

## General

In this paper, the authors systematically review choledochal cysts during pregnancy. The authors do a great job at highlighting why the paper is important. First, there are limited studies looking at choledochal cysts in the perinatal population and second, understanding the behavior of choledochal cysts in the perinatal population can influence clinical management. The authors should be congratulated on the amount of data they reviewed. They found that while rare, choledochal cysts should be included in the differential of liver injury, jaundice and RUQ pain in pregnancy. They offer good perspective on management in these patients based on a the largest data set currently available. This paper would contribute new concepts to the available literature on this under-represented topic in the literature. In general – the following things should be addressed prior to publication to improve the quality of the article (please see attached document for more thorough line by line feedback:):

Dear reviewer, thank you for your time, positive and constructive comments, and for understanding the importance of the topic we studied.

We took into account each of your comments and revised the article accordingly. We answered your questions and gave comments for your line-by-line analysis.

There is inconsistent within the results regarding the number of patients in each model. This is concerning that maybe data analysis should be rechecked for accuracy.

Results, as well as statistical methods, are rechecked and revised.

The authors describe two models they created but present the results for 4. In the models, there is no explanation as to why certain variables were included and how the authors selected those models. There is not Discussion as to whether these models fit the data well (analysis of AUC or C statistic or goodness of fit).

In consultation with statisticians, we reevaluated the statistical methods used. As a result, logistic regression was abandoned for the following reasons related to the necessary assumptions for such statistical analysis. First, our data did not include ten events per variable, a prerequisite for adequate logistic regression modeling. Second, data are collected from case reports over a long period. Therefore, this is not a case-matched study which would be much more appropriate for logistic regression. Third, even in this cohort of a relatively large number of cases, due to non-standardized treatment protocol (treatment depended on a personal clinical decision) and low rate of main outcome variable (only 7 cases of mortality), multivariable statistics would not be useful in describing predictive factors for mortality. Therefore, in this type of research, logistic regression would definitely not imply any causation, and logistic regression models for predicting mortality would be useless. Consequently, our study was mainly based on descriptive statistics, and inferential analysis was done only for correlation and univariate analysis of the “likelihood to receive interventional treatment”, which gave us answers about factors that urged

surgeons/gynecologists to intervene (surgery, drainage or delivery induction). This was achieved by calculating the crude odds ratio with z statistics and p-value. The methods and discussion section shortly mentioned the rationale for all this.

Extraneous text should be removed from sections and authors should ensure the content of each section is correct. For example, results sections should include results from the study and not definitions.

It is corrected according to your suggestions.

The authors should comment on the number of non-English articles were used and how that data was translated or accessed correctly.

It is now mentioned in the methods section and Discussion.

Extraneous text should be removed from sections and authors should ensure the content of each section is correct. For example, results sections should include results from the study and not definitions.

It is corrected according to your suggestions.

The Discussion is far too long as it is currently written - 2843 words not including limitations and conclusions (3143 words total). Removing the first 15 lines may help reduce this and keep text relevant. Further, the Discussion is both somewhat repetitive and also there are instances where new data is presented in the results (e.g. indications for urgent cesarean section).

The discussion section is revised, structured, and shortened to 1968 words

**Note** - This reviewer was unable to review tables and figures as they were not uploaded in the text file or any of the other available files.

Tables and figures are incorporated into the manuscript file.

### **Abstract**

Overall the abstract is Concise and clear ; conclusion portion hit all key points.

Thank you.

Consider changing line 69 to “the aim of this analysis was to systematically review..’

Changed.

Consider clarifying line 70 regarding the meaning of “factors influencing the outcome”

This phrase is removed.

Line 82 – avoid starting a sentence with a number

Corrected.

Line 84 – consider ending sentence at 8.2% and starting a new sentence for fetal mortality.

Corrected.

## **Introduction**

Line 110- Todani’s classification is most widely used for what.

The sentence is rephrased.

Consider cutting down on the #'s in epidemiology – stating that the incidence varies, is higher in the east than west, and offering a general range would convey the same message with fewer words

It is corrected according to your suggestions.

Last sentence – clearly state the aim similar to what was said in the abstract.

It is corrected according to your suggestions.

Consider changing the organization of the introduction – the authors talk about general incidence, then pregnancy, and pathophysiology in pregnancy and then back out to pathophysiology in the general population and then back to pregnancy. Consider talking generally about CC's and then narrowing in on the population of interest.

It is corrected according to your suggestions.

## **Methods**

Authors clearly described methodology – I appreciate the inclusion of the search terms, and the data bases searched.

Thank you.

The authors make a point to say Non-English papers, thus should document how many papers were not in English? How were those papers translated or data accessed in a reliable fashion?

We added an explanation in the methods section and a discussion.

Consider moving this statement from data management, this is more of an analysis statement “The study investigated the influence of the variables mentioned above on maternal outcomes.”

Removed.

Consider clarifying this statement in lines 174-176. “Categorical variables were analyzed using the chi-square test, and correlation analysis or multiple logistic regression was used for an association between dependent and independent variables.” As written it seems to switch back and forth between categorical and continuous variables.

This section is corrected due to changes in statistical methods.

Lines 177-180: Is there a reason that more patients were used I the first logistic model compared to the second model? How did the authors select variables to go into each model (clinical

judgement? Based on chi squared?). Did the authors evaluate how well the models represented the data? **There is inconsistency in the methods and results regarding the number of patients in each model.**

Logistic regression was abandoned due to excellent suggestions from the reviewers.

Grammar

Line 151: postpartum".--> postpartum.”

Corrected.

Results

Line 190 – consider removing “as shown in table 1” and instead referencing table as done for other figures with parentheses.

Corrected.

Line 192 – 193 – “In 88 cases (90.7%), CC was diagnosed during pregnancy and in 9 (9.3%) in the postpartum period.” This seems more like a patient characteristic compared to other figures in the study characteristic section as the patient characteristics section discusses trimester of diagnosis.

It is moved to the appropriate section.

Line 199 “Fourteen had one, and 7 had more than one previous pregnancy.” Be consistent with referencing of numbers. If going to spell numbers, spell all those numbers less than 10.

Corrected.

Line 205 – 207 – The following statement is not a result. “CC triad includes abdominal pain, jaundice, and palpable right upper quadrant abdominal mass, while the Charcot triad includes jaundice, fever, usually with rigors, and right upper quadrant abdominal pain” Consider moving this definition into methods or if you reference it in the introduction that would also be sufficient.

It is moved to the Introduction section.

-

Lines 211-213. – The following statement is not a result: “Before transabdominal ultrasound (US), CC during pregnancy was diagnosed intraoperatively. Since 1976, radiologic methods have enabled preoperative CC diagnosis. In developed countries, MRI became common after 2000”. Consider moving.

It is moved to the Discussion section and rephrased.

Lines 219-222. Laboratory results – since this is only 1 section, consider combining with another section. Further, would update the methods to reflect that only bilirubin was able to be collected. Could consider combining with patient characteristics or including with correlation data.

Corrected.

Consider changing this sentence to be more clear: “Unlike maternal, the fetal outcome was not consistently reported, and a clear description of fetal mortality was noted in only 21 (21.6%) cases” → “Fetal outcomes were not consistently reported and only addressed in 21 cases where the outcome was fetal mortality.”

Corrected.

Lines 256-258 – The following sentence seems more like a discussion or conclusion point. “However, we presumed that the authors would mention fetal or newborn death, so fetal/newborn mortality may be reliable when calculated from the presented data.”

Corrected.

Lines 265-267 – The following sentence is not a result, “Pearson's correlation analyses of continuous parameters (age, cyst dimension, and bilirubin value) were executed on the different numbers of patients according to the presence of analyzed variables.”

It is moved to the Methods section.

Lines 272 – 274 – including why you used two models may be more appropriate in the results. Re-stating the model choice is not needed. **Further there is inconsistency with the number reported in the methods and the results.**

Statistical methods are changed, as mentioned, and we did not use 2 models anymore.

Lines 280 – 281 “In addition, we evaluated the likelihood of receiving surgery regarding cyst size, presence of Charcot triad, and jaundice (Table 8” – this is confusing because jaundice is a part of the Charcot triad. Please explain.

Yes, jaundice is part of the Charcot triad, but here we wanted to evaluate jaundice as a single symptom (without other components of the Charcot triad). However, the new analysis included bilirubin levels of >80 as a variable.

Lines 280-286 – these models are not discussed in the methods while the first two models were.

Statistical methods are changed, as mentioned.

## Conclusions

The first 15 (289 – 305) lines of the conclusion are not related to the authors data or what they found in their study. Consider removing, moving to introduction or discussing these points relative to data in the paper.

This part is removed, and part is added to the Introduction while some of the points are discussed later in the Discussion section.

Line 307 – “CCs are not even mentioned in the differential diagnosis of biliary tract diseases in pregnancy” this sentence seems out of place. Appreciate the referencing of the similar article and the authors explanation of attempts made to access this data.

Removed

Appreciate the referencing of the similar article and the authors explanation of attempts made to access this data.

Lines 314 – 135 “We did not get a response from one author who presented a case in a Slovak journal [106], and we could not translate a case from Aizadehf et al [107].” This comment highlights a previous review point that the authors should state how many articles required translation AND how the articles were translated.

We clarified this in the Methods section.

Lines 367 – 368 – this sentence seems contradictory, “The maternal outcome was as follows: 90 patients (90.7%) regained normal health after delivery and operation, while 7 (9.3%) died. This corresponds to a maternal death rate of 7.2 %.” Consider checking this math and just giving the maternal death rate as the first part of this sentence is discussed in the results.

Yes, it seems like a small error happened here. 7.2% is the correct death rate, and we removed the first sentence, as it is already mentioned in the Results section.

Lines 367-394 – These paragraphs seem very long and detract from the point of the paper. May consider taking this out or summarizing the causes of death. Alternately, presented the causes of death could be done in the results. The authors may then resume text at line 395 which includes good Discussion about how the maternal death rate relative to CC.

We agree with your suggestions. We shortened this section and removed it into the Results section, as we believe it may be important to state the causes of death more thoroughly.

Lines 413- 415 – seem out of place and re-present results as states in the results section.

Removed.

Lines 416-421 mirror earlier comment about Discussion of imaging modality. Please remove the excess text in the results and use this space to discuss differences in diagnosis.

The imaging discussion is revised and rearranged according to your suggestions.

Lines 423 – 425 – this sentence presents results that are not mentioned in the results section. “Urgent CS was indicated mainly due to CC complications (cholangitis, rupture, or peritonitis) or fetal compromise. Outside these indications, severe anemia (2%) and intense abdominal pain (3%) were reasons for CS. In 7 cases, the indication was not specified” Try to avoid presenting new data in the Discussion.

We agree. These results are moved to the Results section.

Lines 430-499 – are these statements a review of literature? Or conclusions the authors derived from the data. If they are review of current literature, consider summarizing and making relations

to your findings. If based on your data, consider referencing this more clearly; this was done well in lines 475-477.

This part of the Discussion is reformulated, shortened, and revised according to your suggestions. Most statements are based on our data, and the reference supports those that refer to literature. The authors of this article propose a treatment algorithm, and it was based on our view of the study results. This is the first such decision-making recommendation for this type of disorder in pregnancy.

## **Limitations**

Lines 508-509 – this sentence again draws concern to how the percentage of papers that were not in the authors primary language and how this data was accessed.

It is explained in the Methods section and the Discussion section.

May move the fetal outcome assumption from results here.

It was moved to the Discussion section and shortly mentioned in the Limitations.

We highlight that line numbers have been changed during the revision, but all changes can be tracked when the original and revised versions are compared.

Once again, thank you for your report and deep analysis of our paper. We look forward to your decision, and if any other changes or explanations are required, we will be glad to provide them.