Article details: 2016-0095	
Title	Frequency and variation of Choosing Wisely recommendations in primary care: a retrospective, population-based cohort study
Authors	Ciara Pendrith MSc, Meghan Bhatia BSc, Noah M. Ivers MD PhD, Graham Mecredy MSc, Karen Tu MD MSc, Gillian A. Hawker MD MSc, Susan B. Jaglal PhD, Lynn Wilson MD, Kimberly Wintemute MD, Richard H. Glazier MD MPH, Wendy Levinson MD, R. Sacha Bhatia MD MBA
Reviewer 1	Patricia Li MD MSc, Clinician scientist, Assistant professor
Institution	Department of Pediatrics, Division of General Pediatrics, Faculty of Medicine, McGill University, Montréal, Que.
General comments (author response in bold)	Overall comments The authors have produced a study demonstrating the utilization rates and variability of 3 Choosing Wisely recommendations in primary care. Although similar work has previously been performed in the US, the authors justify the novelty of the current study, in that the setting differs (the Canadian system being publicly funded with no patient co-payments) and the data involves an entire population. They have also attempted to look at variation across practices, which few studies have done. They have constructed a large and unique study cohort, pre-implementation of the Choosing Wisely campaign in Canada. Given the international adoption/implementation of the CW campaign, the authors have provided a good rationale for the current study. The authors have appropriately listed some important limitations in their research – however, there are additional limitations/assumptions (listed in the major comments) that should be addressed in order to clarify the validity of the study findings. Major comments 1) Definition of low value repeat DEXA scans: Are there other reasons/specific conditions where a repeat DEXA scan may have been indicated/warranted – for example, in cancer protocols or medical conditions? If the latter is possible, how can it be assumed or shown that the majority of repeat DEXA scans in the cohort were not done for "appropriate" reasons? **Due to imprecision in DEXA testing, research suggests that a minimum of two years may be required for a reliably change in bone mineral density. While it is possible that some patients with an accelerated expected rate of bone loss, such as cancer or recent fragility fracture patients, may be receiving repeat scans for appropriate reasons, we suspect that the proportion of such patients in our cohorts affected would be low. We have revised our limitations section to discuss the possibility of capturing appropriate repeated scans.** 2) Definition of back pain with no red flag history: the authors have stated one of the important limitations in that health administrative data does not provide clinical data. This is especially important in considering the validity of the back pain outcome, since it is unclear how many patients actually had constitutional symptoms or infectious symptoms that may have warranted MRI/CT imaging. Can this be addressed with the data available? For example, reference 17 (Schwartz) used ICD-9 codes to identify some of these symptoms/diagnoses [although I do not know if these were correctly used and/or whether they could be used for the current study – 730xx (osteomyelitis), 7806x 7830x 7832x 78079 7808x 2859x (fever, weight loss, malaise, night sweats, anemia not due to blood loss)] **The methods developed for the ICES low back pain report that we used for this study included an extensive list of exclusions to select a conservative population for study. While it is possible that some high risk patients were not excluded, it seems unlikely that a significant proportion of our cohort would have a "red flag" given the long list of exclusions applied. Prior literature has demonstrated that only 4% of low back pain scans are done for red flag conditions across all specialties, and not just primary care (Gidwani et. al, 2016). The diagnostic coding system used by OHIP does not include all ICD-9 codes (as OHIP is a mix of OHIP 8 and 9 codes) and several of the codes used by Schwartz et al. 2014 either are not used by OHIP or have a different definition. Another limitation of the OHIP diagnostic coding system is that it only uses 3 digits, whereas the databases used in the study by Schwartz et al. 2014 use more digits, allowing for greater specificity (e.g. 285 in OHIP vs. 2859x). Please see below:** • **285 = other anaemias** • **730 = osteomyelitis** • **780 = convulsions, ataxia, vertigo, headache, except tension headache and migraine**

	• **783 is not used in OHIP** 3) Comparison of primary care practices: is it safe to assume that the majority of DEXA scans and back imaging were ordered by the family physician and therefore attributable to their practice? In particular for DEXA scans, this goes back to the question regarding the indication for the scan in the first place (e.g. reasons other than routine primary care screening – for example, in the specialty setting for specific conditions). **While we cannot state with absolute certainty that all DEXA scans or low back CT/MRI were ordered by family physicians, it is reasonable to assume that the vast majority were ordered by a family physician. In paper by Jaglal and colleagues published in the CMAJ in 2000, 80.2% of DEXA scans were ordered by family physicians. We do recognize that the limitation of this analysis is that some tests will be ordered or recommended by specialists, and have added to the limitation section of the manuscript that we cannot be certain that all of the tests were ordered only by family physicians.** Minor comments 1) Abstract line 28: the DEXA scan ranges from "4.0% to 54.9%" which is slightly different than stated on page 9 line 8 where it is "55.0%" **Thank you for identifying this mistake. In an earlier draft of this paper, this number had been rounded to 55%. We have since corrected this mistake.** 2) Page 6: Line 48/49 "validated billing algorithm" – please reference **We have added the appropriate citation here.** 3) In the results section: What are results of Cochrane-Armitage test for temporal trends? **During the course of these revisions, we decided against using the Cochrane-Armitage test for trend in favour of a Poisson regression to investigate trends. Where overdispersion was detected, the regression distribution was changed to a negative binomial distribution to produce more precise standard errors. We have revised the methods and results to detail these changes.** 4) Page 10 discussion lines 22-23: the authors state that the rates of DEXA scan were lower in the US study on Medicare patients – this is likely because the US study had a different denominator, i.e. included only those patients with osteoporosis? **The difference in populations studied may account for the difference in rates of repeat DEXA scans. However, since our study included patients receiving their first ever DEXA scan as screening for osteoporosis in addition to patients with osteoporosis, it seems likely that a study in only patients with diagnosed osteoporosis would have a higher rate of screening. We have clarified the discussion to note the difference in populations studied.** 5) Page 15: Table 2 what are the N's? (for example, 394,314 divided by 2,229,113 does not equal 21% for repeat DEXA scans) **The Ns in Table 2 are the cohort denominators for the 2012/13 fiscal year. We have added a footnote to Table 2 to clarify this.** 6) Figure 1 and 2 missing titles on draft provided to reviewer; figure 2 missing x-axis- and y-axis labels and legend **Thank you for your comment. We apologize for these mistakes and have added a page to the manuscript including figure legends, as well as edited the figures.** 7) Appendix 1, line 37-38: hospital admission for dx codes – are they for spine related conditions? **Yes, these diagnostic codes include those for spine related-conditions, conditions of the nervous system and conditions of the musculoskeletal system and connective tissue. Please see below:** • **324.1 = intraspinal abscess** • **334.8 = other spinocerebellar disease** • **334.9 = spinocerebellar disease unspecified** • **335 = anterior horn disease** • **336 = other diseases of spinal cord** • **340 = multiple sclerosis** • **342 = hemiplegia** • **344 = other paralytic syndromes** • **349 = other and unspecified disorders of the nervous system** • **349.81 = cerebrospinal fluid rhinorrhea** • **350-359 = disorders of the peripheral nervous system** • **710-739 = diseases of the musculoskeletal system and connective tissue** • **740-742 = congenital anomalies, nervous system** • **805 = fracture of vertebral column without mention of spinal cord injury**

	• **806 = fracture of vertebral column with spinal cord injury** • **839 = other, multiple and ill-defined dislocations** • **847 = sprains and strains of other and unspecified parts of back** • **950-957 = injury to nerves and spinal cord** I have no competing interests in completing this review.
Reviewer 2	Darren Lau MD PhD
Institution	School of Public Health, University of Alberta, Edmonton, Alta.
General comments (author response in bold)	In the paper, Pendrith et al. describe the rates of 3 "low value" services in Ontario from 2008-2012. The prevalence of cervical cancer screening for women < 21 or > 69 years of age was 8%. 4.5% of patients with low back pain received a CT or MRI within 3 months of initial diagnosis. Of patients with index DEXA scans, 21% received a repeat scan within 2 years. Further, predictors of early repeat DEXA scanning included being female, living in a higher income area, and having a higher risk index scan. The authors found significant regional practice variation (across LHNs, 2.5-8.3% for LBP imaging, 13.8-28.2% for early repeat DEXA scans, and 6.6-10.4% for inappropriate cervical cancer screening), with even higher practice level variation (LBP 0.8-32.6%, cervical cancer screening 0.9-35.2%, repeat DEXA scans 4.0-55.0%). These results suggest opportunities for reducing low value care. Comments - The methods appear to be state of the art. I appreciate the attention to cohort selection, particularly the measures taken to exclude patients with complicated low back pain from the cohort. **Thank you very much.** - Means of identifying patients with a regular family physician were mentioned in the methods. Does having a regular family physician result in reduced use of low value care? **Our data suggests that patients with no family physician had fewer low value Pap smears, but similar numbers of imaging for low back pain imaging and similar low value DEXA scans. While it is difficult to draw conclusions with the available data, it is certainly plausible that higher rates of walk in clinic usage may result in fewer screening Pap smears, but similar degrees of LBP imaging. On possible theory may be that screening, like Pap smears, is done as part of comprehensive primary care, whereas LBP imaging may be conducted on patients who present with the symptom of low back pain, and without comprehensive primary care follow-up, it may be most expedient to order an imaging test. Further research would be needed to better understand provider factors that lead to low value care ordering.** - For those patients with repeat DEXA < 1 year – are these being ordered by the index provider? By a different family physician? Or by a subspecialist? **We were unable to identify who the ordering physician was for DEXA scans. It is possible that some patients receiving a repeat DEXA scan had that scan ordered by a different provider; however, we suspect that the majority of DEXA scans were ordered by a patient's regular primary care physicians. This limitation has been added to the limitations section of the paper.** - Are DEXA scans within 2 years necessarily low value? The optimal duration between tests is unclear, and I have seen both endocrinologists as well as family physicians order repeat DEXA testing in 1 year's time for patients who are deemed at highest risk of fracture, to ensure stability on a bone protective regimen. The Canadian guidelines seem to allow for this, specifying 1-3 years for repeat testing, with the interval to be increased once therapy is shown to be effective. I understand that there are arguments against measuring less than 2 years – the imprecision of the scan may be greater than the radiographic change apparent within 1 year (is this actually well studied? Are there some patients for whom radiographic progression may be quicker?), and the DEXA scan does not have a well validated relation to fracture risk in patients already on treatment. It's complicated, but might the observation that higher risk index DEXA scans having a higher rate of early repeat scanning actually suggest appropriate, as opposed to inappropriate, allocation of DEXA scans? **The utility of repeating a DEXA scan more frequently than every two years is limited by both the imprecision of DEXA scanning and the expected rate of bone loss of a patient. For these reasons, the guidelines put forward by the Canadian Rheumatology Association to Choosing Wisely Canada recommend against repeat scans within two years as this interval may be needed to reliably measure change in bone mineral density. This recommendation is in line with several current osteoporosis guidelines. While some patients may have an accelerated expected rate of bone loss, it seems unlikely that the level of utilization observed would be solely explained by this. These patients likely have secondary causes of osteoporosis and are likely a small minority of patients seen and managed by subspecialists.** - The variability between practices is remarkable. Certainly, practices with DEXA scan rates of 55%, cervical cancer screening rates of 35%, and LBP imaging rates of 33% are marked departures for the norm that would be hard to explain on case mix or clinical benefit. The major value of this work would be to direct further study towards

	understanding why these variations exist, and to target interventions where warranted. **Thank you. We agree that the degree of variation of ordering between practices is quite remarkable and that it seems rather unlikely that a practice with a rate of repeat DEXA scans of 55% is explained by high risk patients. We believe the results of this work can have a significant impact on targeting future quality improvement work in primary care.** Overall thoughts This is a commendable study, and important work from a health reform perspective. **Thank you very much.**
Reviewer 3	Dr. Emily McDonald, Assistant Professor of Medicine
Institution	McGill University, Division of General Internal Medicine, Montréal, Que.
General comments (author response in bold)	Thank you for the opportunity to review your work. The major limitations are the methods section and the statistical analyses. Could a statistician be consulted or someone with a PhD in epidemiology? This could improve the scientific rigour of the study. Was someone at ICES involved in the statistical analyses? **Several clinician-scientists at ICES with many years of experience studying administrative data were involved in this study, as well as two of ICES' data analysts. While there are some limitations in the statistical analyses of this study described in our discussion (e.g. lack of adjustment for patient and provider factors), the purpose of this study was exploratory.** Some comments: how did you choose >/= 15% as a cut-off for "substantial room for improvement" and what is meant exactly by a "high degree of variation"? (this is in the methods section (page 7; lines 38-41). The definitions in the methods need to be tightened up a bit. **The 15% threshold was selected somewhat arbitrarily, but was identified as a utilization rate which would have budgetary impact and would be meaningful for policy interventions. We have since revised the methods to remove the phrase "high degree of variation" as this was previously undefined.** I understand that you looked at patient-level predictors. Did you consider looking at provider-level predictors? i.e. years since graduation or province or institution where provider trained? Or academic vs. community practice? If we identify provider level predictors for overuse this could help target the campaign. If you have OHIP data do you not have some provider information? **We collected data at a primary care practice level and were unable to collect individual provider-level factors. We only had access to information on provider-level predictors at an aggregated level to that of the entire primary care practice. We agree that provider-level data would be very useful for identifying who best to target with interventions, but were unfortunately unable to do so in the present study. This is certainly a limitation of the study and mentioned in our discussion.** In the methods you don't say how you determined that patients with low back pain did not have any red flags. I see that there is more information provided in table 1, but it's not made explicit until you get to the limitations that this information wasn't really available. Do you know the indication for the scan? The methods for the paper need to be a bit more rigorous for publication purposes. This extra information can be provided in an appendix but you ought to list exactly what the exclusion criteria were and not just state for example "certain neoplasms". **An overview of the important red flags was included in Table 1 and we included the full list of exclusions in an appendix for brevity. Unfortunately, indications are not captured by OHIP, so this information was not available. However, the list of exclusions for the low back pain cohort was fairly exhaustive, and given that over 90% of patients with a visit to a family physician for low back pain were excluded for having a red flag, we suspect that we captured a low risk cohort.** How and why did you decide to show patient demographics for the year 2012/2013? Any reason why this year in particular? Need to explain why you chose this year as opposed to any other year. **We selected this year as it was the most recent data available while conducting these analyses.** Was it not possible to obtain data from post CWC recommendations? All of this data predates the campaign and so may be a bit stale dated. Should be clear that we are looking at all data prior to the implementation of the Canadian campaign. I say this because it is possible (and one would hope) that the campaign has had an effect and that the recommendations you target for intervention may have changed in the 3 years

	since the campaign has been in effect. **Unfortunately, when we were conducting this study, we were unable to access data from after the launch of the campaign as not all databases necessary for cohort selection were updated. Part of the purpose of this study is to determine the baseline ordering patterns prior to the launch. We plan on conducting repeat analyses to determine the impact of the campaign in future studies.** In the results was the rate of repeat DEXA scans that increased over the years a statistically significant result? Should include a measure of significance if so. **The change in rate of repeat DEXA scans was not statistically significant. The methods and results have been revised to detail this.** I think you state that you considered interventions to target as being significant if they had a utilization rate over 15% but low back pain scans and PAP tests didn't meet this predefined cut-off. But their overuse on a large scale is likely still significant. On a population basis it is likely that even 5% overuse may be significant especially when we consider all of the downstream investigations that may occur as a result. So I would argue that repeat DEXA scans are not the only target to be improved upon. **We appreciate the reviewer's comment and agree that 5% overutilization at a population level may be significant. However, the 15% cut-off was selected as this was considered the threshold required for changes at a policy level to be cost-effective. Future research is needed to identify what the minimum utilization rate for wide-scale intervention is, but until that threshold is identified, it seems most useful to target low value services with higher rates of utilization.** To further pursue this point, you could use the variability across practices as a "target for best practice". If practice A only uses 2.5% scans and practice B uses 8.5% scans, but their patients are not more likely to die or be diagnosed with, for example, late stage cancer, than perhaps a target for low back pain scans in under 2.5%. **We agree that the variability across practices presents an interesting opportunity to identify the target utilization rate for best practice. We hope the results from this research can be used to study interventions designed to reduce low value care across different primary care practices.** I'm not sure you want to include this sentence: "Since initiatives to reduce low value care will be most effective when targeted towards practices that are both common overall and show great variation, it is essential to conduct preliminary population level analyses prior to launching interventions." The campaign and interventions have been in effect for awhile now without this data. Rather, you may want to say something like, "prior to funneling limited resources into futher campaigning interventions, targets for good practice need to be established; family practices that are high consumers of low value care can now be the object of the next wave of educational campaigns." Or something like that. **Thank you for your suggestion. We have revised this section of the discussion, by removing that sentence and adding the sentence, "Prior to committing limited resources into further campaigning interventions, targets for best practice need to be established; family practices that are consumers of low value care may now be the object of the next educational campaigns."** The great value of your data is that you are now able to establish benchmarks for family practices to target. They may want to audit their own practices or you may want to feed back to them their performance as compared to their peers. **We agree that establishing benchmarks for family practices is critical for influencing behavior change, though we recognize that numerous patient and provider factors may be make it difficult to create one benchmark for all practices. The Choosing Wisely Canada campaign has partnered with provincial and federal agents to disseminate audit and feedback reports to individual primary care physicians, and this data will aid in those efforts.**