



## Article Appraisal

**Article:** The Effect of Pay for Performance in the Emergency department on Patient Waiting Times and Quality of Care in Ontario, Canada: A Difference-in-Differences Analysis

**Date of Journal Club:** Tuesday, February 20, 2018

**Resident Reviewer Name(s) and Residency Affiliation:** Dr. Kali Kennedy, CCFP-EM R3 and Dr. Nathan Stefani, RCPS-EM R1

**Faculty Methodology/Bio-statistics Resource Person:** Dr. Les Vertesi

---

### Background and Study Objective(s):

As wait times escalate in emergency rooms across the country and the issue becomes increasingly publicized, various possible solutions have been investigated. One such solution is pay-for-performance (P4P), an incentive strategy designed to financially award health care institutions that achieve specific outcomes in department efficiency. From 2007 to 2011, the Ontario Ministry of Health and Long-Term Care (MOHLTC) implemented a P4P program involving numerous emergency departments (ED) and hospitals across Ontario. This study sought to assess whether P4P was effective in its aim of reducing ED length of stays (LOS) and whether such a program was associated with any deficits in the quality of patient care.

### Study Design:

This was a retrospective observational study. The program itself consisted of 3 waves, each spanning a single fiscal year, and each hospital that enrolled stayed enrolled during subsequent waves. There were 23 hospitals in the first wave, 23 more joined in the second wave, and an additional 25 joined in the final wave. The study excluded some hospitals (low volume and pediatric hospitals as well as hospitals that had participated in a pilot program to reduce ED LOS) and ended up analyzing data for 60 hospitals overall. They used frequency matched hospitals in Ontario as controls, some of which ended up joining the program in waves 2 or 3 and were thus lost as controls; there were 20 controls for the first wave, 27 for the second and 10 for the third. They looked at 90th percentile ED LOS as their primary outcome and secondary outcomes were median ED LOS, 90th percentile time to initial physician assessment and median time to initial physician assessment. The quality of care measures assessed were chosen using a Delphi analysis and included the number of patients who left without being seen, overall and short-term hospital admission rates, 7 and 30 day mortality, 30 day re-admission among admitted patients, and 72 hour unscheduled ED revisits.

### Results:

A total of 8,010,957 ED visits were recorded in test and control hospitals over the three waves. Only 1.4% of these unique encounters were missing ED LOS data. The researchers performed a difference-in-differences analysis comparing the program hospitals to the controls, which provided them with some small, statistically significant

positive effects of the program which were not consistent between the 3 waves. For example, study hospitals had a significantly smaller increase in the 90th percentile ED LOS in wave 1 (versus the controls) but not in waves 2 or 3. They did observe significant reductions in waves 1 and 3 for ED LOS in nonadmitted patients, however these reductions were small (-6 minutes in wave 1 and -15 minutes in wave 3). There are detailed results given in the paper for every outcome they looked at in each wave; however, none of them are very large, few are significant and almost all of them are only positive effects when presented as a difference-in-differences analysis. There were few significant differences in quality-of-care measures. Waves 1 and 3 had decreased rates of patients who left without being seen, which was a true decrease, and wave 3 had an apparently increased rate of 7 day mortality, again only due to the difference-in-differences analysis; the actual rate was the same pre and post program.

### **Validity of Results:**

There were several issues identified with this study in terms of validity. The study was funded by the MOHLTC, which had a vested interest in publishing positive results from a program they initiated. The study design itself is questionable, particularly regarding control selection and the number of hospitals it included. Using matched controls from another province that was not currently undergoing a pay for performance program would likely have been better. Many of the controls that they used ended up joining the program and were part of the study population in subsequent waves. Controls that did not join the program – why were they not selected? They likely had characteristics that were fundamentally different from study hospitals. This study compiled data from 60 hospitals, effectively eliminating any chance of seeing impressive results; great successes may have occurred and failures as well but in the large pool of data it settles around the mean and we lose the truly interesting results. The use of a difference-in-differences analysis could be described as confusing at best. For example, when they state in the paper that there were “significantly greater reductions in overall 90th percentile ED LOS” there was in reality no reduction in 90th percentile LOS and there was in fact an increase.

### **Generalizability of Results:**

Although this study includes a large number of institutions, they are all localized to a single province under one health authority, limiting the national and international generalizability. Another troubling limitation is the lack of information on how each P4P institution used the financial incentives. If we are to be told there is perhaps a slight benefit to a P4P model, it would be helpful to know what strategies were employed by the program sites to achieve these improved outcomes. Without clear information regarding who received the money, how much they received, and how it was spent, it is difficult to say whether the money itself had an effect at all. Even if we were to be excited about the effects described in the results section, we have no idea how they were achieved or how other institutions might benefit from this data. Finally, the 1-year outcome measures also limit generalizability. Nothing is known of how incentive programs would affect outcomes beyond the short-term. Therefore, when policy makers consider P4P incentives to improve wait times going forward, the important question of sustainability remains unanswered.

### **The Bottom Line:**

Journal Club attendees felt this study was weak in many areas of its study design, limited in generalizability, and although it did address its primary question showing statistically engineered, short-term, modest improvement in some LOS parameters among P4P sites, it does little to provide insight into the ultimate question of ‘how’ monetary incentive programs may relate to reduced ED LOS. As emergency care providers, there is nothing from this study that we can glean to influence patient care or help direct what changes could be made to improve quality and efficiency with which that care is delivered. The authors of the paper are correct in stating further research is warranted, particularly around the effects of P4P on specific ED processes and contextual factors.