The trade-offs of using different physician attribution methods for audit and feedback interventions in general medicine inpatient care

by

Terence Tang
BMath, University of Waterloo, 2005
MD, University of Toronto, 2009

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTERS OF SCIENCE

in the School of Health Information Science

Supervisory Committee

The trade-offs of using different physician attribution methods for audit and feedback interventions in general medicine inpatient care

by

Terence Tang
BMath, University of Waterloo, 2005
MD, University of Toronto, 2009

Supervisory Committee

Dr. Francis Lau, School of health Information Science
Supervisor

Dr. Keith Denny, Director, Clinical Data Standards and Quality, CIHI
Departmental Member

Abstract

BACKGROUND: Audit and feedback interventions have the potential to improve clinical care. Electronically captured administrative and clinical data routinely collected in Canadian hospitals may be used to provide feedback to physicians in general medicine in-patient care. The computation of appropriate quality indicator requires patient care to be attributed to individual physician(s). The appropriate attribution method in contexts where multiple physicians are involved in the care with varying degree of responsibilities that change over time is not straight forward. There has so far been little guidance in the literature of how to best accomplish this. The objective of this study is to identify trade-offs of different physician attribution methods by applying them to the same large clinical dataset.

METHODS: A retrospective cohort study was conducted using the GEMINI dataset consisting of administrative and clinical data of hospitalized patients discharged from General Medicine service between April 1, 2010 and October 31, 2017 extracted from electronic systems at 7 hospitals in the Greater Toronto Area. A set of four quality indicators (length of stay, 30-day re-admission, in-patient mortality, use of advanced imaging) used in an audit and feedback intervention was calculated for each physician using 5 different physician attribution methods: STRICT (only patients with the same admitting, discharging, and most responsible physician with length of stay less than 14 days were included to capture those patients whose care was provided by only 1 physician), ADMIT (attribute care to admitting physician), DISCHARGE (attribute care to discharging physician), MRP (attribute care to most responsible physician), and ANY (attribute care to admitting, discharging, and most responsible physicians). The comprehensiveness and comparability of each attribution method were calculated. The actual differences of the indicator value and physician ranking for each indicator was compared between each pair of attribution methods.

RESULTS: 222,490 hospitalization cared for by 203 physicians were included. STRICT attribution method was least comprehensive, capturing only 40% of patients cared for by a physician), while ADMIT, DISCHARGE, and MRP captured 70% of patients. All attribution methods produced patient populations for individual physicians that were comparable to those seen at each hospital. STRICT attribution method resulted in length of stay values 4.7 to 6.8 days shorter than other attribution methods and had poor rank correlation of physicians when compared to other attribution methods (spearman rank correlation 0.27 to 0.52). Absolute differences for the other 3 indicators were small between all attribution methods, and relative ranking of physicians were reasonably preserved (strong or very strong rank correlation).

INTERPRETATION: Different attribution methods have different comprehensiveness, but all produced mostly comparable patient populations for physicians. Certain attribution method can affect apparent physician performance for some quality indicators but not others. The impact of physician attribution methods deserve consideration during the design of audit and feedback interventions.

Table of Contents

List of Tables

List of Figures

None.

Acknowledgments

I would like to acknowledge the entire GEneral Medicine INpatient Initiative (GEMINI) team for allowing me access to the dataset, and for the many intellectual conversations we have had about this and other topics, especially co-leads. Dr. Amol Verma and Dr. Fahad Razak.

Dedication

To my wife for her unwavering support and to my children whose curiosity constantly inspires me.

Chapter 1: Introduction

Audit and feedback (providing physicians with summary of their own clinical performance over a specified time period [analogous to report cards] with the goal of improving care) is an important strategy in improving quality of healthcare (1, 2). Electronic health records have now made collecting clinical performance data for a large number of physicians for audit and feedback interventions feasible and scalable (3-8). However, producing data alone is not sufficient. A systematic review found audit and feedback interventions have overall led to small improvement in clinicians' rate of compliance with desired practice but with great variability in effectiveness among these interventions (median 4.3% absolute increase in desired practice, with interquartile range 0.5 to 16% from 82 comparisons in 49 studies involving 2310 groups of health professionals and 2053 individual professionals) (2). Research has shown that physicians' trust in the data (including the belief that the data accurately reflect their own performance) is an important factor in the success of audit and feedback interventions (9-12). In this study, the term "physician attribution method" refers to how clinical outcomes or care processes of patients are linked to a particular physician. The physician attribution method used can affect how physicians perceive the feedback as accurately reflecting their own performance or not, and hence the success of the intervention.

Importance of physician attribution in audit and feedback

There is now increasing evidence that quality improvement initiatives should target both individual physicians and hospitals as some studies have found greater variation in care at the physician level when compared to the hospital level in both surgical (13) and general medical care (14, 15). Moreover, both physicians and other stakeholders expect that quality measures have proper attribution (that is, “correctly assigning the measure to the individual, group, or organization responsible for the decisions, costs, and outcomes”) (16).

Challenge in physician attribution for general medicine inpatient care

In some areas of medicine, the appropriate physician attribution method is straight forward. For example, in primary care (17) or surgery, it is logical to link patient care processes and outcomes (e.g. cancer screening adherence or post-operative complications) to the patient’s named family doctor or to the surgeon who performed the surgery. Most physicians will find the physician attribution method reasonable and will trust the data presented as their own performance. However, there are areas of medicine (such as general medicine in-patient care) where the appropriate physician attribution method is less straight forward, where multiple physicians are involved in the care of one patient with varying degrees of shared responsibilities that change over time (15, 18). For example, an elderly patient was admitted to the hospital from the Emergency Department for pneumonia. He deteriorated after 3 days in hospital and passed away in the evening after an unsuccessful resuscitation. If the audit and feedback intervention were to report mortality rate to individual physicians, should this death be linked to the reporting for physician A (who initially admitted the patient and made clinical decision about choice of antibiotics), physician B (who was the attending physician and cared for patients on days 2 and 3), physician C (who was covering in the evening and was in charge of the failed resuscitation attempt), or to none, some, or all of these physicians?

Currently, there is limited literature to guide the selection of appropriate physician attribution methods in these situations. This study attempts to address this gap.

The objective of this study is to compare different physician attribution methods by applying them to the same large clinical dataset to understand how each method affects individual physician's apparent performance on a set of quality indicators in real life. The result of this study will inform how best to select physician attribution methods for use in audit and feedback interventions for general medicine in-patients so that physicians will trust the data and be receptive to making meaningful improvement using the data.

Chapter 2: Methods

This study uses a retrospective cohort design and analyzes physician performance data in the care of general medicine hospitalized patients at 7 hospitals in Ontario from 2010-2017.

In this study, five different physician attribution methods were applied on the same large real clinical dataset to compare how each affect the reported performance of individual physicians. To examine physician performance, 4 quality indicators were calculated. These indicators were selected from the 6 quality indicators used by Health Quality Ontario's General Medicine My Practice report (20) first distributed in 2019 as part of the General Medicine Quality Improvement Network (GeMQIN) initiative (19).

Research Ethics

This study received ethics approval by the University of Victoria Research Ethics board (see appendix A).

Data Sources

The GEMINI dataset consists of administrative and clinical data extracted from hospital electronic systems for all hospitalizations to general internal medicine at 7 hospitals in the Greater Toronto Area with discharge dates between Apr 1, 2010 and Oct 31, 2017 as previously described (21). Data elements include administrative, case costing, laboratory, diagnostic imaging, pharmacy, and clinical (e.g. vital signs) data (21). This study is a secondary analysis of the original dataset.

The original dataset was obtained from 7 hospitals in the Greater Toronto Area including Mount Sinai Hospital, St. Michael's Hospital, Sunnybrook Health Sciences Centre, Trillium Health Partners (Credit Valley Hospital and Mississauga Hospital) and University Health Network (Toronto General Hospital and Toronto Western Hospital). The Decision Support department at each hospital extracted data elements from the local source hospital information systems (which were different at each site). Data extracted was then de-identified locally. The de-identified data was then securely transferred to St. Michael's Hospital (the study's central coordinating site). To prepare a dataset suitable for secondary analysis for this study, a data analyst prepared raw data tables (which did not contain any identifiable information on the physician or patient) containing the quality indicator results for each physician (meeting inclusion criteria below) for each attribution method. Analysis for this study was done from these raw data tables by the author.

Inclusion Criteria

For the purpose of this study, physicians who have cared for at least 100 patients in the GEMINI dataset were included to ensure calculation of physician level quality indicators are meaningful with sufficient sample size per physician, similar to approach used in a previous study (18). The analysis was performed on all patients in the dataset who were either admitted or discharged by a physician of interest (as defined above). Only hospitalizations of patients admitted through the Emergency Department (the usual care pathway) were included to avoid including elective admissions or inter-facility transfers (18).

Physician Attribution Methods

In Canada, three physician roles are routinely collected in hospital administrative data for every hospitalization. These are the admitting physician, discharging physician, and most responsible physician (MRP, defined as the physician responsible for the majority of stay or the most resource use) and reported in the Discharge Abstract Database (DAD) at Canadian Institute of Health Information (CIHI) (22). For each hospitalization, each of these 3 roles may be filled by the same or different physician(s). For the purpose of this study, we considered 5 different physician attribution methods. In addition to each of the 3 roles as a distinct physician attribution method, we also include two additional methods (ANY and STRICT). The purpose of the STRICT method is to capture those patients whose care is likely attributable to only 1 physician, making physician attribution less ambiguous. The ANY method is meant to be the most comprehensive, capturing any patients that were cared for by a particular physician regardless of which role (admitting, discharging, or most responsible) he/she played. The detailed definition of each method is as follows:

STRICT: include only hospitalizations where the admitting, discharging, and most responsible physician are the same individual, and that the length of stay is 14 days or less. This method attributes the hospitalization to that single physician. This method in theory captures those hospitalizations that likely only involve the care of 1 physician (without repeated change over)

ADMIT: Attribute each hospitalization to the admitting physician.

DISCHARGE: Attribute each hospitalization to the discharging physician.

MRP: Attribute each hospitalization to the most responsible physician

ANY: For each physician X, include all hospitalizations where physician X is one or more of admitting, discharging, or most responsible physician. The intention is to capture all care provided by a physician regardless of the role he/she played.

Quality Indicators

For the purpose of this study, we calculated individual physician performance for four selected quality indicators, using each of the 5 physician attribution methods above. The four quality indicators were selected from the six quality indicators currently used by Health Quality Ontario's MyPractice General Medicine reports (20). The first three (hospital length of stay, re-admission within 30 days, and in-patient mortality) are quality indicators commonly used in quality improvement interventions on general medical wards as found in a systematic review (23). They also do not have a clear clinical action linked to the quality indicator. The fourth quality indicator (use of advanced imaging) was selected as a representative example of the remaining three quality indicators (including rate of routine bloodwork and rate of inappropriate transfusion) used in the MyPractice General Medicine reports that specifically examine resource use derived from Choosing Wisely Canada's recommendations for Internal Medicine (24). This indicator has a clinical action linked to the quality indicator, but that is not reliably captured in the electronic dataset (i.e. the ordering physician). Therefore, the 4 quality indicators included are as follows:

1. ***Length of Stay***. This is defined as discharge date/time – admission date/time in days. For each physician, we report on the average length of stay for all patients attributed to that physician.
2. ***Re-admission within 30 days***. We count re-admissions within the GEMINI dataset (i.e. readmission to general medicine at any of the 7 hospitals). For each physician, we calculate the number of patients with a re-admission within 30 days / total number of patients discharged alive attributed to that physician.
3. ***In-patient mortality***. For each physician, we calculate the number of patients who died during the hospitalization / total number of patients attributed to that physician.
4. ***Use of advanced imaging***. For each physician, we calculate the number of CT/MRI/ultrasound tests performed for that physician's patients / total number of patients attributed to that physician

Analysis

As the objective of this study is to compare how different attribution methods affect quality indicators fed back to individual physicians under real world conditions, the analysis focuses on comparing the different attribution methods with each other along a number of dimensions at the individual physician level.

The development of quality indicators has been reviewed by Stelfox and Straus (25, 26), as well as others (27-30). There are 6 proposed criteria to evaluate quality indicators including a) the strength of scientific evidence supporting the measure, b) equal likelihood of individuals in denominator to be included in numerator, c) measure under the control of whom the measure evaluates, d) how well do the measure capture event that is subject of measurement, e) does the measure provide a fair comparison of performance, f) does the measure allow for adjustment to exclude patients with rare performance-related characteristics when appropriate (26). As the focus of this study is on attribution methods (and not the content of the quality indicators themselves), criteria d and e are most relevant and will be included as dimensions of the analysis.

To address criterion d (extent of capturing events of interest; in this case, the care provided by the individual physician), the dimension of ***comprehensiveness*** will be used (see below). To address criterion e (fair comparison of performance), the dimension of ***comparability*** will be used (see below). To quantify the absolute difference of measured performance of individual physicians and the difference in relative ranking of physicians by different attribution methods, the dimension of ***consistency*** (see below) will be used. Finally, in this dataset composed of care at 7 hospitals, the ***generalizability*** of the findings across hospitals and care models will also be assessed.

Comprehensiveness refers to the extent to which care provided by a physician is captured by a particular attribution method. For example, the ADMIT attribution method only includes patients that were admitted by a particular physician. If a physician cares for a patient but he/she was not the admitting physician, then the physician's performance with this particular patient would not have been captured using the ADMIT attribution method when computing a quality indicator for that physician. This is important to understand because ideally, audit and feedback interventions should provide comprehensive feedback to physicians, and if not, any gap in this assessment should be noted so that the data can be interpreted in context.

Note that comprehensiveness is the property of the attribution method and is the same for all quality indicators. However, the comprehensiveness of the attribution method will depend on the model of care (e.g. how often in the model of care does a physician care for a patient that he/she did not admit), and thus may be different across different hospitals. To assess comprehensiveness, the ANY attribution method should include all patients that the physician was involved in (as admitting, discharging, or most responsible physician). Therefore, the comprehensiveness of an attribution method for *each physician* is calculated as follows:

$$\frac{number\ of\ patients\ of\ a\ physician\ included\ with\ that\ attriubtion\ method}{number\ of\ patients\ of\ a\ physician\ included\ in\ the\ ANY\ attribution\ method} \times 100\%$$

The comprehensiveness of each attribution method is then calculated as the *mean of the above quantity across all physicians in the hospital or dataset*.

Comparability refers to the degree to which an attribution method produces patient populations that are comparable across physicians (at the same hospital). This is important to understand as this has implications for benchmarking (i.e. establishing expected performance). This is also an important consideration when comparing performance between physicians (in applications such as ranking). In general medicine care, since patients are admitted from the Emergency Department (i.e. unplanned), over a large number of patients, each physician at each hospital should be exposed to similar patients due to this pseudo-randomization process (18). As such, each physician should see patients with similar baseline characteristics over a large sample size. How different attribution methods affect this pseudo-randomization is important to understand. As with comprehensiveness, this is a property of the attribution method and is the same for all quality indicators. Similarly, comparability will depend on the patient population and model of care at each hospital.

To assess comparability of the patient populations produced, the characteristics of patients produced by each attribution method for each physician will be compared against the same characteristics of all patients at that hospital. The patient characteristics being examined include age, sex, Charlson comorbidity index (31), day of admission, time of admission, previous admission to general medicine within 30 days, Laboratory Acute Physiology Score (LAPS) (32), and top 10 discharge diagnoses. For each physician and for each patient characteristic above, the mean (for continuous variable, e.g. age) and proportion (for binary variables, e.g. male patients) were computed, and the absolute difference between the physician's patients using a particular attribution method and the overall hospital patient population were noted. The median differences (and inter-quartile range) across all physicians of the same hospital / dataset were reported.

Consistency refers to the extent to which different attribution methods give similar results. It is important to understand what difference (if any) different attribution methods have on the calculated value of a quality indicator for individual physician, and the impact on relative ranking of physicians within a group. If different attribution methods produce similar results, then perhaps the simplest method will suffice. This may also increase the confidence of physicians on these indicators. However, consistency does not imply correctness. The key is to understand why different attribution methods give different results. Does one method systematically exclude certain patients (see comprehensiveness above), or is it due to other factors such as handover, communication, or teamwork? Thinking through the reasons for differences is just as important. Consistency is affected by the specific quality indicator, model of care, and patient population.

Consistency of each *pair* of attribution methods for each quality indicator was assessed in two ways. First, the *absolute difference* of the quality indicator values between the attribution methods for each physician in a hospital (or the dataset) was summarized (mean and standard deviation). Second, the *relative ranking of physicians within the same hospital* using different attribution methods were compared and summarized using spearman rank correlation. This was interpreted as weak for values < 0.4, moderate from 0.4 to < 0.7, strong from 0.7 to < 0.9, and very strong for 0.9 to 1 (33).

Generalizability refers to whether findings about attribution methods and impacts on quality indicators are applicable to most hospitals, or whether hospital specific differences impact such generalization. All analyses described above were performed at each hospital so that inter-hospital differences can be detected. Visual inspection by the author was done and differences noted.

Determining the appropriate attribution method

The result of this study will provide data to inform the selection of physician attribution method(s) for quality indicators used in general medicine audit and feedback interventions. However, the results will not by themselves indicate the appropriate attribution method, as different attribution methods will have different trade-offs. The "appropriate" attribution method will depend on many factors including acceptance, feasibility, and usability (26, 28). The results of this study may serve as one source of input for consideration by those devising measurement specifications for quality indicators planned to be used for audit and feedback interventions (26, 28).

Statistical Software

All analyses were performed using statistical software R (version 3.6.3) (34).

Chapter 3: Results

In this section, the dataset within GEMINI that met inclusion criteria for this study will be described. This will be followed by the assessment of comprehensiveness and comparability of each of the 5 attribution methods. The consistency for each attribution method for each indicator will then be presented in turn. Finally, the generalizability insights from this analysis will be presented.

Dataset Derivation

A total of 203 physicians had cared for at least 100 patients in the dataset. The entire GEMINI dataset consisted of 236,575 hospitalizations. Of these, 230,682 (97.5%) hospitalizations involved a physician of interest (one of 203) as either admitting or discharging physician. Of these, 222,490 (94% of original cohort) hospitalizations were admitted through the Emergency Department. Analysis was performed on these 222,490 hospitalizations in this study, cared for by the identified 203 physicians. Table 1 breaks down the number of physicians and hospitalizations included for each hospital.

Table 1. Summary of included physicians and patients by hospital.

Hospital	**Number of physicians included**	**Number of hospitalizations included**
A	24	24,793
B	32	38,326
C	23	28,148
D	38	26,067
E	23	45,507
F	32	31,332
G	31	28,317
All	203	222,490

Comprehensiveness of each attribution method

Using the ANY attribution method as the denominator (as this attribution method captures all patients that were either admitted, discharged, or cared for most responsibly by a particular physician), the STRICT, ADMIT, DISCHARGE, and MRP attribution methods capture a mean of 40%, 70%, 71%, and 69% of each physician's patients across all hospitals. Table 2 outlines the result per hospital as well as across all hospitals. The STRICT attribution method consistently produced the least comprehensive assessment of a physician's performance across each hospital and entire dataset. It captured about 40% of care provided, whereas other attribution methods (other than ANY) captured about 70% of care provided.

Table 2. Mean percentage of a physician's patients included by a particular attribution method (using the ANY attribution method as the denominator).

Hospital	**Number of physicians (N)**	**STRICT** Mean % (sd)	**ADMIT** Mean % (sd)	**DISCHARGE** Mean % (sd)	**MRP** Mean % (sd)	**ANY**
A	24	55% (5.5)	78% (3.3)	78% (4.2)	78% (4.5)	100%
B	32	48% (5.8)	73% (4.4)	76% (3.4)	75% (3.6)	100%
C	23	42% (3.7)	73% (3.3)	70% (3.5)	70% (3.8)	100%
D	38	35% (8.9)	67% (5.7)	69% (6.4)	68% (6.8)	100%
E	23	16% (9.1)	57% (24.1)	60% (16.4)	60% (16.1)	100%
F	32	43% (4.9)	75% (3.7)	70% (3.7)	66% (9.8)	100%
G	31	38% (15.7)	65% (24.8)	73% (8.2)	68% (13.2)	100%
All	203	40% (13.6)	70% (14.4)	71% (8.9)	69% (10.5)	100%

Comparability of each attribution method

Tables 3 to 10 display the median physician-level differences in patient characteristics when comparing each physician's attributed patient population using a particular attribution method when compared against the hospital's included patient population (of the physician's hospital). While differences between different attribution methods do exist across some patient characteristics, the absolute differences are small and may not have meaningful impacts.

For age of patients (see table 3), the median physician-level difference was highest with the STRICT attribution method (median 1.7 years, interquartile range 2.0 years).

Table 3. Median (and IQR) of physician-level absolute differences for patients' mean age in years (when comparing individual physician attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** Mean age	**STRICT** Median age difference (IQR)	**ADMIT** Median age difference (IQR)	**DISCHARGE** Median age difference (IQR)	**MRP** Median age difference (IQR)	**ANY** Median age difference (IQR)
A	66.5	1.5 (1.5)	0.7 (0.4)	0.7 (0.8)	0.8 (1.0)	0.9 (1.1)
B	72.3	1.3 (2.4)	1.1 (1.3)	1.1 (1.7)	1.2 (1.6)	1.4 (0.8)
C	64.2	2.0 (0.9)	0.3 (0.6)	1.0 (0.7)	1.0 (0.6)	0.5 (0.9)
D	69.8	3.3 (2.9)	1.1 (1.3)	1.1 (1.1)	1.1 (1.1)	0.8 (1.4)
E	70.9	0.7 (1.1)	1.0 (1.2)	1.9 (1.4)	2.0 (1.5)	0.5 (0.9)
F	64.3	1.6 (1.1)	0.6 (1.0)	0.8 (1.5)	1.0 (1.4)	0.8 (1.0)
G	69.7	1.7 (1.4)	0.4 (0.8)	0.5 (0.7)	0.5 (1.0)	0.8 (0.8)
All	68.6	1.7 (2.0)	0.8 (1.0)	1.0 (1.4)	1.1 (1.4)	0.9 (1.1)

For patient sex (see table 4), measured as percentage of patients of male sex, all attribution methods produce similar results (with median physician-level differences of between 1 to 2%).

Table 4. Median (and IQR) of physician-level percent difference of male patients (when comparing individual physician attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** Mean % male	**STRICT** Median % male difference (IQR)	**ADMIT** Median % male difference (IQR)	**DISCHARGE** Median % male difference (IQR)	**MRP** Median % male difference (IQR)	**ANY** Median % male difference (IQR)
A	46%	2% (3%)	2% (2%)	2% (2%)	2% (2%)	2% (3%)
B	48%	2% (3%)	1% (1%)	1% (2%)	1% (2%)	1% (2%)
C	57%	1% (2%)	1% (1%)	1% (1%)	1% (2%)	1% (1%)
D	45%	3% (3%)	2% (2%)	2% (2%)	2% (2%)	2% (1%)
E	50%	3% (3%)	1% (1%)	3% (3%)	3% (3%)	2% (1%)
F	52%	2% (2%)	1% (2%)	1% (2%)	2% (2%)	1% (2%)
G	51%	2% (1%)	1% (2%)	1% (2%)	1% (2%)	1% (2%)
All	49.8%	2% (3%)	1% (2%)	2% (3%)	2% (2%)	1% (1%)

For Charlson Comorbidity Score (a measure of patient complexity) (see table 5), the STRICT attribution method consistently produced the highest median physician-level difference in scores across all hospitals. However, the absolute difference in score is small (0.3).

Table 5. Median (and IQR) of physician-level differences for patient's Charlson Comorbidity Score (when comparing individual physician's attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** Mean score	**STRICT** Median score difference (IQR)	**ADMIT** Median score difference (IQR)	**DISCHARGE** Median score difference (IQR)	**MRP** Median score difference (IQR)	**ANY** Median score difference (IQR)
A	1.6	0.2 (0.1)	0.1 (0.0)	0.1 (0.1)	0.0 (0.1)	0.1 (0.1)
B	1.3	0.2 (0.1)	0.1 (0.1)	0.1 (0.2)	0.1 (0.2)	0.1 (0.1)
C	1.2	0.2 (0.1)	0.0 (0.0)	0.1 (0.1)	0.1 (0.1)	0.0 (0.1)
D	1.3	0.3 (0.2)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)
E	1.3	0.3 (0.2)	0.1 (0.1)	0.1 (0.1)	0.1 (0.2)	0.1 (0.1)
F	2.3	0.3 (0.2)	0.1 (0.1)	0.2 (0.1)	0.2 (0.1)	0.1 (0.1)
G	1.9	0.2 (0.1)	0.0 (0.1)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)
All	1.5	0.3 (0.2)	0.1 (0.1)	0.1 (0.2)	0.1 (0.1)	0.1 (0.1)

For percentage of patients admitted on weekend (see table 6), all attribution methods produce similar median physician-level differences (about 2%) and this small difference is unlikely to have impact overall. However, for hospital E, the STRICT and ADMIT attribution methods seem to produce higher median physician-level differences suggesting this may be due to local model of care (possibly involving admission workflows on weekends, including different likelihood of different physicians participating in weekend admissions).

Table 6. Median (and IQR) of physician-level differences of percentage of patients admitted on weekend (when comparing individual physician's attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** % admitted on weekend	**STRICT** Median % difference (IQR)	**ADMIT** Median % difference (IQR)	**DISCHARGE** Median % difference (IQR)	**MRP** Median % difference (IQR)	**ANY** Median % difference (IQR)
A	26%	2% (1%)	1% (1%)	0% (1%)	1% (0%)	0% (1%)
B	28%	2% (3%)	2% (1%)	2% (3%)	2% (3%)	2% (2%)
C	26%	2% (2%)	1% (2%)	1% (2%)	1% (1%)	1% (1%)
D	25%	3% (3%)	3% (4%)	2% (3%)	2% (2%)	2% (3%)
E	24%	6% (7%)	8% (7%)	2% (2%)	2% (2%)	4% (4%)
F	25%	3% (5%)	2% (3%)	2% (2%)	3% (3%)	3% (3%)
G	26%	4% (7%)	3% (3%)	4% (5%)	4% (6%)	4% (5%)
All	26%	2% (4%)	2% (4%)	2% (3%)	2% (3%)	2% (3%)

For percentage of patients admitted during day time hours (see table 7), all attribution methods produce similar median physician-level differences (4 to 5%). Each hospital appears to have different median differences (similar across different attribution methods; but different across hospitals). For example, the median differences for hospital A are between 1% and 2%, while hospital F have median differences between 7% and

8% across different attribution methods. The consistency within a hospital suggest these differences are specific to local model of care or patient population.

Table 7. Median (and IQR) of physician-level differences of percentage of patients admitted during day time (when comparing individual physician's attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** % admitted during day time	**STRICT** Median % difference (IQR)	**ADMIT** Median % difference (IQR)	**DISCHARGE** Median % difference (IQR)	**MRP** Median % difference (IQR)	**ANY** Median % difference (IQR)
A	22%	2% (3%)	1% (2%)	1% (1%)	1% (1%)	1% (2%)
B	20%	6% (3%)	6% (3%)	5% (3%)	5% (4%)	5% (3%)
C	25%	2% (4%)	2% (4%)	2% (3%)	2% (3%)	2% (3%)
D	29%	4% (6%)	3% (4%)	3% (3%)	3% (3%)	2% (3%)
E	26%	5% (12%)	4% (12%)	2% (3%)	2% (3%)	3% (4%)
F	18%	8% (4%)	8% (5%)	7% (3%)	7% (4%)	7% (4%)
G	22%	8% (10%)	8% (8%)	6% (7%)	6% (7%)	6% (8%)
All	23%	5% (7%)	4% (6%)	4% (5%)	4% (5%)	4% (5%)

For percentage of patients admitted to General Medicine in the 30 days prior to hospitalization (see table 8), all attribution methods produce similar median physician level differences (about 1%) without significant hospital variation.

For LAPS as a proxy measure for patient acuity (see table 9), the median physician-level differences were generally higher with the STRICT attribution method (1.9) compared to all other methods (0.8 to 1.0). However, a difference in LAPS score of 1.9 is still small.

Table 8. Median (and IQR) of physician-level differences of percentage of patients admitted to GIM 30 days prior to admission (when comparing individual physician's attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** % admitted to GIM in prior 30 days	**STRICT** Median % difference (IQR)	**ADMIT** Median % difference (IQR)	**DISCHARGE** Median % difference (IQR)	**MRP** Median % difference (IQR)	**ANY** Median % difference (IQR)
A	12%	1% (2%)	1% (1%)	1% (1%)	1% (1%)	1% (1%)
B	10%	1% (2%)	1% (1%)	1% (1%)	1% (1%)	1% (1%)
C	13%	1% (1%)	1% (1%)	1% (1%)	1% (1%)	1% (1%)
D	8%	2% (1%)	1% (1%)	1% (1%)	1% (1%)	1% (1%)
E	6%	1% (1%)	1% (1%)	1% (1%)	1% (2%)	1% (1%)
F	14%	1% (3%)	1% (1%)	1% (2%)	1% (2%)	1% (1%)
G	12%	2% (3%)	1% (2%)	1% (1%)	1% (2%)	1% (1%)
All	11%	1% (2%)	1% (1%)	1% (1%)	1% (1%)	1% (1%)

Table 9. Median (and IQR) of physician-level differences of LAPS (Laboratory Acute Physiology Score) of patients (when comparing individual physician's attributed population to hospital patient population) using different physician attribution methods.

Hospital	**Hospital overall population** Mean LAPS	**STRICT** Median score difference (IQR)	**ADMIT** Median score difference (IQR)	**DISCHARGE** Median score difference (IQR)	**MRP** Median score difference (IQR)	**ANY** Median score difference (IQR)
A	22.0	2.3 (2.1)	1.7 (2.7)	1.7 (3.5)	1.6 (3.5)	1.4 (2.7)
B	23.6	2.5 (1.7)	0.7 (1.9)	0.8 (0.9)	0.8 (0.9)	0.9 (1.1)
C	21.7	2.0 (1.5)	0.9 (1.4)	0.9 (1.1)	0.8 (1.1)	0.8 (1.1)
D	22.1	2.6 (1.3)	0.8 (0.9)	1.2 (1.2)	1.2 (1.3)	0.9 (0.8)
E	21.1	1.2 (1.5)	0.5 (1.1)	1.8 (1.0)	1.8 (1.0)	0.9 (1.5)
F	15.4	1.6 (1.1)	0.7 (0.7)	0.5 (1.0)	0.5 (0.9)	0.5 (0.8)
G	14.2	1.2 (1.3)	0.7 (0.6)	0.6 (0.9)	0.6 (0.8)	0.6 (0.8)
All	20.2	1.9 (1.7)	0.8 (1.2)	1.0 (1.3)	0.9 (1.4)	0.8 (1.1)

For discharge diagnoses (see table 10), the STRICT attribution method consistently produces the highest median physician-level differences across all top 10 diagnoses (0.6 to 1.2%) than other attribution methods. However, the absolute differences remain small.

Table 10. Median (and IQR) of physician-level differences in percentage of each top 10 discharge diagnosis of patients (when comparing individual physicians' attributed population to hospital patient population) using different physician attribution methods.

Diagnosis	**Overall**	**STRICT** Median % difference (IQR)	**ADMIT** Median % difference (IQR)	**DISCHARGE** Median % difference (IQR)	**MRP** Median % difference (IQR)	**ANY** Median % difference (IQR)
Heart failure	5.0%	0.9% (1.4%)	0.7% (0.9%)	0.8% (1.1%)	0.8% (1.1%)	0.6% (0.9%)
Pneumonia	4.9%	1.0% (1.4%)	0.7% (0.9%)	0.8% (1.1%)	0.9% (1.2%)	0.6% (0.8%)
UTI	4.5%	1.0% (1.4%)	0.8% (1.0%)	0.8% (1.1%)	0.9% (1.1%)	0.8% (0.9%)
COPD	4.4%	0.9% (1.2%)	0.6% (0.9%)	0.6% (1.0%)	0.7% (1.0%)	0.4% (0.8%)
Stroke	4.1%	1.2% (2.3%)	0.8% (1.7%)	0.9% (2.0%)	0.9% (1.8%)	0.8% (1.8%)
Delirium, dementia, cognitive impairment	3.3%	1.1% (1.2%)	0.5% (0.8%)	0.6% (0.9%)	0.6% (1.1%)	0.7% (0.9%)
GI bleeding	2.7%	0.7% (1.0%)	0.4% (0.6%)	0.5% (0.9%)	0.6% (0.8%)	0.4% (0.5%)
Sepsis	2.5%	0.9% (0.9%)	0.6% (0.8%)	0.6% (0.9%)	0.6% (0.9%)	0.5% (0.8%)
Diabetes and complications	2.3%	0.6% (0.9%)	0.4% (0.6%)	0.4% (0.6%)	0.4% (0.6%)	0.3% (0.4%)
GI infection	2.3%	0.6% (0.9%)	0.4% (0.6%)	0.4% (0.6%)	0.5% (0.7%)	0.3% (0.5%)

Consistency of quality indicators with different attribution methods

As the consistency between different attribution methods is specific to a particular quality indicator, results for each quality indicator are presented in turn below.

Quality Indicator 1: Average Length of Stay

Table 11 summarizes how different attribution methods affect the actual calculated average length of stay (bottom left triangle) and how they affect relative ranking of physicians (top right triangle) for this quality indicator. The STRICT attribution method consistently gives average lengths of stay that are 4.7 to 6.8 days shorter than all other attribution methods. Pairwise differences of other attribution methods ranged from 0 to 2.1 days. The STRICT methods also had the weakest rank correlation with other attribution methods (spearman rank correlation 0.27 to 0.52). DISCHARGE and MRP have the strongest correlation (0.87). The MRP attribution method has moderate to strong correlation with all other attribution methods (0.52 to 0.87).

Table 11. The mean pairwise differences in average length of stay across all physicians using each pair of attribution methods (bottom left triangle) and the spearman rank correlation (top right triangle) for each pair of attribution method.

	STRICT	**ADMIT**	**DISCHARGE**	**MRP**	**ANY**
STRICT (3.8 days)		0.47	0.45	0.52	0.27
ADMIT (8.9 days)	-5.0 days ($p < 0.01$)*		0.55	0.56	0.77
DISCHARGE (8.9 days)	-4.7 days ($p < 0.01$)*	0.4 days ($p=0.01$)*		0.87	0.76
MRP (8.8 days)	-4.7 days ($p < 0.01$)*	0.4 days ($p=0.01$)*	0.0 days ($p=0.62$)		0.67
ANY (11.3 days)	-6.8 days ($p < 0.01$)*	-1.7 days ($p < 0.01$)*	-2.1 days ($p < 0.01$)*	-2.1 days ($p < 0.01$)*	

LEGEND: Row labels contain the average length of stay for each attribution method in brackets. For mean pairwise differences in average length of stay (bottom left triangle), negative value indicates that the attribution method in column label is shorter than attribution method in row label. * p-value (<0.05) for pairwise comparison using t-test between the pairwise attribution methods. Colour (top right triangle) indicates strength of correlation: 0.5 to < 0.7 (yellow or moderate); 0.7 to < 0.9 (blue or strong); 0.9 to 1 (green or very strong).

Quality Indicator 2: Re-admission within 30 days

Table 12 summarizes the result of how different attribution methods affect percent re-admission within 30 days (bottom left triangle) and how they affect relative ranking of physicians (top right triangle) for this quality indicator. For all pairwise comparisons of attribution methods, the absolute differences are small (0% to 0.5%), although some are statistically significant. Also, physician ranking is relatively preserved comparing against all pairs of attribution methods with all showing strong or very strong correlation (0.83 to 0.97). The strongest correlation was between MRP and DISCHARGE attribution methods (0.97).

Table 12. The mean pairwise differences in percent re-admission within 30 days across all physicians using each pair of attribution methods (bottom left triangle) and the spearman rank correlation (top right triangle) for each pair of attribution method.

	STRICT	**ADMIT**	**DISCHARGE**	**MRP**	**ANY**
STRICT (12.6%)		0.88	0.88	0.89	0.83
ADMIT (13.1%)	-0.1% (p = 0.27)		0.87	0.85	0.94
DISCHARGE (13.5%)	-0.5% (p=0.01)*	-0.4% (p=0.01)*		0.97	0.95
MRP (13.5%)	-0.4% (p=0.01)*	-0.3% (p=0.03)*	0.0% (p=0.51)		0.93
ANY (13.6%)	-0.4% (p=0.03)*	-0.3% (p=0.01)*	0.0% (p=0.54)	0.0% (p=0.88)	

LEGEND: Row labels contain the average percent readmission within 30 days for each attribution method in brackets. For mean pairwise differences in percent re-admission within 30 days (bottom left triangle), negative value indicates that the attribution method in column label is less than attribution method in row label. * p-value (<0.05) for pairwise comparison using chi-square test between the pairwise attribution methods. Colour (top right triangle) indicates strength of correlation: 0.5 to < 0.7 (yellow or moderate); 0.7 to < 0.9 (blue or strong); 0.9 to 1 (green or very strong).

Quality Indicator 3: In-patient mortality

Table 13 summarizes the result of how different attribution methods affect in-patient mortality (bottom left triangle) and how they affect relative ranking of physicians (top right triangle) for this quality indicator. The STRICT attribution method consistently reports lower in-patient mortality (by 1 to 2%) against all other attribution methods. All other attribution methods had pairwise differences between 0.1 and 1%. Despite the fact that the STRICT method consistently reports a lower absolute in-patient mortality rate, the relative ranking among physicians (when compared to other attribution methods) was preserved (spearman rank correlation between 0.70 and 0.84). In fact, most pairwise comparison of attribution methods showed similar physician ranking (strong to very strong correlation) with the exception of the comparison of DISCHARGE and ADMIT attribution methods. The strongest pairwise rank correlation was between MRP and DISCHARGE (0.96).

Table 13. The mean pairwise differences in in-patient mortality across all physicians using each pair of attribution methods (bottom left triangle) and the spearman rank correlation (top right triangle) for each pair of attribution method.

	STRICT	**ADMIT**	**DISCHARGE**	**MRP**	**ANY**
STRICT (4.6%)		0.70	0.84	0.83	0.75
ADMIT (6.3%)	-1.9% ($p < 0.01$)*		0.67	0.73	0.89
DISCHARGE (5.5%)	-1.0% ($p < 0.01$)*	0.8% ($p < 0.01$)*		0.96	0.85
MRP (5.5%)	-1.1% ($p < 0.01$)*	0.8% ($p < 0.01$)*	-0.1% ($p=0.08$)		0.88
ANY (6.5%)	-2.0% ($p < 0.01$)*	-0.1% ($p=0.09$)	-1.0% ($p < 0.01$)*	-0.9% ($p < 0.01$)*	

LEGEND: Row labels contain the average in-patient mortality for each attribution method in brackets. For mean pairwise differences in in-patient mortality (bottom left triangle), negative value indicates that the attribution method in column label is less than attribution method in row label. * p-value (<0.05) for pairwise comparison using chi-square test between the pairwise attribution methods. Colour (top right triangle) indicates strength of correlation: 0.5 to < 0.7 (yellow or moderate); 0.7 to < 0.9 (blue or strong); 0.9 to 1 (green or very strong).

Quality Indicator 4: Use of advanced imaging

Table 14 summarizes the result of how different attribution methods affect the number of advanced imaging tests ordered per patient (bottom left triangle) and how they affect relative ranking of physicians (top right triangle) for this quality indicator. The STRICT attribution method consistently reports fewer advanced imaging tests per patient (0.4 to

0.5 tests fewer) than other attribution methods. Other attribution methods give consistent results with pairwise difference of 0 to 0.1 tests per patient. Despite that, all attribution methods (including STRICT) seem to preserve relative ranking of physicians with strong or very strong spearman rank correlation (0.73 to 0.96).

Table 14. The mean pairwise differences in average number of advanced imaging tests per patient across all physicians using each pair of attribution methods (bottom left triangle) and the spearman rank correlation (top right triangle) for each pair of attribution method.

	STRICT	**ADMIT**	**DISCHARGE**	**MRP**	**ANY**
STRICT (1.1 tests)		0.73	0.84	0.83	0.73
ADMIT (1.6 tests)	-0.4 tests ($p < 0.01$)*		0.78	0.73	0.92
DISCHARGE (1.6 tests)	-0.4 tests ($p < 0.01$)*	-0.0 tests (p=0.55)		0.96	0.89
MRP (1.6 tests)	-0.4 tests ($p < 0.01$)*	-0.0 tests (p=0.96)	0.0 tests (p=0.02)*		0.84
ANY (1.8 tests)	-0.5 tests ($p < 0.01$)*	-0.1 tests ($p < 0.01$)*	-0.1 tests ($p < 0.01$)*	-0.1 tests ($p < 0.01$)*	

LEGEND: Row labels contain the average number of advanced imaging tests per patient for each attribution method in brackets. For mean pairwise differences in average number of advanced imaging tests per patient (bottom left triangle), negative value indicates that the attribution method in column label is less than attribution method in row label. * p-value (<0.05) for pairwise comparison using t- test between the pairwise attribution methods. Colour (top right triangle) indicates strength of correlation: 0.5 to < 0.7 (yellow or moderate); 0.7 to < 0.9 (blue or strong); 0.9 to 1 (green or very strong).

Generalizability

For generalizing ***comprehensiveness*** (see table 2), the STRICT attribution method consistently produced the least comprehensive assessment of a physician's performance across each hospital. Although the exact comprehensiveness differs between hospitals, this pattern is consistent across all hospitals, making this result generalizable.

For generalizing of ***comparability***, tables 3 to 9 display the overall and per hospital analysis of physician-level differences of each attribution method for age, sex, Charlson comorbidity index, percent admitted on weekend, percent day-time admission, percent admitted to GIM in prior 30 days, and LAPS for acuity respectively. For patient characteristics (age, sex, Charlson comorbidity index, percent admitted to GIM in prior 30 days, and LAPS), there were not any clear differences between hospitals.

However, for characteristics that may be affected by models of care and local workflow (percentage of weekend admissions and percentage of day time admissions), there were observed differences. For physician-level differences in percentage of patients admitted on weekend (table 6), the STRICT and ADMIT attribution methods produce physician-level patient population that were more dissimilar at hospital E than at other hospitals (hospital E had mean difference of 6 and 8%, whereas other hospitals were in the range of 2% to 3%). For other attribution methods, hospital E had similar physician-level differences as other hospitals. Also, for percentage of admission during day time (table 7), all attribution methods seem to produce similar discrepancies at each hospital but the actual amount of dis-similarity at the physician-level seem to differ between hospitals (e.g. hospital A had the lowest physician-level differences at 1-2% whereas hospital F had the most physician-level differences at 7-8%).

Chapter 4: Discussion

Based on 222,490 general medicine hospitalizations cared for by 203 physicians over 7 years at 7 hospitals, the STRICT physician attribution method was the least comprehensive (captured only about 40% of patients cared for by a physician in some capacity). However, all attribution methods produced patient populations for individual physicians that were comparable to those seen at each hospital. Although minor differences did exist (maximum differences 1.7 years for age, 2% for patient sex, 0.3 for Charlson comorbidity, 1% for admission to general medicine in prior 30 days, 1.9 for LAPS), these small differences are unlikely to have meaningful impact. For the quality indicator average length of stay, the STRICT attribution method consistently produced length of stay values that were 4.7 to 6.8 days shorter than other attribution methods and also had poor rank correlation of physicians (spearman rank correlation 0.27 to 0.52) against all other attribution methods. For the other quality indicators (re-admission within 30 days, in-patient mortality, and rate of advanced imaging), although there were differences between different physician attribution methods, the absolute differences tended to be small, with most pairwise comparisons of different attribution methods showing strong or very strong rank correlation of physicians. Looking across hospitals, local model of care differences may have led to patient characteristics differences generated by different physician attribution methods, but this appeared to be affecting only those characteristics that were related to such processes (% of admission during weekend, % of day time admission), but not to other characteristics (e.g. age, sex, comorbidity). For percentage admission during weekend, model of care at hospital E may have led to different level of involvement in weekend admissions for individual physicians within their group. For percentage of daytime admissions, differences between hospitals may be reflective of local admission processes (e.g. a dedicated day time admission team) at different hospitals.

Contribution to the Literature

The effect of physician attribution methods on quality indicators have been previously reported. For example, Mehrotra and colleagues found that based on four US commercial insurance health plans data (not exclusive to hospital care), 17% to 61% of physicians would have been assigned to a different cost category based on which one of 12 attribution rules used (35). Moreover, the percentage of episodes of care assigned to a physician (similar to the comprehensiveness measured used in this study) varied substantially between the 12 attribution rules ranging from 20% to 69% (35). Chang and colleagues examined attribution methods in the US primary care context and found that for low value cervical cancer screening, attribution methods affected both the absolute quality indicator value and physician performance ranking (36). Dowd and colleagues examined physician attribution in the specific context of the US Medicare Physician Quality Reporting System (PQRS) and found that the physician who provided most out-patient visits to a particular patient supplied only 50% of such visits on average, making physician attribution challenging (37). These three studies were all in the US and mostly ambulatory (not exclusively hospital care) context, with key source of attribution based on physician billing data.

This study contributes new knowledge to the literature by systematically examining the effects of using different physician attribution methods on the apparent performance of physicians on 4 quality indicators in the general medicine in-patient setting (as opposed to the previous studies that mostly examine ambulatory performances) (35-37). Previous studies in the ambulatory settings have described that some attribution methods have poor comprehensiveness in capturing care provided by a particular physician with estimates of 20% to 69% in one study (35) and 50% in another (37). This study shows that similar concerns exist for some attribution methods in the in-patient setting as well (STRICT attribution method 40%; ADMIT, DISCHARGE, MRP 70%).

Studies in the ambulatory setting have shown that choice of attribution methods seem to have affected apparent physician performance substantially (35, 36). While this may be true for some quality indicators in the in-patient setting (e.g. STRICT attribution method for average length of stay), this study shows that for other quality indicators (re-admission in 30 days, in-patient mortality, use of advanced imaging), the absolute differences tended to be small and the relative ranking among physicians seemed to be preserved between different attribution methods. One possible explanation is that assignment of patients to physicians in general medicine in-patient care is a pseudo-randomized process (as these are unplanned hospitalizations from the Emergency Department with physician assignment based on work schedule) (15, 18). Any systematic bias introduced by the attribution method may affect all physicians equally.

Selecting an appropriate attribution method

The findings of this study were meant to inform the selection of attribution methods for those involved in designing and implementing physician audit and feedback interventions for general medicine in-patient care. As demonstrated, there are trade-offs to each attribution method.

Trade-offs for the STRICT attribution method

The STRICT method was intended to capture care reasonably attributable to only a single physician (the admitting, discharging, and most responsible physicians were the same individual and the length of stay was relatively short such that repeated changeover of physicians is unlikely to meet this STRICT condition). It has the ***theoretical advantage*** of physicians believing that the feedback is reflective of their own performance (as no one else was involved in the care of these patients). Previous studies suggested physician's trust in that data as their own is important for audit and feedback interventions to be successful and to be able to drive quality change (9-12). In fact, this was the attribution method used to deliver the first round of General Medicine MyPractice Reports (20) to physicians at 7 Ontario hospitals in 2019. However, this study showed that this attribution method was the ***least comprehensive***, capturing a mean of only 40% of patients that a physician may have cared for (table 2). This means that 60% (majority) of the care provided by a physician would not have contributed to the calculation of a physician's performance.

The STRICT method may select a subset of each physician's patients that are systematically different than other patients. At the physician level, the age, Charlson comorbidity index, and LAPS using the STRICT attribution method do differ the most from the hospital's overall patient population when compared to all other attribution methods. However, their absolute differences are small and may not lead to meaningful impact. Differences in patients' sex, percent admitted on weekend, percentage of day time admissions, percentage admitted to GIM in previous 30 days do not seem to differ from other attribution methods. In terms of diagnoses (table 10), the STRICT attribution method differs the most (1.2% difference) for the group of conditions defined by delirium, dementia, and cognitive impairment. It may be the case that patients with these diagnoses have generally longer length of stay (38) and may less likely to be include by the STRICT attribution method due to its restriction on length of stay.

In terms of specific quality indicators, the STRICT attribution method resulted in average length of stay that were 4.7 to 6.8 days shorter at the individual physician level than when other attribution methods were used to calculate the average length of stay for physicians (see table 11). This is not surprising as the STRICT attribution method includes hospitalizations that had a short length of stay (less than 14 days). However, the ranking of physicians per hospital using the STRICT attribution method also had poor correlation with other attribution methods. This may suggest a physician's performance for length of stay using the STRICT attribution method may be different than if calculated by other methods.

For the other three quality indicators, the STRICT attribution method at the physician level tend to have minimally lower re-admission rates (by 0.1 to 0.5%), lower in-patient mortality (by 1 to 2%), and less use of advanced imaging (by 0.4 to 0.5 tests per hospitalization) than with other attribution methods. This may be due to the specific patient characteristics captured using the STRICT attribution method. However, unlike with length of stay, the STRICT attribution methods have strong correlation in terms of physician ranking when compared to other attribution methods, indicating, that it may still be a fair representation of a physician's general performance.

Trade-offs for the MRP attribution method

The most responsible physician (MRP) has been defined by Canadian Institute for Health Information (CIHI) to be the physician who is responsible for the longest duration of stay or greatest resource use (22). It is assigned by a trained health information professional by reviewing the hospital record after a patient's discharge. There is, therefore, an ***intuitive appeal*** to using this attribution method when computing quality indicators for audit and feedback intervention as the attributed physician should be the "most responsible" one among all physicians involved as determined by a trained reviewer (and may be perceived as less arbitrary).

In this study, the MRP attribution method was more comprehensive than the STRICT method, and performed similarly to the ADMIT and DISCHARGE attribution methods, all capturing about 70% of patients a physician was known to be involved in in any capacity (table 2). The MRP attribution method also produces, for individual physicians, patient populations that are quite similar to hospital's overall population in all patient characteristics examined (see tables 3 to 10). The quality indicators generated may be appropriate for benchmarking.

For specific quality indicators, it has strong or very strong correlation with other attribution methods for readmission within 30 days (rank correlation 0.85 to 0.97, with absolute difference between 0 and 0.4%), in-patient mortality (rank correlation 0.73 to 0.96, with absolute difference between 1.1% and 0.1%), and use of advanced imaging (rank correlation 0.73 to 0.96 with absolute difference of 0.4 to 0 tests per hospitalization). It may suggest that this attribution method reasonably capture a physician's performance. In terms of average length of stay, like all other attribution methods, it produces different results than STRICT (a mean of 4.7 days longer across all physicians, rank correlation 0.52) as explained above. It is most consistent with DISCHARGE (absolute difference 0 days, rank correlation 0.87). In most cases, the discharging physician decides when the patient leaves the hospital (and may influence length of stay this way) but this in itself may not explain this completely as the discharging physician may not have influence over the length of stay that has already been accrued prior to his/her involvement in patient care.

Trade-offs for the ANY attribution method

The ANY attribution method is the most comprehensive as it captures all patients that the physician cared for in any role (admitting, discharge, or most responsible). However, in reality, this does not capture all patients that the physician has had an involvement as the administrative roles (admitting, discharging, most responsible) only capture a subset of care providers. For example, if a patient has had an 8 week stay, and physicians changed every 2 weeks, that there may actually be 4 physicians involved but the administrative data will only capture a subset (at most 3) of these. In addition, there is often overnight and weekend coverage that may not be captured by administrative datasets at all. Nevertheless, it is the most comprehensive of all the attribution methods considered in this study.

In terms of comparability, the ANY attribution method produces patient population for each physician that are quite similar to that of the local hospital for all characteristics examined.

In terms of specific quality indicators, it produces results consistent with other attribution methods for re-admission within 30 days (rank correlation 0.83 to 0.95; absolute difference of 0 to 0.4%). For in-patient mortality and use of advanced imaging, it is consistent with all other attribution methods in terms of ranking of physician (rank correlation 0.75 to 0.89 and 0.73 to 0.92 respectively) but differ the most from the STRICT method (absolute difference of 2% in in-patient mortality and 0.5 tests per patient for used of advanced imaging). For average length of stay, it did differ from STRICT, but also from other attribution methods (1.7 days longer than ADMIT, 2.1 days longer than DISCHARGE, and 2.1 days longer than MRP) with generally poor rank correlation (0.27 to 0.77).

Attribution methods in context across hospitals

The patients captured by each attribution method (comprehensiveness) are sensitive to local models of care. In our study, we included 7 hospitals, with a mix of academic and community teaching hospitals in the Greater Toronto Area. There are differences in model of care. For example, some models of care involve a dedicated physician in the Emergency Department admitting patients rather than the attending physician taking turn admitting patients from the Emergency Department. In the former case, the admitting physician may less likely to also be the most responsible and discharging physicians, thereby making the STRICT method have less comprehensiveness at these hospitals. From table 2, hospitals D and E have particular low comprehensiveness (35% and 16% respectively) and likely reflect model of care differences.

Despite this, it is reassuring to know that patient populations at the physician level produce by the different attribution methods are comparable to the overall local hospital patient population for most patient characteristics (age, sex, Charlson Comorbidity Index, admission to GIM 30 days prior, illness severity), except for those expected to be associated with model of care differences (percentage of patients admitted on weekends and day time admissions).

Informing future physician audit and feedback interventions

This study highlights the need to consider physician attribution methods when implementing physician audit and feedback interventions. This is especially true in contexts where attribution methods are not straight forward (e.g. as in the case of general medicine in-patient care where multiple physicians are involved to varying degrees and their involvement change over time). The result of this study highlights that there are trade-offs with each attribution method and that the “optimal” method likely depends on a number of data and non-data related factors.

Although the analyses performed in this study do not in themselves answer the question of what is the most appropriate physician attribution method, they certainly produce results that are very valuable in informing these discussions. For example, in the case of the General Medicine Quality Improvement Network (GeMQIN) in Ontario, our results suggest that the current method of physician attribution method (STRICT) captures only a small proportion of patient care delivered by physician to the population of interest. For hospitals with certain model of care (e.g. those with dedicated admission physician who do not become the attending physician of the patient), this is even more true. These analyses with real data also show that there are other viable attribution methods that have improved comprehensiveness, similar comparability and consistency with existing attribution method (e.g. MRP as described above). The GeMQIN program has therefore decided that in the next iteration of MyPractice Report (scheduled to be released in early 2021) to use the MRP attribution method for the two hospitals that showed particular low comprehensiveness as shown in this analysis.

The results also highlight that model of care differences do affect performance of attribution methods. This raises the question of whether attribution method should be different based on local contextual factors. If so, how does this affect the use of these quality indicators for benchmarking across hospitals with different models of care.

The results also indicate that attribution methods can have different performance for different quality indicators. In the analyses presented, length of stay is a particular problematic one. The attribution method used can generate vastly different results and the results from different attribution methods for the same physician are not consistent. It raises the question of what the indicator truly measures. It also raises the issue of whether each indicator should use an attribution method that is most appropriate, and that different quality indicators of the same intervention may need different attribution methods. If so, does this complexity add to the challenge of interpretation by the physician and whether that is actually helpful or not. These questions remain unanswered.

On the other hand, the result of this study can also be used by those implementing audit-and-feedback interventions to reassure physicians that for most indicators (with the exception of length of stay), different attribution methods for general medicine in-patient care across multiple hospitals do seem to produce consistent results (in terms of physician ranking) and only minor differences in quality indicator values.

How to incorporate information produced by this study into the design process of future audit and feedback interventions, possibly using consensus methods (26), is an area of future research.

Strengths and Limitations

The strengths of this study include use of a large data set (more than 222,000 hospitalizations) from 7 hospitals over 7 years that reflected data generated through real patient care. The 7 hospitals are a mix of academic and community teaching hospitals with a variety of models of care, which made comparing performances of attribution methods with different models of care possible. The methods employed in this study can be applied to other audit-and-feedback interventions, organizations, and quality indicators to inform selection of appropriate physician attribution methods of future audit-and-feedback interventions.

Limitations of this study include the fact that attribution methods considered are only the ones feasible within the GEMINI dataset (which consists of both administrative and clinical data). Other studies on physician attribution have used billing data to ascertain physician involvement in care (35-37) but were not available within the GEMINI dataset. Also, even though the dataset was derived from 7 hospitals with different models of care, these hospitals were all in the urban or sub-urban setting. It is unclear whether the findings of this study are generalizable to other provinces in Canada, or to rural areas where the model of care may be different still (e.g. involving patient's own community family physician).

Chapter 5: Conclusion

In general medicine in-patient care where multiple physicians are involved with varying degree of responsibilities that change over time, physician attribution method used when implementing an audit and feedback intervention deserve thoughtful considerations. Different physician attribution methods can produce different results for individual physician.

For general medicine in-patient care, different physician attribution methods have different comprehensiveness in capturing care provided by an individual physician, but all seem to produce similar patient populations suitable for comparison with only minor differences. The quality indicators of 30-day re-admission, in-patient mortality, and use of advanced imaging seem to produce results with only minor differences and consistent physician ranking when different physician attribution methods were used. The quality indicator of length of stay did not.

Understanding the comprehensiveness, comparability, consistency, and generalizability of the attribution method(s) used can inform thoughtful decision of the appropriate attribution method for a particular audit and feedback intervention. How to incorporate such information into the design process of audit and feedback interventions is an area of future research.

Bibliography

1. de Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua. 2009;21(2):119-29.

2. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012(6):CD000259.

3. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503-27.

4. James B. Information system concepts for quality measurement. Med Care. 2003;41(1 Suppl):I71-9.

5. Persell SD, Wright JM, Thompson JA, Kmetik KS, Baker DW. Assessing the validity of national quality measures for coronary artery disease using an electronic health record. Arch Intern Med. 2006;166(20):2272-7.

6. Roth CP, Lim YW, Pevnick JM, Asch SM, McGlynn EA. The challenge of measuring quality of care from the electronic health record. Am J Med Qual. 2009;24(5):385-94.

7. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. Implementation science : IS. 2017;12(1):61.

8. Patel S, Rajkomar A, Harrison JD, Prasad PA, Valencia V, Ranji SR, et al. Next-generation audit and feedback for inpatient quality improvement using electronic health record data: a cluster randomised controlled trial. BMJ Qual Saf. 2018;27(9):691-9.

9. Brehaut JC, Colquhoun HL, Eva KW, Carroll K, Sales A, Michie S, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. Annals of internal medicine. 2016;164(6):435-41.

10. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. Implementation science : IS. 2006;1:9.

11. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implementation science : IS. 2014;9:14.

12. Paciorkowski N, Pruitt C, Lashly D, Hrach C, Harrison E, Srinivasan M, et al. Development of performance tracking for a pediatric hospitalist division. Hosp Pediatr. 2013;3(2):118-28.

13. Xu T, Makary MA, Al Kazzi E, Zhou M, Pawlik TM, Hutfless SM. Surgeon-Level Variation in Postoperative Complications. J Gastrointest Surg. 2016;20(7):1393-9.

14. Gutacker N, Bloor K, Bojke C, Walshe K. Should interventions to reduce variation in care quality target doctors or hospitals? Health Policy. 2018;122(6):660-6.

15. Tsugawa Y, Jha AK, Newhouse JP, Zaslavsky AM, Jena AB. Variation in Physician Spending and Association With Patient Outcomes. JAMA Intern Med. 2017;177(5):675-82.

16. Krumholz HM, Keenan PS, Brush JE, Bufalino VJ, Chernew ME, Epstein AJ, et al. Standards for measures used for public reporting of efficiency in health care: a scientific statement from the American Heart Association Interdisciplinary Council on Quality of Care and Outcomes research and the American College of Cardiology Foundation. J Am Coll Cardiol. 2008;52(18):1518-26.

17. Hysong SJ, Smitham K, SoRelle R, Amspoker A, Hughes AM, Haidet P. Mental models of audit and feedback in primary care settings. Implementation science : IS. 2018;13(1):73.

18. Verma AA, Guo Y, Jung HY, Laupacis A, Mamdani M, Detsky AS, et al. Physician-level variation in clinical outcomes and resource use in inpatient general internal medicine: an observational study. BMJ Qual Saf. 2021;30(2):123-32.

19. Health Quality Ontario. Ontario General Medicine Quality Improvement Network (GeMQIN) 2021 [Available from: https://www.hqontario.ca/Quality-Improvement/Quality-Improvement-in-Action/The-General-Medicine-Quality-Improvement-Network.

20. Health Quality Ontario. M Practice: General Medicine 2021 [Available from: https://www.hqontario.ca/Quality-Improvement/Practice-Reports/MyPractice-General-Medicine.

21. Verma AA, Guo Y, Kwan JL, Lapointe-Shaw L, Rawal S, Tang T, et al. Patient characteristics, resource use and outcomes associated with general internal medicine hospital care: the General Medicine Inpatient Initiative (GEMINI) retrospective cohort study. CMAJ Open. 2017;5(4):E842-E9.

22. CIHI. Discharge Abstract Database (DAD) Metadata 2016 [Available from: https://www.cihi.ca/en/types-of-care/hospital-care/acute-care/dad-metadata.

23. Pannick S, Davis R, Ashrafian H, Byrne BE, Beveridge I, Athanasiou T, et al. Effects of Interdisciplinary Team Care Interventions on General Medical Wards: A Systematic Review. JAMA Intern Med. 2015;175(8):1288-98.

24. Choosing Wisely Canada. Eleven things physicians and patients should question (Internal Medicine) 2020 [Available from: https://choosingwiselycanada.org/internal-medicine/.

25. Stelfox HT, Straus SE. Measuring quality of care: considering measurement frameworks and needs assessment to guide quality indicator development. J Clin Epidemiol. 2013;66(12):1320-7.

26. Stelfox HT, Straus SE. Measuring quality of care: considering conceptual approaches to quality indicator development and evaluation. J Clin Epidemiol. 2013;66(12):1328-37.

27. Campbell SM, Kontopantelis E, Hannon K, Burke M, Barber A, Lester HE. Framework and indicator testing protocol for developing and piloting quality indicators for the UK quality and outcomes framework. BMC Fam Pract. 2011;12:85.

28. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua. 2003;15 Suppl 1:i5-11.

29. Perera R, Dowell A, Crampton P. Painting by numbers: a guide for systematically developing indicators of performance at any level of health care. Health Policy. 2012;108(1):49-59.

30. Rubin HR, Pronovost P, Diette GB. From a process of care to a measure: the development and testing of a quality indicator. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua. 2001;13(6):489-96.

31. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173(6):676-82.

32. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. Med Care. 2008;46(3):232-9.

33. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg. 2018;126(5):1763-8.

34. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2017.

35. Mehrotra A, Adams JL, Thomas JW, McGlynn EA. The effect of different attribution rules on individual physician cost profiles. Annals of internal medicine. 2010;152(10):649-54.

36. Chang E, Buist DS, Handley M, Pardee R, Gundersen G, Reid RJ. Physician Service Attribution Methods for Examining Provision of Low-Value Care. EGEMS (Wash DC). 2016;4(1):1276.

37. Dowd B, Li CH, Swenson T, Coulam R, Levy J. Medicare's Physician Quality Reporting System (PQRS): quality measurement and beneficiary attribution. Medicare Medicaid Res Rev. 2014;4(2):doi: 10.5600/mmrr.004.02.a04.

38. Fick DM, Steis MR, Waller JL, Inouye SK. Delirium superimposed on dementia is associated with prolonged length of stay and poor outcomes in hospitalized older adults. J Hosp Med. 2013;8(9):500-5.

Appendix A: Ethics Approval

University of Victoria
Office of Research Services I Human Research Ethics Board
Michael Williams Building Rm B202 PO Box 1700 STN CSC Victoria BC V8W 2Y2 Canada
T 250-472-4545 I F 250-721-8960 I uvic.ca/research I ethics@uvic.ca

Certificate of Approval - Annual Renewal

PRINCIPAL INVESTIGATOR	**Francis Lau** (Supervisor)	**ETHICS PROTOCOL NUMBER** Chair/Vice-chair - delegated	**19-0208**
PRINCIPAL APPLICANT	**Terence Tang** **Master's student**	ORIGINAL APPROVAL DATE	16-Jul-2019
UVIC DEPARTMENT	**Health Information Science**	APPROVED ON	22-Jun-2020
		APPROVAL EXPIRY DATE	15-Jul-2021

PROJECT TITLE **How do different methods of attributing patients to physicians affect quality indicators derived from electronic health record data for general medicine physicians**

RESEARCH TEAM MEMBERS
Jay (Jung Hae) Young - Co-Investigator, St. Michael's Hospital
Amol Verma - Co-Investigator, University of Toronto and St. Michael's Hospital
Adina Weinerman - Co-Investigator, University of Toronto and Sunnybrook Health Sciences Centre
Janice Kwan - Co-Investigator, University of Toronto and Sinai Health System
Shail Rawal - Co-Investigator, University of Toronto and University Health Network
Lauren Lapointe-Shaw - Co-Investigator, University of Toronto and University Health Network
Fahad Razak - Co-Investigator, University of Toronto and St. Michael's Hospital
Scott MacDonald - Co-supervisor, University of Victoria

DECLARED PROJECT FUNDING **None**

DOCUMENTS INCLUDED IN THIS APPROVAL
thpReb.pdf - 21-Jun-2019
GEMINI CTO approval 2018.pdf - 21-Jun-2019
MSH GEMINI Renewal Approval 2018.pdf - 21-Jun-2019

CONDITIONS OF APPROVAL

This Certificate of Approval is valid for the above term provided there is no change in the protocol.

Modifications
To make any changes to the approved research procedures in your study, please submit a “Request for Modification” form. You must receive ethics approval before proceeding with your modified protocol.

Renewals
Your ethics approval must be current for the period during which you are recruiting participants or collecting data. To renew your protocol, please submit a “Request for Renewal” form before the expiry date on your certificate. You will be sent an emailed reminder prompting you to renew your protocol about six weeks before your expiry date.

Project Closures
When you have completed all data collection activities and will have no further contact with participants, please notify the Human Research Ethics Board by submitting a “Notice of Project Completion” form.

Certification

This certifies that the UVic Human Research Ethics Board has examined this research protocol and concluded that, in all respects, the proposed research meets the appropiate standards of ethcis as outlines by the University of Victoria Research Regulations Involving Human Participants.

.