collect data, a focus on hospital performance rather than individual surgeon performance, and similar methods of data analysis and determination of outliers. Moreover, SOMIP allowed individual HA hospitals to benchmark their performance against the overall HA average.

As all surgical patients have a different health status, their operation outcomes will likewise differ. Appropriate adjustment for different patient risks is essential when interpreting hospital mortality rates. To adjust for different risk factors, over 100 patient risk factors were captured for each enrolled patient. For NSQIP, one risk-adjustment mortality model was constructed for all operations. In contrast to NSQIP, separate models were devised for emergency and elective operations in SOMIP.

Hospital outliers can be identified by O/E ratios if the confidence interval of the O/E ratio is greater than 1, meaning that after balancing the different risks of hospital patients, their clinical outcomes are most likely different from the rest. From the results of this study, it was encouraging to find a significant trend of reduction in crude mortality rates and O/E ratios for both elective and emergency operations at HA hospitals over the past 5 years. According to the significant reduction in expected odds ratios over the years, this improvement is genuine and not due to patient selection.

There are several possible reasons behind the changes: public identification as a poor performer is a strong incentive for change in HA hospitals; sharing best practices on perioperative patient care is an important educational activity that takes place annually through the SOMIP Forum; the HA Head Office invests more resources into deficient hospitals; and changing attitudes towards managing surgical complications by other colleagues from the intensive care unit are also helpful. All of these may have contributed to the change.

Limitations

The SOMIP has a number of limitations. The coverage of SOMIP is not as complete as that of NSQIP, since many surgical departments such as orthopaedics and
neurosurgery are not included. Monitoring is done by retrospective annual case collection (from 1 July to 30 June) because it takes 13 months to complete case enrolment and an additional 3 months for data verification, model building, and statistical analysis. Because of the small number of events, the current programme is not able to determine the risk-adjusted outcomes of individual operations or surgeons. Furthermore, because this programme relies heavily on the HA central electronic database, it is not easy to extend it to hospitals without this information infrastructure. Although the trend of reduction in mortality was statistically significant, we were not able to demonstrate a causal relationship with SOMIP implementation.

Potential issues with data quality may have affected the outcomes. In the inter-rater reliability test, the nurse reviewers were not blinded and this may have caused information bias. Also, quality of data collection in the initial 2 years may have been unreliable. As a result, the 5-year cumulative comparison analysis for emergency and elective operations commenced from 2011/2012, rather than 2008/2009. Furthermore, data definitions are updated regularly in the operation manual and could have affected the time trend analysis. Nonetheless, the SOMIP team considered changes in data definitions to be minor and did not expect a significant impact on the risk models.

Mortality could be influenced by many factors; ensuring risk adjustments are adequate and appropriate would be a challenge. Disease factors, stage of disease, and treatment options may not be fully taken into account by the risk-adjusted models, and data readiness and availability are further constraints. Surgeon skill and experience was another aspect that could not be accommodated and was difficult to adjust for. In the HA, surgical operations are performed by a team; therefore, it would be difficult to separate individual surgeon experience and credentials from those of the whole team.

Conclusion
From 2008 to 2016, the HA’s overall crude mortality rates and risk-adjusted O/E ratios showed a significant trend of reduction for both emergency and elective operations. The SOMIP enhances understanding of surgical performance and helps identify areas for improvement. It allows individual HA hospitals to benchmark their performance against the overall HA average through risk-adjusted O/E ratios.

Acknowledgements
We thank the SOMIP Steering Committee of the Hospital Authority, the Coordinating Committee of Surgery, and the Biostatistics team of The Chinese University of Hong Kong for their contributions and helpful comments on this manuscript.

Declaration
The authors have no conflicts of interest to disclose.

References
