Peer Review Overview

Manuscript Title: Phase II Trial of Atezolizumab in BCG-Unresponsive High-Risk Non-Muscle Invasive Bladder Cancer: SWOG S1605

1st Decision letter

Reference: EURUROL-D-23-00287
Title: Phase II Trial of Atezolizumab in BCG-Unresponsive High-Risk Non-Muscle Invasive Bladder Cancer: SWOG S1605
Journal: European Urology

Reviewer #1
In this study, Black et al., reported efficacy and safety results from results from a phase 2, open-labeled study using atexolizumab for the treatment of BCG unresponsive bladder cancer. The authors conclude that the efficacy of the treatment was modest and the use of atezolizumab should be balanced against its treatment-related toxicity. The study did not meet the prespecified threshold for interim analysis of >=7/25 CRs at 6mo via bladder mapping biopsy. In the 74 CIS containing patients, a CR of 27% was found. For the papillary only cohort, 18mo EFS was 49%. These results were compared to previously reported trial results using pembrolizumab, nadofaragene, N803, and vicinium. Although an important trial in the BCG unresponsive setting, this trial does not move the needle in further understanding how best to treat this disease. The results were very similar to the previous monotherapy trials with the toxicity profile very similar to that reported for pembrolizumab.
Comments:
1.The authors report 166 patients were treated, and it was only subsequently found that 37 patients did not meet inclusion criteria. Was there no steering committee oversight throughout the trial to review patient eligibility criteria for enrolment as the trial was accruing?
2.9 patients with persistent or recurrent CIS/Ta at 3mo were continued on treatment. Was this a planned treatment for this cohort? If not all patients with CIS/Ta HG at 3mo were treated, what were the determinants for patients to continue treatment?
3.The use of blue light cystoscopy was at the discretion of the treating physician. If not used consistently, this may skew the trial results. For instance, if a patient were to be diagnosed and undergo complete resection using blue light cystoscopy at baseline, and then assessed using white light cystoscopy, he may be erroneously diagnosed with no recurrence due to the lower sensitivity of white light. What proportion of the patients underwent blue light cystoscopy at baseline vs. at 6mo biopsy?
4.It would be useful for the urological community at large to understand the timeline of adverse events, particularly the onset of death related events, to better maintain vigilance on patients treated with immune checkpoint blockade.

Reviewer #2
This is a very well-written paper and presents results of SWOG 1605, which have been presented at international meetings. The NMIBC research community has been awaiting this paper and it will be well-received and I am sure highly referenced as there are only a handful of registration trials for BCG unresponsive bladder cancer published to date. Pembrolizumab and nadofaragene firadenovac have been published and this study will add to those high impact studies even though it was terminated early because the study target sample size was met.
Comments:
1.Please provide the initial statistical plan for the number of patients expected to be enrolled. What was the original criteria for statistical significance for the CR rate in CIS patients? Was there a pre-planned lower 95% CI bound that had to be excluded for efficacy?

2.On page 9, CR is defined as absence of high-grade tumor on bladder biopsy and absence of upper tract or urethral disease. Can the authors also present CR based on bladder pathology alone? The number of upper tract and prostatic urethral recurrences in the CIS cohort are provided, so this should be easy to add as a calculation.
3.Since only one patient with high-grade recurrence on mandatory biopsy had no reported abnormality on cystoscopy or cytology and since the detection rate of missed HG tumour in the nadofaragene trial was only 3.3% with mandatory biopsy, it is hard to know whether mandatory biopsy is truly needed for future CIS bias. Can the authors comment whether future studies should, in their opinion, mandate biopsy at a preset time point based on this date?

Reviewer #3
This is an important study. The team are to be congratulated for completion and documenting it so stringently. I am sure it will be of interest to the community.
Comments:
1.Please add line numbers to help with peer review.
2.Who administered the Atezolizumab? The author's list does not really reflect Medical Oncology engagement. This is important given your key concern about safety. Could it be that Urologists oversaw administration and did not monitor patents safely etc.,
3.Methods: How did you address blue light cystoscopy?
4.Methods: Path CR- did you use cytology/washings/barbotage as well as biopsy? Please Caplin- with respect to Supp Table 2.
5.Results: Do you have screening numbers? How many were screened, declined, judged ineligible?
6.Results: In figure 1, there is 1 death. But 3 are described in the Safety results text. Also, in Table 3- what are these Immune/drug mediated Grade 5's not Adverse events of Special Interest. Can you define Special Interest?
7.Table 2: I do not understand the footnote **One patient has Nx in the TanyN1-3 row. Please elaborate, esp. given Table S2 (4 N1+ cases).
8.Can you add the OS figure as a third panel for Figure 2? It is good to see it in the main paper.
9.Discussion: It is concerning that 4 patients have N+ disease at RC. Does this need highlighting in the Abstract?
10.The reasons for declining RC at entry are revealing. Most cases were Patient choice- who presumably would have had RC without this trial.
11.Please rename the Table S3 supplementary file. It is 'MIBC no numbers'

1st Author Response Letter
Response to comments from Editors and Reviewers:

Reviewer #1
In this study, Black et al., reported efficacy and safety results from results from a phase 2, open-labeled study using atexolizumab for the treatment of BCG unresponsive bladder cancer. The authors conclude that the efficacy of the treatment was modest and the use of atezolizumab should be balanced against its treatment-related toxicity. The study did not meet the prespecified threshold for interim analysis of >=7/25 CRs at 6mo via bladder mapping biopsy. In the 74 CIS containing patients, a CR of 27% was found. For the papillary only cohort, 18mo EFS was 49%. These results were compared to previously reported trial results using pembrolizumab, nadofaragene, N803, and vicinium. Although an important trial in the BCG unresponsive setting, this trial does not move the needle in further understanding how best to treat this disease. The results were very similar to the previous monotherapy trials with the toxicity profile very similar to that reported for pembrolizumab.
Reply: We believe that this trial provides important additional clinical trial evidence with respect to the use of single agent anti-PD(L)1 immunotherapy in patients with BCG-unresponsive NMIBC. After early enthusiasm for pembrolizumab in this disease state, our results underscore the modest benefit and the significant risk of toxicity.

Comments:
1.The authors report 166 patients were treated, and it was only subsequently found that 37 patients did not meet inclusion criteria. Was there no steering committee oversight throughout the trial to review patient eligibility criteria for enrolment as the trial was accruing?
Reply: The trial sites independently determined patient eligibility and registered patients on the trial without

requiring individual approval by one of the study co-chairs. The source documents were provided by the sites and reviewed by the study co-chairs to evaluate eligibility after the patients had been registered and had started treatment with atezolizumab. After recognizing the high rate of ineligibility, we instituted a mandatory screening of eligibility as described below:
"Initially a high number of ineligible patients were enrolled in the trial due primarily to misinterpretation of disease criteria. After introduction of a mandatory screening of eligibility by the study chair the subsequent ineligibility rate was reduced." (Line 158)

2.9 patients with persistent or recurrent CIS/Ta at 3mo were continued on treatment. Was this a planned treatment for this cohort? If not all patients with CIS/Ta HG at 3mo were treated, what were the determinants for patients to continue treatment?
Reply: This point was addressed in the Supplementary Methods:
"A study amendment after enrollment of 54 CIS patients and 51 Ta/T1 patients without CIS enabled patients with persistent CIS or high-grade Ta tumor at 3 months to continue on atezolizumab with a mandatory biopsy 3 months later. Enrollment after this amendment included 47 CIS patients and 20 Ta/T1 patients without CIS. Treatment was discontinued for all T1 or MIBC recurrences." (Line 4)

3.The use of blue light cystoscopy was at the discretion of the treating physician. If not used consistently, this may skew the trial results. For instance, if a patient were to be diagnosed and undergo complete resection using blue light cystoscopy at baseline, and then assessed using white light cystoscopy, he may be erroneously diagnosed with no recurrence due to the lower sensitivity of white light. What proportion of the patients underwent blue light cystoscopy at baseline vs. at 6mo biopsy?
Reply: We recognize this issue and have followed FDA guidance on addressing the use of blue light cystoscopy in the trial. We encouraged consistent use of either white light or blue light cystoscopy longitudinally in an individual patient. However, we did not capture its use as a variable that we are able to report. Anecdotally from review of source documents only a small number of centers used blue light cystoscopy at the time of TURBT, and those centers that used it used it consistently in all patients over time. We did not recognize any patients who underwent flexible blue light cystoscopy for surveillance.

4.It would be useful for the urological community at large to understand the timeline of adverse events, particularly the onset of death related events, to better maintain vigilance on patients treated with immune checkpoint blockade.
Reply: We have added the timing of death in these three patients in the text of the manuscript:
"Another 3 (1.8%) patients died, including one due to immune-related myasthenia gravis followed by respiratory failure (after 1st cycle atezolizumab), one due to immune-related myositis (10 months after discontinuing atezolizumab (4 cycles) due to high grade recurrence) and one due to sepsis (after 13th cycle atezolizumab)." (Line 260)

Reviewer #2
This is a very well-written paper and presents results of SWOG 1605, which have been presented at international meetings. The NMIBC research community has been awaiting this paper and it will be well-received and I am sure highly referenced as there are only a handful of registration trials for BCG unresponsive bladder cancer published to date. Pembrolizumab and nadofaragene firadenovac have been published and this study will add to those high impact studies even though it was terminated early because the study target sample size was met.
Reply: The CIS arm of the trial was not terminated early, and the results from the CIS arm define the primary endpoint of the trial.

Comments:
1.Please provide the initial statistical plan for the number of patients expected to be enrolled. What was the original criteria for statistical significance for the CR rate in CIS patients? Was there a pre-planned lower 95% CI bound that had to be excluded for efficacy?
Reply: This information is provided in the Supplementary Methods:
"The sample size was calculated based on the assumption of a CR rate at 6 months in CIS patients of 30% as the null hypothesis and a 50% CR rate as the alternative hypothesis based on the early AUA/FDA guidance16. With a 1-sided alpha=0.05 and 96% power the target accrual was 70 evaluable patients with CIS." (Line 20)
A similar description is provided for the endpoint EFS in Ta/T1 patients. (Line 23)

2.On page 9, CR is defined as absence of high-grade tumor on bladder biopsy and absence of upper tract or urethral disease. Can the authors also present CR based on bladder pathology alone? The number of upper tract and prostatic urethral recurrences in the CIS cohort are provided, so this should be easy to add as a calculation.
Reply: In the Results we have specified the following:
"In the CIS cohort, the sole site of recurrence was the upper tract in 2 patients and the prostatic urethra in 1 patient after 9-12 months." (Line 236)
The timeline (9-12 months) means that these were recurrences after primary CR. The CR rates at 3 and 6 months are identical whether considering bladder only or the entire urinary tract.
To clarify the point, however, we have added this text:
"The CR rate at 3 and 6 months was therefore not affected by recurrences outside the bladder." (Line 237)

3.Since only one patient with high-grade recurrence on mandatory biopsy had no reported abnormality on cystoscopy or cytology and since the detection rate of missed HG tumour in the nadofaragene trial was only 3.3% with mandatory biopsy, it is hard to know whether mandatory biopsy is truly needed for future CIS bias. Can the authors comment whether future studies should, in their opinion, mandate biopsy at a preset time point based on this date?
Reply: We addressed this point in some detail in the Discussion but are unable to provide a recommendation for future trials based on the results of this trial:
"The impact of the mandatory biopsy at 6 months in CIS patients is uncertain. Since all patients proceeded to biopsy without a preceding cystoscopy, it is impossible to determine what proportion of these patients would have had a for-cause biopsy due to abnormal cystoscopy. Many patients had subtle erythematous changes noted at the time of biopsy that may not have required biopsy, depending on the treating urologist's judgment. This, however, also reinforces the potential value of mandatory biopsy in trials where the cystoscopic appearance is subjective." (Line 307)

Reviewer #3
This is an important study. The team are to be congratulated for completion and documenting it so stringently. I am sure it will be of interest to the community.
Comments:
1.Please add line numbers to help with peer review.
Reply: Line numbers have been added.

2.Who administered the Atezolizumab? The author's list does not really reflect Medical Oncology engagement. This is important given your key concern about safety. Could it be that Urologists oversaw administration and did not monitor patents safely etc.,
Reply: The co-chair of the study was a medical oncologist. Atezolizumab was administered almost exclusively by medical oncologists at all trial sites. Monitoring and safety checks were conducted by the same medical oncologists.

3.Methods: How did you address blue light cystoscopy?
Reply: In line 155 we state:
"White light and blue light cystoscopy could both be used but investigators were encouraged to use the same technique throughout the trial in any one patient."
The FDA has recommended that blue light cystoscopy be handled in this manner.

4.Methods: Path CR- did you use cytology/washings/barbotage as well as biopsy? Please Caplin- with respect to Supp Table 2.
Reply: Since all patients had a mandatory biopsy, we used only the biopsy results to determine the CR rate at 6 months. This is defined accordingly in the protocol. It is unclear what Reviewer #3 is requesting in the second half of this statement.

5.Results: Do you have screening numbers? How many were screened, declined, judged ineligible?
Reply: We did not capture these numbers.

6.Results: In figure 1, there is 1 death. But 3 are described in the Safety results text. Also, in Table 3- what

are these Immune/drug mediated Grade 5's not Adverse events of Special Interest. Can you define Special Interest?
Reply: Figure 1 (CONSORT diagram) indicates the reason for coming off trial. Two of three deaths occurred in patients who came off trial for other reasons (recurrence or other AE) so these are not captured in the CONSORT diagram. Nonetheless, we have added a footnote to indicate how these additional deaths fit in this diagram. The reviewer is correct that two of the deaths met criteria of "special interest". This is defined by a long list of criteria established by Genentech. Since this is not of general interest to the reader, we have removed AEs of special interest from Table 3.

7.Table 2: I do not understand the footnote **One patient has Nx in the TanyN1-3 row. Please elaborate, esp. given Table S2 (4 N1+ cases).
Reply: We have moved the double asterisk to clarify this. The patient with pNx disease (i.e., the patient underwent cystectomy without lymph node dissection) is included in the pTisN0 cohort. Four patients had TanyN1-3 disease in both Table 2 and Table S3.

8.Can you add the OS figure as a third panel for Figure 2? It is good to see it in the main paper.
Reply: We have changed the figures accordingly. We have added reference to Figure 2C in the text:
"Overall survival for all eligible patients stratified by CIS status is depicted in Figure 2C." (Line 255)

9.Discussion: It is concerning that 4 patients have N+ disease at RC. Does this need highlighting in the Abstract?
Reply: In the abstract we state:
"Twelve of 129 eligible patients progressed to muscle-invasive or metastatic disease." (Line 85)
This includes the 4 N+ patients. If we were to add more detail about patients with N+ disease at RC we would have to remove other text to accommodate the word count. We would like to maintain the current content.

10.The reasons for declining RC at entry are revealing. Most cases were Patient choice- who presumably would have had RC without this trial.
Reply: Similar proportions of patients selecting trial over RC have been reported in other trials in this space. However, in the absence of a trial many of these patients receive alternative intravesical therapies, and not necessarily RC.

11.Please rename the Table S3 supplementary file. It is 'MIBC no numbers'
Reply: We have renamed file names to remove any words that the reviewer may find extraneous.

2nd Decision letter

Reference: EUROL-D-23-00287
Title: Phase II Trial of Atezolizumab in BCG-Unresponsive High-Risk Non-Muscle Invasive Bladder Cancer: SWOG S1605
Journal: European Urology

Reviewer #1
1.It would be important for the authors to specify what was done for those patients enrolled who were then deemed to be ineligible.
2.Were any AEs/SAEs encountered specifically in this patient cohort?
3.It would be important to report based on available information what the percentage of patients who were diagnosed using blue light cystoscopy was on trial.

Reviewer #4
Please see the European Urology guidelines for the presentation of statistics: https://doi.org/10.1016/j.eururo.2018.12.014. Attend to each of the guideline points below carefully and describe point by point what changes were made so that your response can be assessed by the statistical editor. In particular, see 6.4, 4.2 and 4.3.
Please also see the European Urology guidelines for reporting figures and tables https://www.sciencedirect.com/science/article/pii/S030228382030316tt?via%3Dihub. It will be helpful for the authors to read these in full. However note the following. In particular, see the comments on Kaplan Meier

plots, including avoiding censoring marks when there is a risk table, truncating each curve when the number of patients is less than 5 - 10, and keeping other text and results off of the figure plot.
Major Comments:
1.Much of the text is inconsistent or outright self-contradictory.
2.The trial is repeatedly described as Phase II.
a. Conclusions are drawn about efficacy, not the need for further research.
b. The sample size is described as small, although it is many times larger than a typical phase II.
c. Line 215 mentions a "target sample size" but there is no mention of this in the methods.
3.The first line of the discussion states that the trial "demonstrated the efficacy of atezolizumab". The next sentence but one states that the study must be considered "a negative trial" because prespecified thresholds were not met. The conclusions in the text and the abstract then make general reference to the efficacy of the treatment and the need to balance this against harms.
4.With respect to prespecified thresholds, these are not described in terms of the statistical design.
Accordingly, the authors need to decide:
- Is this phase II or not? If it is, give a sample size calculation and a decision rule, but do not draw conclusion about efficacy only about further research.
- What are the conclusions of the trial? Thumbs up or thumbs down.

2nd Author Response Letter
Response to comments from Editors and Reviewers:

Reviewer #1
1.It would be important for the authors to specify what was done for those patients enrolled who were then deemed to be ineligible.
Reply: Only eligible patients were included in the analysis of efficacy – this is stated in the manuscript (line 164-165).
Ineligible patients were allowed to continue atezolizumab. A statement to this effect has been added (line 159): “Ineligible patients were allowed to continue therapy once started if deemed appropriate by the treating investigator. They were followed for disease assessment and follow-up per-protocol.”

2.Were any AEs/SAEs encountered specifically in this patient cohort?
Reply: Any patient who received a dose of atezolizumab (n=166) was included in the safety analysis (line 164-165). We did not separate AEs by eligibility status. Because the primary reasons for ineligibility were amount and timing of prior BCG and re-resection of T1 disease, it would not be expected that AEs would vary by eligibility status.

3.It would be important to report based on available information what the percentage of patients who were diagnosed using blue light cystoscopy was on trial.
Reply: We did not prospectively capture data with respect to bluelight cystoscopy so we cannot provide this information. We do not know if bluelight cystoscopy was used to detect BCG- unresponsive CIS prior to referral to a trial center if a patient was referred after diagnosis. None of the other similar trials in this disease space have reported on impact of blue light cystoscopy. As we suggested in the first round of reviews, the general impression was that bluelight was used in very few patients, and it was used consistently at different time points in those patients in whom it was used. However, the data were not systematically collected and so are not available to report in this manuscript.

Reviewer #4
Please see the European Urology guidelines for the presentation of statistics: https://doi.org/10.1016/j.eururo.2018.12.014. Attend to each of the guideline points below carefully and describe point by point what changes were made so that your response can be assessed by the statistical editor. In particular, see 6.4, 4.2 and 4.3.
Reply: We have reviewed this document.
Section 4.2. We do not see that we have violated this recommendation about descriptive statistics.
Section 4.3. We had indicated median and range for some variables (age, number prior BCG doses, days since last BCG) in Table 1. These have been changed to median and interquartile range.
Section 4.16. We have truncated the three KM plots when the number at risk in any group drops below 5 patients.

Section 6.4 discusses small sample size as a "pseudolimitation". We cite small sample size as a limitation in our study (line 347), and we justify why we state this (line 347-348). We state that the small sample size in our trial leads to wide 95% confidence intervals which suggests uncertainty in the magnitude of the efficacy. This is the primary reason why these trial results did not lead to FDA approval of this agent for patients with BCG-unresponsive NMIBC. Most other trials in this disease state enrolled approximately 100 patients. If we observed similar results in 100 patients this agent would likely be under review by the FDA now. We therefore strongly believe the small sample size is an important limitation to cite. We also refer the reviewer to the "golden rule" in the Eur Urol statistical guidelines which states: "Break any of the guidelines if it makes scientific sense to do so."

Please also see the European Urology guidelines for reporting figures and tables https://www.sciencedirect.com/science/article/pii/S030228382030316tt?via%3Dihub. It will be helpful for the authors to read these in full. However, note the following. In particular, see the comments on Kaplan Meier plots, including avoiding censoring marks when there is a risk table, truncating each curve when the number of patients is less than 5 - 10, and keeping other text and results off of the figure plot.
Reply: We are familiar with these guidelines and have reviewed this document again. We have modified the KM plots accordingly. Numbers summarizing the outcomes from each KM plot have been added to the legend.

Major Comments:
1.Much of the text is inconsistent or outright self-contradictory.
Reply: We are reporting these trial results in the context of the established clinical trials landscape for patients with BCG-unresponsive NMIBC. As we clearly describe in the Introduction, there have been several important clinical trials completed in this disease state, and our trial is very similar to these other trials. The reviewer appears to be unfamiliar with the trial landscape and the FDA Guidance statement (reference 22) to which all of these trials adhere. There is nothing inconsistent or self-contradictory in our reporting.

2.The trial is repeatedly described as Phase II.
Reply: Indeed, because this is a Phase II trial. The reviewer is implying that it is not a Phase II trial but does not suggest what they think it is or offer justification for it to be considered something else. Although atezolizumab has been tested in many other settings, including metastatic and muscle-invasive bladder cancer, there were no preceding results on efficacy in BCG- unresponsive NMIBC. In this context, a single arm trial testing efficacy of atezolizumab primarily in 74 patients with BCG-unresponsive CIS is most appropriately referred to as a Phase II trial. The 65 patients with Ta/T1 disease without CIS are a separate cohort with a different endpoint.

a. Conclusions are drawn about efficacy, not the need for further research.
Reply: The NCI defines a Phase II trial as:
"A study that tests whether a new treatment works for a certain type of cancer or other disease (for example, whether it shrinks a tumor or improves blood test results)."
We have indeed used this Phase II trial to define efficacy of atezolizumab in patients with BCG- unresponsive NMIBC, which is consistent with this NCI definition of Phase II trial.
The FDA Guidance statement on BCG-unresponsive NMIBC (reference 22), which was based on deliberations between experts in the field and regulatory experts at the FDA, states that the efficacy of a drug as determined in a single arm Phase II trial of patients with BCG-unresponsive CIS can be used for regulatory approval of that drug. The results are compared to historical controls, with clear efficacy thresholds set by the FDA Guidance document. The prior approval of pembrolizumab in this setting further refined the efficacy bar. Since the FDA defines these as registration trials, there is no need for further research. Some trials in this disease state are referred to as Phase III trials if there was a preceding smaller Phase II trial.
One main reason for us to target European Urology with this manuscript is because the urologic reader is familiar with this trial design in this disease state. It is indeed different than in some other disease states.

b. The sample size is described as small, although it is many times larger than a typical phase II.
Reply: There is no uniform rule about sample size for a Phase II trial. We disagree with the reviewer's statement and could provide many examples of Phase II trials with a larger sample size, including the Keynote 057 that led to regulatory approval of pembrolizumab for patients with BCG-unresponsive NMIBC (reference 11). It is important to recognize that the primary analysis is in 74 patients, and the 55 patients with Ta/T1 disease represent a secondary cohort. We justify in the text and in response to the reviewer comment above why we think that the relatively small sample size is a limitation in this study.

c. Line 215 mentions a "target sample size" but there is no mention of this in the methods.
Reply: These details are provided in the Supplementary Methods (lines 19-31). Since this reviewer and one or two reviewers in the first round were not able to find the Supplementary Methods (despite two clear references in the corresponding sections of the main text), we have now moved these to the main body of the Methods (lines 187-204 and lines 208-220).

3.The first line of the discussion states that the trial "demonstrated the efficacy of atezolizumab". The next sentence but one states that the study must be considered "a negative trial" because prespecified thresholds were not met. The conclusions in the text and the abstract then make general reference to the efficacy of the treatment and the need to balance this against harms.
Reply: The trial demonstrated the efficacy of atezolizumab in patients with BCG-unresponsive CIS with or without Ta/T1 and in patients with BCG-unresponsive Ta/T1 without CIS. We report these numbers clearly in the results and reiterate the key numbers in the second line of the discussion immediately following the quoted text above. We then discuss the impact of this level of efficacy on clinical practice. We conclude that this level of efficacy does not meet pre- specified criteria, and this is a negative trial. We discuss based on the results of this trial the need to carefully consider the balance between marginal efficacy, risk of progression with delayed cystectomy and considerable toxicity for this class of drugs, which have been previously approved by the FDA and other regulatory bodies (e.g., Health Canada).

4.With respect to prespecified thresholds, these are not described in terms of the statistical design. Accordingly, the authors need to decide:
Reply: The prespecified thresholds are described in the statistical design as indicated above, and we state in the results that we did not meet these thresholds (line 270-271 in Results and 300-302 in Discussion). We also describe that the trial failed to pass the futility analysis.

- Is this phase II or not? If it is, give a sample size calculation and a decision rule, but do not draw conclusion about efficacy only about further research.
Reply: Please refer to responses above. This is all outlined clearly in the manuscript, and we have now made it easier for the reader to find important information on the sample size.

- What are the conclusions of the trial? Thumbs up or thumbs down.
Reply: This is a negative trial as described above – yet the results are similar to a previously reported trial with pembrolizumab. We therefore highlight the need to balance marginal efficacy with considerable toxicity with this class of drugs in this treatment indication.

3rd Decision letter

Reference: EURUROL-D-23-00287
Title: Phase II Trial of Atezolizumab in BCG-Unresponsive High-Risk Non-Muscle Invasive Bladder Cancer: SWOG S1605
Journal: European Urology

Reviewer #4
1.The authors say: "Section 4.2. We do not see that we have violated this recommendation about descriptive statistics." But in table 1, the authors state that e.g., in CIS, 95% of the patients were white and 5% were non-white and the reasons for no cystectomy by 7% ineligible and 93% patient choice.
2.With respect to Phase II, sample size and conclusions, the authors are quite correct that this statistical editor is unaware of the landscape of in NIMBC, but then that is exactly the point. The most typical Phase II trial in cancer is indeed much smaller than 100 patients and is used to determine the value of a Phase III trial, not for regulatory approval. If this is different in the particular setting of this study, then the authors could add a sentence or two describing that in the introduction.
3.I asked "With respect to prespecified thresholds, these are not described in terms of the statistical design." The authors state "The prespecified thresholds are described in the statistical design as indicated above". The section above refers to lines 187 to 220 and to the supplementary methods. The supplementary methods appear to have been deleted, and there is no reference to thresholds in line 187 to 220.
4.With respect to the conclusions the authors say:

a. We did not meet our prescribed thresholds.
b. Our results compare favourably with other trials.
5.Now take the question "Is Joe a good golf player?" and say that to meet the definition of "good" he has to shoot a 72. Joe shoots a 74 and you conclude he didn't meet the prespecified threshold, but his score compares favorably with other golfers. This suggests that either 72 is not the right threshold and you shouldn't have chosen it in the first place, or that 72 is the right threshold and you are trying to get around that.
6.It is of particular note that there is no reference to a failure to meet prespecified thresholds in the abstract or in the conclusion.

3rd Author Response Letter
Response to comments from Editors and Reviewers:

Reviewer #4
1.The authors say: "Section 4.2. We do not see that we have violated this recommendation about descriptive statistics." But in table 1, the authors state that e.g., in CIS, 95% of the patients were white and 5% were non-white and the reasons for no cystectomy by 7% ineligible and 93% patient choice.
Reply: We interpreted the instructions in Section 4.2 to apply to written text and not tables. We believe that these numbers add to clarity, consistency and uniform formatting in Table 1 that enhance readability but have removed them at the reviewer's request. If there are more than two variables (e.g., performance status and stage) we have retained the percentage for the last variable so that the reader does not have to do the math. Strictly speaking this goes against the instructions and we can remove if requested.

2.With respect to Phase II, sample size and conclusions, the authors are quite correct that this statistical editor is unaware of the landscape of in NIMBC, but then that is exactly the point. The most typical Phase II trial in cancer is indeed much smaller than 100 patients and is used to determine the value of a Phase III trial, not for regulatory approval. If this is different in the particular setting of this study, then the authors could add a sentence or two describing that in the introduction.
Reply: The second paragraph of the Introduction was written to provide this background. The FDA guidance document was cited for additional details for those unfamiliar with this trial design. We added "phase 2 or 3" to this description since this appears to be the point of contention.

"The clinical trials landscape for second-line bladder-preserving therapy in patients who are unresponsive or relapse after BCG has advanced rapidly[11-15] since the development of a consensus definition of BCG-unresponsive high-risk NMIBC and the standardization of optimal clinical trial design[16-18]. A joint panel of the American Urological Association (AUA) and the Food and Drug Administration (FDA) determined that a single-arm phase 2 or 3 trial assessing investigational agents in BCG-unresponsive CIS would be adequate to demonstrate evidence of patient benefit to justify introduction of new therapies into clinical practice[16]. Importantly, it is believed that CIS cannot be completely resected at the time of diagnosis, so that the complete response (CR) rate to the experimental treatment can be interpreted as a demonstration of drug efficacy. Without an effective standard bladder-preserving therapy available for these patients, there is no appropriate comparator arm for a randomized trial, and the single arm trial design has been considered the most feasible option. Outcomes after second-line single agent intravesical chemotherapy have been poor[9,19], although sequential gemcitabine and docetaxel has more recently been adopted into routine practice based on results from a retrospective multicenter series[20]."

3.I asked "With respect to prespecified thresholds, these are not described in terms of the statistical design." The authors state "The prespecified thresholds are described in the statistical design as indicated above". The section above refers to lines 187 to 220 and to the supplementary methods. The supplementary methods appear to have been deleted, and there is no reference to thresholds in line 187 to 220.
Reply: It is not clear what additional thresholds the reviewer is requesting. The relevant text is reproduced below.
"The sample size was calculated based on the assumption of a CR rate at 6 months in CIS patients of 30% as the null hypothesis and a 50% CR rate as the alternative hypothesis based on the early AUA/FDA guidance[16]. With a 1-sided alpha=0.05 and 96% power the target accrual was 70 evaluable patients with CIS. We aimed to enroll 65 evaluable patients with Ta/T1 disease. Assuming a null hypothesis of 20% EFS at 18 months, this provided 71% power to detect an alternative hypothesis of 30% EFS at 18 months using a one-sided alpha=0.05. We specified that the regimen would significantly improve EFS relative to historical trial data if the lower bound of

the 90% confidence interval for EFS at 18 months excluded 20%[16]. To enroll 135 evaluable patients, the total target sample size was 148 to allow for 10% ineligibility and drop out. The total sample size accrual goal was amended twice during the trial to account for the high rate of ineligibility of registered patients resulting in a final target sample size of 202 patients."

We described also the futility analysis in the Methods:

"Because there were no existing safety or efficacy data available for atezolizumab in patients with NMIBC when this trial was initiated, we planned an interim futility analysis based on the CR rate at 6 months in the first 25 evaluable CIS patients. A favorable evaluation allowing continuation to full accrual required ≥7 CRs in the first 25 patients, which was equivalent to testing the alternative hypothesis of a 50% CR rate with a one-sided alpha=0.02. Patient accrual was allowed to continue beyond 25 patients in the first stage while the 6-month biopsy data matured."

4.With respect to the conclusions the authors say:
a. We did not meet our prescribed thresholds.
b. Our results compare favourably with other trials.
5.Now take the question "Is Joe a good golf player?" and say that to meet the definition of "good" he has to shoot a 72. Joe shoots a 74 and you conclude he didn't meet the prespecified threshold, but his score compares favorably with other golfers. This suggests that either 72 is not the right threshold and you shouldn't have chosen it in the first place, or that 72 is the right threshold and you are trying to get around that.
Reply: The analogy could be extended. We determined that Joe has to shoot 72 to be a good golfer. But others have said that Bob only has to shoot 74 to be a good golfer. If Joe and Bob both shoot 74, Joe did not meet our pre-specified threshold but he has a similar score as fellow golfer Bob. There is no uniformly accepted threshold of what score makes a good golfer. In the case of the our trial, it is not only the score (ie the CR rate) but also the sample size which determines the confidence that the CR rate is "true". We explain in the Discussion that this is the main limitation that prevents us from labeling this a positive trial, since the response rates are very similar to other similarly designed trials. Since our trial is negative and pembrolizumab was approved with similar results, the reader can judge which trial used the more appropriate threshold.

The reviewer is suggesting that we are trying to overstate our results and put them in a more favorable light, but we are actually trying to do the opposite. Importantly, we do not use the word "favorably" (we wrote "similar to"). We acknowledge that this is a negative trial, but we also state that the rate of significant toxicity and the risk of progression are concerning, and we believe that these results are relevant to this entire class of drugs. There was widespread enthusiasm for the adoption of pembrolizumab when Keynote 057 was published, but we are issuing an implicit warning for pembro and other anti-PD(L)-1 agents:

"The results of this trial underscore the need to balance carefully treatment-related toxicity and the risk of progression with the modest efficacy of systemic immunotherapy observed in patients with BCG-unresponsive CIS…"

6.It is of particular note that there is no reference to a failure to meet prespecified thresholds in the abstract or in the conclusion.
Reply: We have added this in both locations

Accept Letter

Dear Dr. Black,

We are pleased to inform you that your above-mentioned revised manuscript has been accepted for publication in EUROPEAN UROLOGY. We will now forward it to our Publishing Department where it will undergo a desk editing process to ensure the highest quality publication.

If you have not already, you will soon receive a letter detailing the modifications that have been made by the copyeditor and those that need to be addressed when you receive the proofs from the Publishing Department.

We at the Editorial Office of EUROPEAN UROLOGY would like to personally thank you for your interest and support to the Journal and we do hope that you continue submitting valuable manuscripts to us in the future.

Thank you once again for your interest and collaboration to European Urology, The Platinum Journal.

Yours sincerely,

James Catto
Editor-in-Chief

-------- End of Review Comments --------