

Article details: 2016-0151	
Title	Impact of restricting diagnostic imaging reimbursement for uncomplicated back pain in Ontario: a population-based interrupted time-series analysis
Authors	Benjamin Fine MD, Susan E. Schultz MA MSc, Lawrence White MD, David Henry MB ChB
Reviewer 1	Dr. Christopher Naugler
Institution	Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alta.
General comments (author response in bold)	This population-level observational study examines changes in diagnostic imaging requests following an administrative change in Ontario. The topic is important in the context of an increasing number of similar interventions across Canada. The paper is well-written and the analyses are appropriate. Thank you we appreciate these comments
Reviewer 2	Dr. Steven G. Morgan
Institution	University of British Columbia, School of Population and Public Health, Vancouver, BC
General comments (author response in bold)	This is a relatively simple but useful analysis of the impact of what might be considered "evidence-based delisting" of a payment for a specific diagnostic test. There is some ambiguity about what specific measures are being reported at various points in the paper. 1. Authors argue for the use of tests per physician, not tests per patient, as the measure of testing volume. Though the argument for this seems reasonable, it would be worth assuaging concerns of readers by noting whether (and how) tests per patient might differ as a measure over the period. The intervention was aimed at the test-ordering physician, so the physician is the natural unit of analysis. The issue of multiple imaging tests per patient is a separate consideration, but was not the specific focus of the policy change, which was aimed at physicians. In addition, we looked at only 3 imaging modalities and the two scanning techniques are often preceded by plain X-rays, which would constrain our ability to observe any intervention effect. In response to other comments we have added values for the variation in test ordering per physician to Table 1. 2. Related to the measure of use, it was sometimes unclear whether specific measures reported were tests per physician or per physician ordering tests. The table appears to report per physician ordering and yet the narrative appears to be per physician overall. We apologise for the rather loose language and have tidied this up in the re-draft of the manuscript. The denominator used in the analysis of ordering rates is the number of physicians ordering at least one test during the calendar year. We have added footnotes to the Tables and to the relevant subheading in the Results. 3. Figures depicting the time series data would benefit from inclusion of predicted post-intervention trend and/or the confidence interval around such a prediction. We understand the reasoning here but unfortunately several of the figures are quite crowded and extrapolated regression lines will make them very messy. It would be relatively easy with Figure 2 but in this case we believe the eyeball is as accurate as a regression slope. Rather than interfere with the figures we have added to 95% confidence intervals to the parameters in Table 2. In the marked-up version of the revised manuscript these are in red text. 4. Did the authors model other specifications of the intervention? A priori, one might argue that an intervention of this source might have a level effect that that it would not in and of itself change the trend in test use. Thus, one wonders if the intervention is perhaps over-specified. We agree that an intervention of this type can have a 'one time' effect and might not change the underlying trends in usage, which may be influenced by other interventions (eg Choosing Wisely campaigns). We have made this point more clearly in the Discussion (see below). We don't see an alternative to segmented regression analysis, but the results need to be interpreted carefully. We waited for 3 years after the intervention to have some confidence in our ability to discern the trends and we have addressed the limitations in the study. 5. Table 1 requires a bit of tidying up so that it is clearer what the final change from baseline refers to. Thank you; It Table 1 we have clarified that the final change refers to the difference between the final study year and the baseline annual average. 6. The authors make the argument that the reduction in x-rays by FPs is "not trivial" in financial terms. It would appear that they have sufficient data to quantify the billable value of the reduced number of tests, possibly even with suitable confidence intervals around such a "savings" estimate (in quotes because there will no-doubt be some substitution effects that have not been modelled with precision in this study, which one would only need to mention in a paper of this sort). We have added a new section covering the costs to the Ontario Health Insurance in page 8. These don't include the 'technical fees' paid to institutions to cover the infrastructure costs for MRI and CT. As Prof Morgan indicates it is hard to do a proper economic analysis as we have not estimated substitution effects and clinical outcomes. 7. I think the assessment of limitations is fair and that the authors are in a position to argue that the time-series analysis method is not fatally flawed even in the presence of some misclassification of the outcome measure (i.e., the roughly 20% of scans that were for patients with red flags unidentifiable with the administrative data). The rationale being that bias would only occur if the prevalence of red flags were somehow causally related to the intervention. We agree with this and have added a comment to the Discussion in page 12.
Reviewer 3	Dr. Emily McDonald
Institution	Division of General Internal Medicine, McGill University, Montréal, Que.
General comments (author response in bold)	This is a well written manuscript that examines an intervention to reduce the overuse of medical imaging for uncomplicated low back pain. In the province of Ontario, in 2012 there was a provincial policy change and government reimbursement for such imaging was removed. Imaging was reimbursed only for "suspected pathologies" with examples given such as infection, tumour, osteoporosis, fracture, radicular syndrome, cauda equina and inflammatory processes. Studies for LBP were looked at per practitioner per month before and after the intervention and non-LBP studies were used as a control as they were not subject to the same policy change. A few questions/comments:

1. What was the range for number of annual tests/physicians ordering these imaging tests? Is it reasonable to include those who order only one test per year? Should orderers be divided into "high orderers" ex. 10 more studies per year vs. "low orderers" (ex. <10 studies/year). Maybe there would be a greater impact on those who were outliers and ordered more studies than their peers. I would imagine that if you are a physician who only orders one MRI or one lumbar xray per year this intervention would not likely have affected your practice. However, if you order 10 or 20 of these a year then the impact might be much greater. Might be interesting to look at.
This is a good point. Dr McDonald's argument has potentially actionable policy changes – eg to 'target' only high ordering physicians. And, of course, there are several potential stratifications or sub-group analyses that could be carried out, including examination of intervention effects in older v younger physicians, male v female, older v younger patients, those with and without co-morbidity, or physicians that have high general test-ordering practices, or not, and so on. Our sole objective in this report was to measure the overall impact of this policy and we did not have a set of hypotheses regarding potential effect modifiers. This would require quite extensive new analyses that we are unable to undertake at present.

2. Similarly, I think you should be able to demonstrate, in the figures, when looking at number of imaging studies, the variability around the mean?
It is important to state that we are capturing all relevant imaging tests covered by OHIP, and all test ordering physicians over the study periods, so these are true mean values, not estimates derived from a small sample. Estimated standard errors will be very small and will not be a useful addition to the graphs. However, we do agree however that the ranges of test ordering practices are not fully captured in the current manuscript and we have added maximum and minimum values to key parameters in Table 1 (red text). As can be seen ordering patterns are very variable.

3. Which specialties (in the SP category) order this type of imaging? Could you list the top 5? Would be interesting to know (is it oncology? is it neurosurgery? is it internal medicine? potentially very different practice populations...but grouped together in this study).
Thank you for this comment. We have added text to the revised manuscript listing the highest ordering specialties on Page 8.

4. Differences in patient populations between family practice and other specialists may also explain the differences observed between the groups in reduction in certain types of imaging.
We agree and have added a comment to the Discussion (Page 10) to make this clearer.

5. Might want to mention when the choosing wisely recommendations came into being addressing low back pain as this potentially had an impact on imaging orders unrelated to the government policy change. Might want to soften the last sentence related to the impact of the intervention in the strengths of the study section. A lot of changes were concurrently taking place related to overuse in diagnostic imaging that may have had an impact. Would have been great to compare to Quebec where this policy was not in place. But RAMQ data can be difficult to access!
All good points. The intervention has to be seen against a background of several initiatives that may have been modifying practice and influencing the pre- and post-intervention trends. We are confident that the acute inflections that correspond in time with the intervention represent a true effect. However, the underlying trends may have been influenced by other factors (including Choosing Wisely) and we have made this point more clearly in the Discussion (Page 13).

6. It's true a major limitation of the study is that we don't know if appropriateness improved or not. You could for instance remove all patients who had spine surgery in the 18 months following their imaging. That would be elegant but may be beyond the scope of such an article.
Dr Gibson's suggestion makes sense. Excluding surgical cases will eliminate a proportion of the appropriately imaged patients. To do this in a comprehensive way would require exclusion of a broader range of patients – eg those who bone metastases who subsequently received chemotherapy or radiotherapy or those who went on to receive treatment for osteoporotic fracture. This is a big undertaking and as Dr Gibson notes is beyond the scope of this work.

7. Out of curiosity, how is this intervention practically applied? There is a sentence that slips in at the end in the conclusion saying it was not enforced. Maybe could add a bit more detail on this in the methods or introduction.
This is essentially an 'honour system'. There is no direct policing of the policy. The restriction provides a reminder to physicians and reinforces educational initiatives. We believe that many physicians are aware of the low value of imaging in uncomplicated low back pain and the policy provides them with an additional argument when discussing it with patients. We have added this point to the Discussion (page 11).

8. I wonder if physicians got wise to the intervention and so they knew what to write on the requisition in order for it to be compensated. Might be why the change isn't necessarily as sustained as you'd like it to be. Could add this into the discussion. Sometimes it's nice to discuss why or why not an intervention like this works. I imagine the point is that other provinces should have a policy like this in place and so you want to discuss why this intervention was a success and how it could be improved upon for others to benefit from.
It is likely that many tests continue to be ordered for patients who do not have 'red flags' in their histories or findings. As noted above there is no policing of the policy, and physicians who are convinced of the value in particular patients will find a way around the restriction. As noted earlier the policy makes a broad statement about the lack of utility of imaging in uncomplicated LBP and reinforces educational messages from other sources. The impact on lumbar spine X-rays appears sustained over the 3 years after the intervention (in contrast to the effects on ordering of single segment MRIs), which may indicate that this modality was already viewed as being of marginal value and more easily 'sacrificed' than MRI, which is being reserved for more serious cases. We think these points are covered adequately in the Discussion.

9. Small point but I think the grammar in the last sentence is off a bit. ...may help physicians practice "choosing wisely" or else ...may help physicians "choose wisely" sounds a bit better.
Thank you – we now refer to this in a new (penultimate) paragraph (Page 13).

Reviewer 4	Dr. Mamdouh M. Shubair
Institution	University of Northern British Columbia, School of Health Sciences, Prince George, BC

<p>General comments (author response in bold)</p>	<p>The topic of the paper is a specialized imaging topic; while seemingly important within the context of clinical practice related to Family Physicians (FPs) and Specialists's imaging practice, the body of the paper lacks clarification in several aspects. Most notably:</p> <p>1. Only one broad objective is stated. No clear rationale and background sufficient to warrant this "investigation" (study). What's only mentioned is how imaging tests may have been affected by the 2012 Ontario government announcement of withdrawal of insurance coverage for uncomplicated low back pain. Is this a sufficient reason to warrant this study? If so, this needs further elaboration. We disagree that this is a 'specialized imaging topic'. This is a study of imaging tests ordered in the hundreds of thousands by family practitioners and specialists. Restrictive coverage decisions are often used alongside educational initiatives such as 'Choosing Wisely', and it is important to determine the independent effects of the different components of such programs. We believe that the introduction section makes these points clearly.</p> <p>2. The Methods and Results sections are written in more technical terms and most notably the results are presented only in terms of "Annual numbers" of imaging tests but no rationale as to when/how/why these tests are being done by each of "Specialists" and "Family Physicians"? In other words, what warrants these imaging tests for "uncomplicated" low back pain (given the no-severity/seriousness of such pain)? If it is a serious clinical problem, then it needs to be specified as such and what, if left untreated, it may lead to -- as far as complications are concerned? We have tried to make it clear in the Introduction that uncomplicated low back pain is common, self-limiting and that unnecessary imaging is costly and exposes patients to unnecessary radiation (for some tests). We have also pointed to the fact that unnecessary investigation of uncomplicated LBP is the subject of international initiatives, such as Choosing Wisely. There is not much more we can add to this point. We agree that there may be differences in the types of patients seen by family practitioners and specialists and we added this point in the Discussion (see above). However, this is speculative. We have included all specialists in our analyses, in order to measure the overall impact of the restriction. Had we restricted the analyses to orthopaedic surgeons we would have been justified in saying that there were more severe and resistant cases that might be suitable for surgery. However, we don't feel confident in that statement for the analyses we conducted.</p> <p>3. While the study denotes some strengths and limitations, the Discussion section needs further elaboration WHY this is an important study. For example, L-spine imaging is done for "uncomplicated low back pain" -- what is the reason this is done? In other words, is it common clinical practice to do so? Further, the authors note that "some imaging studies were done for reasons other than low back pain". If this is the case, what are these underlying reasons? We have addressed the 'importance' of the study earlier in our responses. We have also noted earlier, and in the manuscript, that the imaging tests have been shown to be of no value in uncomplicated back pain. They probably continue to be ordered in the mistaken belief that they will point to a remediable cause or provide reassurance regarding the absence of serious pathology. In the case of lumbar spine X-rays they do neither, and in the case of MRI they may identify non-significant 'abnormalities' and lead to further unnecessary procedures. We believe these points are covered in the manuscript.</p> <p>4. The Conclusions the authors note that the 2012 Ontario Intervention policy resulted in a significant decrease in the numbers of imaging tests of the spine performed over three years in Ontario. What about more recent data from the last couple of years (for example, 2015-2016)? What is the trend there in terms of imaging tests, and is it really due to the 2012 Ontario policy or could it be referral bias by family doctors and/or specialists to do these imaging tests? We disagree. The visible inflections in the curves, their timing, and the parameters obtained by segmented regression analysis all point strongly to an impact of the 2012 policy change. There is certainly a fairly sudden change in the referrals for imaging, that corresponds to the timing of the policy. We don't understand how a sudden 'referral bias' could account for this.</p> <p>Therefore, the paper needs major revision, including in the Discussions citing what other previous studies have done comparably on such topic (Canadian studies or other Jurisdictions). To our knowledge there has not been a similar evaluation of such a policy in Canada, or elsewhere and we have included the relevant Canadian references.</p>
<p>Reviewer 5</p>	<p>Dr. Lee Green</p>
<p>Institution</p>	<p>Department of Family Medicine, University of Alberta, Edmonton, Alta.</p>
<p>General comments (author response in bold)</p>	<p>The abstract accurately describes the content of the manuscript.</p> <p>1. The introduction is short and to the point. It establishes the importance of the question and its relevance to current research sufficiently well. An estimate of the cost of unnecessary imaging in Canada would be a good addition; a nice to have not a need to have. We agree on the value of estimated cost savings (see above in response to Professor Morgan) and have added this to the revised manuscript (Page 8).</p> <p>2. In lines 34-39, the authors might consider referencing some of the literature on withdrawal of payment for other unnecessary interventions, e.g., the so-called "DESI" drugs, post-operative intermittent positive-pressure breathing treatments, etc. in the US. Again a nice to have not a need, if page space allows. This is a good point. As Dr Green notes there is a large literature on payment withdrawals going back to the 1980s involving a range of interventions, including drugs and surgery. We felt that we should limit our scope to interventions designed to reduce use of diagnostic tests. Imaging for low back pain is a substantial topic in its own right and, as noted above, is the subject of educational interventions. We feel that references 15 to 18 cover the topic sufficiently.</p> <p>3. The methods are described in sufficient detail. The choice of physician rather than population as the denominator is appropriate. Specifying the actual billing codes would be another nice to have. Agreed - we have added the Ontario billing codes to the text (Page 5).</p> <p>4. The modeling appears appropriate though I might like to see a bit more detail. We have added more detail regarding the technique and the rationale for using segmented regression analysis (Page 6).</p>

5. Including all physicians who ordered the test at least once in that calendar year is a reasonable approximation to exclude those who never order (e.g., FPs whose practice is largely OB/peds) while not excluding the likely substantial number of months with zero ordering by chance (as ordering is a low-frequency event for many FPs in particular). Reporting the percent of zero-event months might be helpful.

We discussed this at some length. To include all physicians with any OHIP billings during a year would have led to a large number of zero cells. As Dr Green notes we elected to study physicians who had ordered one or more tests in a calendar year. This was a pragmatic decision that linked the policy to those most likely to be affected. Of course, among those groups there is very variable ordering practice (see above and additions to Table 1) and there were many zero-event months' over the course of the study and these are included in the calculation of the event rates. We can add the proportions of zero-event months to the results, but we are unsure of the significance of the statistic or how it might be expected to change as a result of the coverage restriction.

6. The results presented are clear, but there is considerable redundancy between the text and Tables 1 and 2. The results section could be shortened, as the Tables present the information well.

This is a significant point. We struggled with the balance between interpretation and repetition. Tables 1 and 2 are now quite dense with the additions of ranges and confidence intervals (see above). To reiterate, these tables (and Figure 1) are concerned with different aspects of annual test ordering. Figures 2-4 deal with segmented regression analyses of monthly test-ordering, so there is little true redundancy. We have re-read the text in the light of Dr Green's comments and feel that a significant proportion provides necessary interpretation of some fairly complex data rather than regurgitation of data.

7. The discussion addresses the important points raised by the study, and handles both strengths and limitations clearly and sufficiently.

The tables are well done and the figures useful.

We appreciate these additional comments from Dr Green.