Article details: 2017-0097	
Title	Patient characteristics, resource utilization, and outcomes associated with general internal medicine hospital care from 2010 to 2015 – The General Medicine Inpatient Initiative (GEMINI) Retrospective Cohort Study
Authors	Amol A. Verma MD MPhil, Yishan Guo MSc, Janice L. Kwan MD MPH, Lauren Lapointe-Shaw MD, Shail Rawal MD MPH, Terence Tang MD, Adina Weinerman MD MHSc, Peter Cram MD MBA, Irfan A. Dhalla MD MSc, Stephen W. Hwang MD MPH, Andreas Laupacis MD MSc, Muhammad M. Mamdani PharmD MPH, Steven Shadowitz MDCM MSc, Ross Upshur MD MSc, Robert J. Reid MD PhD, Fahad Razak MD MSc
Reviewer 1	Zachary Bouck
Institution	Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, Ont.
General comments (author response in bold)	1. Would have been interesting to investigate differences in patient populations, outcomes, and care utilization by site (n = 7) or at least by organization (n = 5, given the clustered nature, e.g. both TGH and TWH fall under UHN). Could minimally compare variability in demographics and outcomes among sites and/or organizations via coefficient of quartile deviation (alternative to coefficient of variation where data is skewed and likely non-normal in its distribution). **We agree that this would be interesting but we deliberately chose not to make comparisons between hospital sites in this manuscript for the following reasons: 1. We feel that there is already a substantial amount of data being presented in this manuscript and there is insufficient space for a thoughtful treatment of the comparison between sites. 2. Given that there are only 7 participating sites, there is insufficient sample to perform site-level comparisons in a multi-level model (patients nested within providers nested within hospitals). There are likely to be differences between the sites with respect to demographics and outcomes using a statistic like the coefficient of quartile deviation, given the differences in populations served (e.g. St. Michael's is an urban inner-city facility, Sunnybrook is an urban center that caters to a relatively wealthier population, and Trillium is a community hospital). However, to carefully understand hospital-level differences, we would need to adjust for differences in patients and providers between sites, ideally in a multi-level model. We reviewed statistical approaches with a methodologist with expertise in multi-level modeling (SV Subramanian at the Harvard School of Public Health) who felt that a random effect at the hospital level given marked site heterogeneity would not be appropriate without a larger sample of hospitals (i.e. level 3 variables). Thus, we felt it would be unfair to present hospital-level differences and we feel that this falls outside the scope of our current manuscript.** Specific comments: 2. Introduction (p.5, lines 88-89) - this sentence needs work, fairly awkward. Would also recommend using a different transition word as 'thus' is used shortly after in line 96. **Have deleted.** 3. Methods (p.6, line 106) - specify retrospective 'cohort' study. **Revised. Line 107.** 4. Methods under data collection heading (p.7, line 133) - unclear how the data was standardized (adjusted for which factors)? Or is this a reference to the fact that the administrative health data is routinely collected across these sites in the same manner, in which case I would use a different adjective versus 'standardized'. **Yes, this refers to the fact that administrative data is collected with standardized definitions across sites. Given the potential for confusion, we have simply deleted the word 'standardized'.** 5. Methods, under data collection heading (p.7, line 138-140) - potentially address this in limitations under Discussion. Possibility that some of these admissions were for the same patient and may have qualified as re-admissions within 30 days of discharge (i.e. would count towards numerator for one of your outcomes) you had the patient identifying information. As a result, you may have underestimated some outcomes. Could have alternatively excluded these 1,678 hospitalizations; however, given the large sample size, the effect on the descriptive statistics presented is likely minimal. **Included discussion of this in limitations section (lines 349-352).** 6. Measures and Statistical Analysis (p.9, line 175)- very minor, but redundant to write 'CIHI DAD database' as DAD stands for Discharge Abstract Database. **Revised.** 7. Measures and Statistical Analysis (p.10, line 197-199) - mentioned ECGs and TTEs earlier (p. 8, lines 157-158) but they are not listed among the outcomes captured. Would remove reference to these tests from the earlier section as they were not captured and are not mentioned after p.8. **Deleted reference to these.** 8. Measures and Statistical Analysis (p.11, lines 219-220) - despite the simplicity of analysis (descriptive only), should state how variables were summarized (i.e. use of median and interquartile range to summarize continuous variables due to skewed distributions). **Revised.** 9. Results (p.11, line 223) - typo in heading, should read 'Demographics and Coexisting Medical Conditions'. **Revised.** 10. Results, under patient outcomes and resource utilization (p.12, line 243) - interventional radiology procedure was neither referenced nor previously defined in methods (p. 10, lines 196-199). Revise methods to list interventional radiology procedure as an outcome in the appropriate section and define what procedures are included. **Revised.** 11. Results (p. 12, lines 246-247) - possibly incorporate this lone sentence into an earlier paragraph. **Revised.** 12. Results (p.12, line 251) - cut 'by', should be 'proportion of all hospital patients cared for by GIM increased between 10.4% and...'. **Revised.** 13. Results (p.12, line 253) - very minor again, would recommend changing to 'median acute length-of-stay' for

specificity. **Revised.**

14. Results (p.12, line 249-259) - would have been interesting to test these outcomes via regression to see if the changes over time were statistically significant.
We considered this, but specifically chose not to perform inferential statistical testing because this was not a hypothesis-driven question and we felt it would be inappropriate to report statistical significance. We therefore chose to only report descriptive statistics.

15. Results (p.13, lines 275-276) - not an average (mean) value for either the length of stay or cost per hospitalization variables. Please change 'average' to 'median' in both instances so it reads "longest median acute length-of-stay…and the highest median cost per hospitalization". **Revised.**

16. Interpretation (p.14, line 286) - could potentially argue that your results, in showing a small minority of GIM patients belonging to these CMG use a disproportionately high percentage of resources, should be selectively targeted by a potential intervention versus apply a broader intervention to the entire GIM inpatient population. May be a more cost-effective intervention with high return on investment.
This is a good point. Given the requested structure of this section (ie. paragraph 1 summarizes main findings), we have removed the discussion of the implications for quality improvement here.

17. Interpretation (p.14, line 292) - very minor, spell out 2 and 7. **Revised.**

18. Interpretation (p.14, line 295-299) - "with 6 coexisting conditions on average". Again, your median value was 6, this does not represent the mean/average number of comorbidities but rather the 50% percentile. **Revised.**

19. Interpretation (p.15, line 309) - remove "average". May revise to say that interquartile ranges and frequency counts alone are insufficient to characterize the full extent of the observed variability. It would have been interesting to create a model adjusted for site (possibly a multilevel model with hospitalization-level data and a random intercept for site) to explore the variability in these outcomes by site.
We have revised the language of this line. See our response to comment 1 describing why we felt it was not appropriate to explore variability in outcomes by site. This would be interesting, but our dataset, with only 7 sites, is not well suited to site-level comparisons.

20. Interpretation (p.14, line 314-316) - I think this should be emphasized more. Again, it may be different among sites. Some of the sites may have outperformed others in this regard, for e.g. some of the sites may not have adjusted well to the influx of GIM patients and had increased usage and poorer patient outcomes as a result. However, since the data is pooled across sites, these inferences and insights cannot be made.
We agree that further exploration of these trends over time would be interesting. These are the very avenues for further research that we believe our current manuscript illuminates. We state, "further work is needed to understand the drivers of changes in resource utilization over time." (lines 331-332)

Reviewer 2	Mr. Charlie Tan
Institution	Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Ont.
General comments (author response in bold)	1. GIM as a system of care likely varies significantly at teaching hospitals compared to non-teaching hospitals. At medium to large community centres (e.g., Kitchener, Guelph, Niagara region), Internal Medicine may act as a consult service to a family medicine based hospitalist service, or admit more complicated patients requiring a higher level of care (separate from a hospitalist service). Therefore, the findings here are not necessarily representative of all GIM hospital patients. Greater consideration should be given to the generalizability of the study to non-CTU GIM services. **We agree that this is an important limitation. One point of clarification: we have defined 'GIM' as general internal medicine hospital services that include both internal medicine and family medicine, as described in the participants section of the methods (lines 127-132). We suggest that our study may be generalizable to large community hospitals because 2 of the 7 participating sites (ie. the Trillium Health Partners hospitals) are large community hospitals. Thus, we write in the limitations section: "Third, our study was conducted in 7 teaching hospitals in Canada's largest metropolitan area. Our patient sample is broadly similar to other studies in GIM in Canada, the US, and Europe(29–32) and is likely generalizable to other urban and suburban academic health centres and large community hospitals, but may not be representative of patients or practices in other settings." (lines 343-347)** 2. Depending on the admitting subspecialty services at each hospital, the diagnoses admitted to the GIM service may differ. For example, if a hospital has an admitting Neurology service, no active strokes would be admitted to the GIM service. Similarly, hospitals with an admitting Cardiology service may see fewer CAD/CHF on the GIM service. Were there any differences in the most common discharge CMGs between the participating hospitals? **At one institution (University Health Network), there were relatively few stroke patients because they are predominantly cared for by the neurology service. However, apart from this difference there were no major differences in the most common CMGs between participating hospitals.** 3. May have been interesting to look at the most resource-intensive diagnoses (above a certain threshold prevalence) along with the most prevalent conditions. **This would have been an interesting approach. We felt that it was advisable to focus on the most prevalent conditions because they are the least likely to suffer from misclassification resulting from poor validity of administrative diagnostic codes.**