Article details: 2020-0188	
Title	The real world cost-effectiveness of bariatric surgery for the treatment of severe obesity: a cost-utility analysis
Authors	Erica L.W. Lester MD MSc, Raj S. Padwal MD MSc, Daniel W. Birch MD MSc, Arya M. Sharma MD DSc, Helen So PhD, Feng Ye MSc, Scott W. Klarenbach MD MSc
Reviewer 1	Dr. Mon Tun
Institution	Department of Pediatrics, University of Alberta, Edmonton, Alta.
General comments (author response in bold)	This manuscript aims to assess the real world cost-utility of obesity therapy from two perspectives: publicly funded healthcare payer and societal perceptive. The societal perceptive helps detect cost shifting between sectors. It is justifiable in principle but difficult in practice. **We agree that the societal perspective is difficult on a practical level. However, we believe it is indelibly relevant in the context of obesity. Obesity and its sequalae become more prevalent in our society annually: as reported by Statistics Canada in 2018, 26.8% of Canadians were classified as obese, with another 36.3% classified as overweight. The impact of this disease process is seen at a societal level, with issues such as absenteeism and presenteeism being widespread. A perspective that incorporations societal cost is especially relevant in a country with that has both a publicly funded health system and a high prevalence of obesity.** (p. 3) The cost-utility analysis was conducted for 2 year, 10 year and lifetime. The model included 2 important obesity related health events, diabetes and hypertension but not the cardiovascular events such as angina, heart failure. The authors concluded surgery is cost-effective from healthcare payer perspective and cost saving from societal perspective and good value for money for treatment of obesity for lifetime. The authors did not mention type of bariatric surgery performed for the study population. Were all the patients offered the same type of bariatric surgery? If different types of surgery were performed, would it result a change in the findings? **This is a very valid point. Due to word limitations, we removed much of the details of the APPLES study, in favor of referencing the original protocol and results.** **The surgical patients had a combination of Roux-en-Y gastric bypass, gastric sleeve and adjustable gastric band placement. Which procedure was undertaken was decided via an informed consent conversation between the patient and the surgeon. This is relevant to our cost utility analysis, as it reflects the real-world nature of the study: the participants and surgeon decided which procedure to undertake just as they would have had they not been part of a study.** **We have added more information for the reader with regards to the surgical arm in the methods section:** **All patients were deemed surgical candidates, as per the NIH guidelines.13 Of the 150 of patients that had surgery, 48 had an adjustable gastric band, 51 sleeve gastrectomy and 51 Roux-en-Y gastric bypass. All surgeries were conducted laparoscopically.**[1] **With regards to how this distribution of surgical approaches impacts our findings, it is likely that the inclusion of the adjustable gastric band diminished the clinical and cost effectiveness of the surgical option. Recent follow-up studies have demonstrated that the gastric band is less effective**

	than the other procedures included here. Corollary to these studies is a reduction in the use of gastric bands in favor of sleeve gastrectomy and bypass. A higher percentage application of these later surgeries would likely result in more excess weight loss and reduction in comorbidities without a significant increase in complication rate. This would likely translate to a more attractive ICER for surgery when compared to the other treatment arms. **The authors did consider conducting multiple subgroup analysis, with one approaching the types of surgeries employed. However, we felt that this manuscript would be more powerful if it reported on the APPLES cohort as a whole. Furthermore, as we are sure the reviewer is aware, it can be challenging to describe a Markov model with multiple subgroup analysis within 2500 words.** (p. 4) Did the authors employ half-cycle correction for the Markov model? **Thank you for noticing this omission. We did employ half cycle correction, but did not mention it in the manuscript. The following has been added to the methods section:** **Markov models using 10-year and lifetime time horizons were created. Half cycle correction was applied.** (p. 5) The authors did not mention the WTP used for the cost-utility analysis. **We did not report a threshold in the methods; to our knowledge there is no explicitly stated WTP threshold for Canada. Rather, we opted to report at which WTP value each treatment alternative becomes the preferred option, and to provide the cost effectiveness acceptability curves for the reader. Given that there are three alternatives presented in this study, providing the results across a broad range of WTP thresholds allowed us to provide the reader with information regarding the appropriate alternative at any given WTP, rather than simply one that we chose. As is illustrated by the curves, the results are quite different at various thresholds.** **To clarify this, we have added the following statement to the methods: "The willingness-to-pay(WTP) value at which an intervention (medical or surgical) became cost effective was reported for each perspective and time horizon".** (p. 7) The authors explained in details about the transitional probabilities, costs, utility estimation and assumptions in Supplementary files 1 & 2. In Supplementary file 2 (page 31), the authors mentioned employing only one friction period if patients were unemployed for 2 separate periods in a year. Were the 2 unemployed periods maximal? How did the authors assign the friction period for a person with more than 2 unemployed periods? It would be beneficial to report the friction period and the unit cost used to value lost productivity. **A friction period of 20.1 weeks, corresponding to the Canadian average length of unemployment in 2016, was applied. If a patient met criteria to have a friction period applied once during the study, it was applied. If they later gained employment, but became unemployed again during the 2 year study, no additional friction period was applied i.e. there was a maximum of one friction period applied to each person during the 2 year study period. This was to maintain a conservative estimate, and to avoid double counting**

	secondary to purposely short term work. As nearly all patients listed as pregnant had EI and sick weeks documented, no additional time was friction time was assigned for maternity leave. **This likely underestimated the cost of the standard care and medical treatment groups from the societal perspective, as these patients were more likely to be unemployed (as is reported in the supplementary file, the percent of persons incurring a friction period was 40%, 35% and 17%, for the standard care, medical and surgical groups respectively).** **The friction period (20.1 weeks) was multiplied by the Canadian average number of hours worked per week in 2016 (30.5), and multiplied by the Canadian average hourly wage (as of April 2016) by gender.** **We have added the following explanatory statements to supplementary file 2 to clarify this issue:** **The friction period (20.1 weeks) was multiplied by the Canadian average number of hours worked per week in 2016 (30.5), and multiplied by the Canadian average hourly wage (as of April 2016) by gender.** (Supp file 2) PSA was conducted with Monte Carlo simulation with 1,000 iterations. The findings will be more robust with conducting 10,000 iterations and one-way, two-ways or three-ways probabilistic sensitivity analysis. One-way sensitivity analysis was conducted for different time horizon for both perspectives but it would be helpful to conduct scenario analysis. **We have reconducted the PSA using 10,000 iterations, and have updated the manuscript to reflect this. The results were unchanged.** **We are not aware of three-way probabilistic sensitivity analysis: we will assume you meant deterministic analysis.** **The fourth edition of the CADTH HTA guidelines do not recommend deterministic sensitivity analysis. For this study in particular, the new guidelines were released just as we had completed the work and were preparing the manuscript: in fact, the new guidelines were introduced at the same CADTH symposium that this study was presented at. Given this timeline, the group discussed the approach, and felt that the one-way deterministic work should be included, as it provides a decision maker with relevant and important information about application to their system. For example, if the population they are considering treatment for is almost entirely made up of patients with diabetes, they could see how the variables pertaining to diabetes impacted the model.** **Given the limitations of presenting models such as this one in manuscript format, only the most important and relevant sensitivity analysis can be included. As there were no obvious scenario analysis that would contribute significantly to the results of the study, we have not included any. If the reviewer feels that there is a specific scenario analysis that is more important for the reader than the other sensitivity analysis included, the authors can remove other aspects and include scenario analysis. Of note, face, internal and external validity were assessed for the model presented.** (p. 7) The discount rate for sensitivity analysis was mentioned (page 9). A sentence or two on the discount rate used and in the models would be beneficial. **We have added the following statements within the methods section for**

	clarity surrounding the choice of discount rate for the model (base case) and the sensitivity analysis. **Statistical Analysis** **Models** **Costs and effects were assessed at a discount rate of 5%, as per convention.** **Sensitivity Analysis** **Discount rates of 3% and 0% were applied in sensitivity analysis.** (pp. 5 and 7) In supplementary file 2 (Page 32), the authors mentioned that the cost associated with death in the medical therapy arm was applied beyond 2 years but in Table 3, the cost estimates for mortality associated with standard care and surgery were listed. **Our apologies if Table 3, or the supplementary files, are unclear or difficult to read. Table 3 reports the same cost of death for medical and surgical therapy. While the death state in all three arms were included in the table for completeness, the values are identical in the medical and surgical groups. As there were deaths that occurred in the standard care arm, that has a different cost value applied. We can remove the cost of death in the surgical group row from the table if the reviewers feel this would improve the manuscript.** (na)
Reviewer 2	Laurie Twells
Institution	School of Pharmacy and Faculty of Medicine, Memorial University, St. John's, Nfld.
General comments (author response in bold)	This is a very well conducted study that uses data from the prospective APPLES study along with administrative data from the Health Quality Council of AB and published cost data to conduct a modelling exercise. Very robust two year data that compares health and economic outcomes of 500 patients (bariatric surgery, medical management, standard care) are used to model out to 10 years and lifetime to determine the cost effectiveness of bariatric surgery from a public health care system perspective and a societal perspective. The study provides more evidence that bariatric surgery for the treatment of severe obesity is cost effective on both fronts due to its impact on improving comorbid conditions, improving quality of life and increasing life expectancy. In addition, increased work life productivity is taken into account and health care use when examining the impact from a societal perspective. The data is in line with other health economic evaluations published that primarily use American or European data. Thus it is good to see Canadian data being used to support our health policy decisions as these health care systems are not comparable. Some minor recommendations: be consistent in the use of severe or extreme obesity - choose one. Use first person language where appropriate.