Cervical priming is a frequently utilized health care service, and is often delivered as inpatient care with some unattractive features. There is limited empirical evidence which can guide how best to develop alternative models of priming care. The authors analyze data from a discrete choice experiment which elicits the patient’s preferences for specific aspects of priming. Each choice set consists of 3 alternatives: two alternatives describing new forms of priming care (outpatient priming and enhanced inpatient priming) and one describing the prevailing form of inpatient priming.

Major Compulsory Revisions

1. “A multinomial logit (MNL) model with a panel specification” (p.7): I think further clarification is needed here. While there is a variant of MNL (random parameter or “mixed” MNL) which can address the panel dimension of data, Table 4 and related discussion do not provide information that this variant is what the authors have estimated. One possibility is that the authors have estimated the usual cross sectional MNL model that ignores the panel dimension of their data, and computed sandwich standard errors clustered at the respondent level. If the authors have estimated mixed MNL, a more specific discussion of their model specification needs to be provided. If the authors have estimated the usual MNL model and reported clustered standard errors, their empirical strategy needs to be explicitly described as so with due acknowledgement of its limitations.

2. “To achieve the most parsimonious model possible, without compromising model fit, each variable that was non-significant was removed and the model re-estimated. Model fit parameters, and Log Likelihood, were assessed after each re-specification.” (pp.7-8) where the model fit parameters seem to be “the likelihood ratio test statistic for the global test of zero model coefficients, the McFadden’s pseudo R-squared and Akaike’s information criterion.” (p.7): This summary of specification search appears incomplete. On one hand, quite a few coefficients in Table 4 are statistically insignificant at any conventional level. On the other hand, it would not be possible to base model selection on all those fit measures simultaneously. For instance, when some variables are dropped, the pseudo R-squared always declines whereas the AIC may go up or go down, regardless of whether dropping them leads to a statistically significant decline in the Log Likelihood.

3. The authors state that “categorical variables were effects coded.” (p.8) They
report both the beta coefficients as well as the exponential functions of those coefficients. The latter transformed coefficients are interpreted as the effects on the odds ratios. The following comments I make may be wrong because I’m not familiar with effects coding. But I wish to point out some issues, and ask the authors to re-consider interpretation of the transformed coefficients if my comments are not wrong. With dummy coding, it is true that \( \exp(a \text{ coefficient on a particular dummy variable}) \) represents how the odds ratio changes in response to a change in that variable from its base level coded as 0 to another level of interest coded as 1. Since moving from dummy to effects coding entails re-normalization of the model coefficients, I find it unlikely that the effects on the odds ratio can be computed in the same way when effects coding is used. For example, suppose that a 2-level attribute is effects-coded. Changing the resulting variable from the base level to the new level involves a 2 unit increase (from -1 to 1) which means that the associated change in the odds ratio is \( \exp(2 \cdot \text{coef on this variable}) \) instead of \( \exp(\text{coef on this variable}) \). For a more complicated case of a \((K+1)\)-level categorical attribute, with the use of effects coding, the base level corresponds to the case when all \( K \) resulting variables take the value of \(-1\); to compute how the odds ratio changes when this attribute changes from the base level to a target level of interest, all \( K \) coefficients will need to be taken into consideration instead of the coefficient on the target level alone: the variable representing the target level increases by 2 units from -1 to 1, while the remaining \( K-1 \) variables representing other levels of the same attribute increase by 1 unit from -1 to 0.

4. The comment 3 also applies to Table 5. The footnoted formulas are exactly the same as what they would have been with the use of dummy coding. With the use of effects coding, the numerators may need to be suitably revised to capture the willingness to trade off the frequency and duration of travel for a change in a particular attribute from the base level to another target level.

5. The authors’ DCE is an unlabelled choice experiment wherein the only information that differentiates the 3 presented alternatives is their program characteristics. I have 3 related questions about the model specification. First, why have the alternative-specific constants (ASCs) been included? I’m thinking that except when it is assumed that the patient has realised option A always mimics outpatient priming and option B always mimics enhanced inpatient priming, it is difficult to associate the ASCs with average unexplained tendencies to choose those priming services over basic inpatient priming. Second, why have all coefficients been specified as alternative-specific coefficients? I find it difficult to see why the patient’s tastes for generic attributes like travel time and familiarity with the midwife would change depending on whether the option is labelled A or B. Third, why only the ASCs are assumed to vary with patient characteristics? Given the DCE design, I find it more natural to think that the patient’s choice behaviour is driven by her preferences for the underlying program characteristics, instead of the alphabetical labels. I’m thinking that a more natural way to enter the patient characteristics into the utility function is to let the coefficients, each of which represents the patient’s taste for variation in a particular program characteristic, change with the patient’s observed characteristics.
6. “Our results suggest that outpatient priming was preferred over either enhanced inpatient priming or basic care” seems unwarranted as an unqualified conclusion. Given the results in Table 4, whether outpatient priming is the most likely choice will depend on the characteristics of this option and enhanced priming, as well as patient characteristics.

Minor essential revisions
7. In Table 2, a circle needs to be added to Choose Option A
8. In Table 4, the lower and upper bounds of CI are always displayed as positive numbers even when in fact they are negative.

Discretionary Revisions
9. I’m thinking that “overall value of alternative priming options” can be easier to interpret if the index numbers are transformed into choice probabilities.
10. Throughout this paper, the results are often summarised as: (a) the patient is more likely to choose Option A (or B) over option C when a particular program characteristic of A (or B) is present, and (b) the patient with a specific personal characteristic is more likely to choose A (or B) over C. I’m thinking that the issue of “more likely” than what or whom can be better clarified. In the present context, “more likely” may mean two different things: (i) A (or B) is less likely to be chosen than C without the relevant program or patient characteristic which increases the odds in favour of A (or B) (ii) A (or B) is already more likely to be chosen than C and the program or patient characteristic in question makes it even more likely to be chosen than C.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests.