C. Di Rocco
Editor-in-Chief, Child’s Nervous System

Re: Cover letter for Revised “Evaluation Extent of Resection in Pediatric Glioblastoma: A Multiple Propensity Score Adjusted Population-Based Analysis (CNSY-D-14-00115)

December 16, 2015

Dear Dr. Di Rocco:

We would like to thank you for considering our revised manuscript entitled “Evaluation Extent of Resection in Pediatric Glioblastoma: A Multiple Propensity Score Adjusted Population-Based Analysis (CNSY-D-14-00115) for publication in Child’s Nervous System. We are extraordinarily appreciative to the reviewers and the editor for looking at our manuscript/study and providing this additional valuable feedback. We have used the reviewers' comments as suggestions for improving our manuscript and we have implemented these changes to our manuscript. We believe that these additional changes have significantly strengthened and clarified our manuscript and we hope the manuscript and its important conclusions will be now suitable for publication in your distinguished journal.

Following is a letter from the authors, responding in a point-by-point manner to the reviewers' comments, and a revised version of the manuscript incorporating the appropriate changes. We have made all our efforts to meet the reviewers' concerns in keeping with the letter and spirit of the comments as suggested and overall we have:

• Further discussed the limitations of the study given its retrospective nature and limitations in SEER database specifically with regards to the limited information on patient’s functional status, radiotherapy, and molecular pathology.
• Highlighted the consistency in results between separate analyses using 4 different statistical methods in the evaluation of all patients with pGB and in pGB patients excluding the “unresectable cases” such as DIPG patients.
• Advocated for the importance of the use of the multiple propensity score and simplified its explanation.
• Addressed the questions on the use of intraoperative MRI, “acceptable sequelae” of the surgery, and impact of radiation therapy in those cases.
• Slightly modified the figures (just visually, not the content) to improve them for the reader.
Reviewer #1:
We thank this reviewer for all of the insightful comments regarding our manuscript. The reviewer’s detailed analysis and recommendations were very helpful in our revision in improving the manuscript. Our itemized responses to this reviewer’s comments and the changes we made to improve the manuscript are as follows:

1. The authors report on a large retrospective series of glioblastomas in children, originating from all locations in the central nervous system (including the spinal cord), collected from a multicenter registry. They analyzed overall survival after gross total resection (GTR), partial resection, biopsy, and no treatment; the impact of irradiation was not assessed. They conclude "if deemed safe and feasible, a GTR should be the goal for pediatric patients with glioblastoma ". These conclusions are sound but not new, and a bit disappointing for such a large series. Performing a GTR for a diffuse glioma infiltrating the pons or the spinal cord, or presenting with metastases, is obviously out of the question. The real question is: when a significant resection can be done (in supratentorial or cerebellar GBM), does the benefit of resection justify taking risks of even sacrificing function (hemianopia could be acceptable for example). Considering the large number of patients, I regret the authors missed the opportunity to split the series in order to exclude cases for which GTR was out of the question, and select those for which resective surgery could be proposed, in order to evaluate at which price and for what benefit. I am unsure whether the dataset from the SEER series allows this kind of analyze, but this clearly should be the goal of a prospective multicenter trial.

The reviewer raises some excellent points and we are very grateful for the insight. There have been indeed numerous studies that evaluate the impact of extent of resection in the adult population. However, the same statement is not accurate for the pediatric population. There have been previous small-scale studies evaluating the impact of extent of resection on survival in pediatric HGG. Those, groups often mixed pediatric GBMs with anaplastic astrocytomas, which have many behavioral differences in parenchyma infiltration, molecular, genetic properties and survival rates. We wanted to specifically focus on pediatric GBM as a subgroup. Additionally, we conducted a larger cohort analysis with the use of propensity score, which allowed us to recreate a quasi-randomized setting to analyze the above hypothesis. This simulates randomization in retrospective analysis, which is particularly important in this study question given the inability and infeasibility of performing a prospective randomized study. We agree, that a prospective randomized trial is the gold standard. The SEER database is limited in the information available on the characteristics of the patients reported in the database, a specific one is the post operative neurological status. This is a limiting factor very well pointed by the reviewer and we included it in the limitations section. In regards to the differences in supratentorial vs. brainstem (diffuse infiltrative pontine glioma) in terms of their resectability and the effect of whether including them both in the study may effect the results, we addressed this by providing a separate analysis (all 4 statistical analysis) after excluding the pontine glioma population (population less amendable to resection). The effect of EOR on survival was concordant with and without the brainstem cases. We would like to emphasize that our statistical method, the multiple propensity score model, has the ability of only comparing cases that are comparable. This is the advantage of our approach above the
'standard' logistic regression approaches. When adding the anatomical location variable into the model it takes into account the location of the cases and compares it only to cases in the same location. Our analyses demonstrated the importance of the anatomical location variable, as it was returned by the model as a significant prognostic factor (as mentioned in the Results section). With this statistical approach, our intention was to 'purify' the effect of EOR in our cohort of pediatric GBM patients. To check that our EOR findings still hold up after exclusion of all the brainstem cases, we have re-run all the analyses without the brainstem cases.

(Line 188) "[...] we have repeated all analyses after exclusion of brainstem cases. In the sample without brainstem cases, the observed survival, univariate and multivariate analyses, and the final mPS model were all consistent with the findings from the entire sample"
Also please see Table 4 for detailed p values.

2. Retrospective analysis of a centralized multicenter register from 1988 to 2009. This method incurs inherent flaws, for example: how was the extent of resection established (imaging? Operative chart?). Imaging has made such advances over the period of the study that the evaluation of extent of preoperative tumor extension and dissemination, as well as the postoperative evaluation, are certainly influenced. Pathology is also a major issue to be discussed, considering the lack of centralized review, but also the length of the period considered and the vagrancies of tumor classifications during this period. Since interobserver as well as intraobserver variability is notoriously high in GBM (much higher than the figures reported from reference 32), this should be a major concern. The postmortem diagnosis concerns mostly diffuse intrinsic pontine gliomas (DIPG), and is also a problem, since the mutation and degeneration rate of these tumors is high and the postmortem grade may not reflect the initial pathology. Contrary to the authors' assertion, in our experience with biopsies of DIPG, the diagnosis is more often anaplastic oligodendroglioma than GBM.

Thank you for your thoughtful comments. We agree with the limitations of a retrospective analysis and did address specific concerns in the discussion and methods section. With regards to the specific concerns in your comments, the extent of resection was established using the SEER Program Code Manuals where surgical procedures are coded by trained coders according to established guidelines which are coded into 5 categories: no surgery, biopsy, partial resection, gross total resection and surgery unspecified. We acknowledged the limitations of the above method of assessing EOR in our discussion including concerns of cases before and after the assessment of extent of resection with modern quantitative volumetric assessment.

(Lines 285-290): "When carefully reviewing SEER's surgical procedure codes, it appears that the differentiation between excisional biopsy and partial resection is not based on quantitative measurement and effectively there is overlap. Also, cases dating from the pre-MRI era could have been misclassified as a GTR, since this is difficult to assess subjectively. However, the survival benefit of GTR for a pooled population that also contains PR cases would underestimate the treatment effect for patients who truly underwent a GTR. Ideally, all pre- and post-operative images should be acquired using a standardized protocol and reviewed centrally to ensure a uniform classification of EOR in the entire cohort. Data obtained from a retrospective
nationwide cancer registry, such as SEER, will carry (intra)institutional differences that can influence EOR coding.”

In the same section, we addressed the concern on the lack of central hisopathological review. HGG could have been overestimated but while looking specifically at GBM compared to anaplastic astrocytoma, previous studies have shown that central review of institutional diagnosis were more likely to reclassify anaplastic astrocytoma compared to GBM which is the focus of our study.

(Line 294): “Equally important is the lack of central histopathological review. Without central review the efficacy of treatment in patients with HGG could have been overestimated because of significantly better OS in patients with low-grade glioma (LGG).[1, 2] Fouladi et al. demonstrated that after institutional diagnosis of high-grade glioma, 28% had LGGs on central review.[2] However, glioblastoma patients were only reclassified in 8% of the cases, which was considerably lower than anaplastic astrocytomas (37%) and other eligible malignant gliomas (34%).[2]”

It is most likely that most diffuse intrinsic pontine gliomas were diagnosed on post mortem examination given the risk of infeasibility of biopsy or significant resection. Prior studies have shown concerns that post mortem analysis may not accurately reflect the biology of these tumors at diagnosis due to additional mutations induced by radiation and chemotherapy (Robinson et al 2014). A recent study has observed that the mutational spectrum of the untreated biopsy sample was not significantly different from that of the autopsies. (Taylor et al 2014). Newer efforts of attempting to obtain biopsy samples of DIPGs allowed for further molecular and genetic studies that classify DIPG into an oligodendroglial group and a mesenchymal subgroup with the oligodendroglial subgroup carrying an expression profile most similar to the proneural group of adult GBM. (Puget et al. 2012). Yet, even with more updated information regarding the molecular profile of DIPG, there is still no consensus on DIPG subgrouping. Therefore, in our study, to prevent confounding of DIPG being miscategorized as pGBM variant, repeat analysis with all 4 statistical models were carried out with DIPG being excluded from the total sample and the results of the impact of EOR on OS were consistent even with the removal of DIPG from the cohort as can be observe in Table 4.

(Lines 184-190): “Diffuse intrinsic pontine gliomas cases, an aggressive brainstem astrocytic tumor with a poor outcome, usually end up with the histopathological diagnosis of glioblastoma.[3] Our analyses included brainstem cases, and while the mPS model allows for condensing multiple confounding variables into scores that can reduce treatment assignment bias and increase precision, we have repeated all analyses after exclusion of brainstem cases. In the sample without brainstem cases, the observed survival, univariate and multivariate analyses, and the final mPS model were all consistent with the findings from the entire sample (Table 4).”

3. The statistical method chosen (multiple propensity score) is abstruse, as shown by the recourse to online complimentary resources, and, I think, beyond the competence of most readers. The interest of this method for the comprehension of the results is not obvious, judging from table 3 in which univariate, multivariate, and multiple propensity score
analysis appear basically to show the same results. The interest of using other tests like the Independence of Irrelevant Alternative with Hausman's test is not explained. For a neurosurgeon with a basic training in statistics like me, it also looks (wrongly maybe) suspect of circularity since the data gathered from the series (the treatment that was administered) are used to control whether the same cases were allotted to the proper treatment group. I am also worried that these technical difficulties make the paper difficult to read and blur its main message. I think the authors should advocate the interest of this complex statistical method using language adapted to the journal's audience. Also, in the legends of figure 1, a brief summary of the method would be welcome in order to explain to the lay reader what these figures show: what does the mPS (0-1 scale shown in the figures ordinates) represent? If the point made by these figures is to show the overlap between the box-plots, is it necessary to duplicate the text?

More details and simpler language have been used to further explain the statistical method of propensity score analysis and its relevance and advantage in addition to univariant and multivariant analysis. The purpose of the Hausman test has also been addressed in the text. The propensity score of 0-1 in the figures are explained with the additional text. To note, the very well pointed fact that the univariant, multivariate and propensity scores show the same results, is not a circularity but a positive result that supports our conclusions.

(Lines 112-120): “Given the infeasibility of prospective randomized studies to evaluate the impact of EOR on survival in pGB patients and our reliability on observational data, propensity score matching can be a tool to control confounding factors in evaluating the effects of various EOR categories. Traditionally, the PS is estimated by logistic regression where a dichotomous treatment variable is the outcome and the covariates are the predictor variables in the model. [4] A propensity score is an estimate of the likelihood that any given individual would be in a particular treatment group given a set of measured characteristics, in our case age, period of diagnosis, anatomical location of tumor, and tumor extension These variables were selected after multiple univariate Cox regression analyses with each potential confounder as independent variable. Variables were selected if p value in univariate analysis was less than 0.1 (Table 2, indicated as “†”). In our study, propensity probabilities are calculated for each patient from the data set in order to estimate their chance of receiving GTR, PR, or Bx. (see Online Resource 1) We then evaluated the distributions of mPS by EOR groups to check for sizeable overlap demonstrating that the groups are comparable. Box-plot charts were used to visualize the comparison of the mPS distributions (Figure 1). This allows us to create a “quasi-randomized” experiment with retrospective data.”

4. References are collected using a computer-generated list, not formatted for publication in the journal (example: ref 30).

Reference formatting has been corrected accordingly

5. Figure 2 shows survival curves; please indicate whether these curves show overall survival or event-free survival

Overall survival has been specified in the figure legend
Reviewer #2:
We thank this reviewer for the positive comment that “paper is well written and well documented”, “statistical analysis using four complementary approaches is robust and original… statistic originality is a strong point of the manuscript” and that “the authors provide a comparative effectiveness study for different extent of resection groups in the largest series of pediatric GB patients”. Our itemized responses to this reviewer’s comments and the changes we made to improve the manuscript are as follows:

Major points:
1. In table 2, 7 patients with brainstem location benefited from a gross total resection, and 13 patients had a partial resection. Should these cases be considered as DIPG? If this is the case, this number of patients appears to be particularly high. It is unusual to consider that gross total resection is feasible in patients with DIPG. Considering the fact that DIPG is not a surgical disease and is clearly different from supratentorial HGG locations, I suggest removing these patients with DIPG from statistical analysis.

The reviewer has a very good point. We did perform a separate statistical analysis excluding the brainstem gliomas, as “Online Resource Table 2”. Since this important point was brought up by both reviewers, and given the importance of the matter we moved that table to the paper as Table 4. We hope that the editors agree with this addition given the importance of this table and the reviewer comments.

As stated in the paper, the results were concordant with and without the brainstem cases.

(Line 188) “[…] we have repeated all analyses after exclusion of brainstem cases. In the sample without brainstem cases, the observed survival, univariate and multivariate analyses, and the final mPS model were all consistent with the findings from the entire sample”

Minor points:
2. The most prevalent primary sites of pGB were the frontal lobe (n=59) and brainstem (n=45). I think this pattern is quite unusual, since the prevalence of DIPG is usually more important. Have the authors an explanation for that?

In the majority of papers evaluating the epidemiology of pGB, the location of the tumor is usually categorized into supratentorial versus brainstem with the majority of the tumor located in the supratentorial space. Previous statistics showed distribution of pGB to be around 50-65% supratentorial, 13-15 % brainstem, with the rest in the cerebellum or deep brain structures. In the supratentorial space, frontal lobe usually is the most prevalent followed up either temporal or parietal. The location of GBM is predominantly supratentorial in the literature, as in our series. We have added the appropriate references to the text.

(Line 141) “The most prevalent primary sites of pGB were the frontal lobe (n=59) and brainstem (n=45), with the majority of tumors located in the supratentorial space which is consistent with previously published statistics [23, 40]. ”
3. There is a discrepancy on the number of irradiated patients. In paragraph "results", it is noted that 78.4% received radiotherapy while it seems to be 71.4% in table 1. Could the authors clarify this point?

Thank you for noticing the discrepancy. It was an error in translating the table data to the text. It should be 71.4%

(Lines 146-147) “Also, the majority of pGB patients received radiotherapy (71.4%).”

4. In table 1, number of patients corresponding to the period 1995-1999 seems to be significantly lower than that of the next period 2000-2004. Have the authors an explanation for that?

Thank you for this detailed observation. In 2001, the SEER Program significantly expanded coverage by including additional geographical areas (http://seer.cancer.gov/registries/). Therefore, the number of patients after 2000 is markedly higher.

5. Analysis of the potential benefit of surgical resection is not relevant in patients with DIPG.

See above regarding separate analysis excluding DIPG. In brief, we did perform a separate statistical analysis excluding the brainstem gliomas.

(Line 188) “[...] we have repeated all analyses after exclusion of brainstem cases. In the sample without brainstem cases, the observed survival, univariate and multivariate analyses, and the final mPS model were all consistent with the findings from the entire sample”

6. The SEER data lacks information on radiotherapy techniques and dose schedules, and on chemotherapy regimens. The poor chemosensitivity is well known in HGG, so the lack of data on this subject is not really a problem. However, the lack of data on radiation therapy is a weakness of the manuscript, but we understand the reason. It would have been interesting to watch the recurrence pattern of irradiated patients (field, border?, Difference according to tumor sites and in the invasion of peri-ventricular zones?). Authors could add a sentence about it.

Thank you for the insightful discussion. The above issues have been addressed in the manuscript paragraph 8 in the discussion section. The lack of radiotherapy data in SEER is a major limitation to the study given that radiotherapy is a standard of care in the treatment of pGB and further evaluation of differences in efficacy between different dose schedule and dose amount would be insightful.

(Line 326) “Given the invasive nature of pGB and high local failure rate after surgery, adjuvant radiation therapy remain a standard of care post resection. An optimal dose of radiation has not been established and the exact dose schedule are often a bit arbitrary and comparisons of
various total dose and dosing schedules and their impact on survival are limited. Previous publications have shown that patients who receive less than 5000 cGy tend to have sore outcome than compared to standard 5400-6000 cGy. Furthermore, in patients with residual tumor after resection, higher dose of radiation may lead to significant improvement in progression free and overall survival in pGB[5]. Future studies with specific radiation treatment details could address the effect of radiation dose and dose schedule on recurrence pattern on irradiated patients and on survival.”

7. Recent advances in the biology of HGG have shown the diagnostic value of histone H3.3 mutations (H3.3K27m, H3.3G34) leading to loss of trimethylation and the prognostic impact of H3.3K27m mutation. It is possible that the mutation status is linked to the quality of surgery. Indeed the presence of the H3.3K27m specific mutation leads to DIPG (or thalamic HGG) diagnosis. These locations are considered inoperable. Could we imagine that the feasibility of surgical resection is related to the biology of the tumor? Authors could add their ideas about it.

This is a very interesting point brought up by the reviewer. In fact many molecular and pathological markers have been found to be associated with various subgroups of pGB. It is unfortunate that the SEER database does not have detailed pathological and molecular study results at this time. We addressed specifically the histone H3.3 mutations in the manuscript.

(Line 301) "With advances in molecular and genetic analysis, specific molecular markers have been implicated in the pathogenesis, prognosis, and treatment of pediatric high grade gliomas in recent studies ([6], [7, 8]). In particular histone H3.3 mutations (H3.3K27m, H3.3G34) have been identified as molecular drives for subgroups of pGBM. For instance H3.3K27 has been associated with tumors in the midline structures as the thalamus, deeper structures as the basal ganglia and brain stem and has been associated with worse overall survival compared to H3.3G34 which are more commonly found in tumors of the cerebral hemisphere and are associated with better overall survival. These findings suggest that molecular biology of the tumors may define clinically and biologically distinct subgroups of pGBM and predictor of survival with possible future impact on therapeutic trial design. However, it is difficult to differentiate whether these markers are possible predictors of survival due to basic molecular differences in tumor genesis or could be due to the location of the tumors with thalamic, basal ganglion and pontine tumors being less amendable to surgical resection with increased surgical risks therefore leading to worse prognosis. Unfortunately the SEER database does not make molecular and pathology studies readily available to allow us to further evaluate these questions. Future studies should be conducted to further address these issues. “

8. The authors conclude that GTR is independently associated with improved survival for pediatric patients with glioblastoma. This could encourage the use of intra-operative MRI to assist the surgeon. What is the authors’ opinion on this issue?

As intraoperative imaging, ultrasound, and navigation technique become more widely available and more accurate, it is important to consider the utility of these modalities to assist the surgeons
in a large but safer EOR. We have added our opinion on this matter to the discussion section

(Line 247) “As more studies show that the EOR as one of the most consistent predictors of survival, craniotomies for resection of pediatric glioma are becoming more refined in an attempt to maximize resection and minimize complications. Recent use of intraoperative MRIs has also been shown to improve the occurrence and size of residual tumor compared to no intraoperative imaging [9] Intraoperative ultrasound has also become a routine adjunct to stereotactic resection in recent years. Ultrasound has been shown to be efficacious in localizing, defining boarders of, and differentiating the tumor from cyst or necrosis and guiding the surgeon in detecting residual tumor [10] Historically, intraoperative MRIs have been used for lower grade tumors with difficult to define gross boarders and lack of enhancement on imaging to differentiate from normal parenchyma in adult gliomas. Further studies in the efficacy of intraoperative imaging on high grade gliomas would be encourage the use of intraoperative imaging for pGB resections.”

9. Finally, the conclusion of the article raises the question of "acceptable" sequelae of surgery. Authors could add their ideas about it.

Patient selection, discussion and ways of limiting complications, and balance of EOR and safety are discussed in paragraph 3 of the discussion section.

(Line 258) "With increased effort in increasing extent of resection and even with routine use of intra-operative navigation, digital tractography, cortical mapping, and intra-operative neuro-monitoring improving safe radical surgery, it is still important to understand the associated risks to glioma surgery for optimal selection and counseling of patients and their families on who may benefit the most from the surgery and to weight the benefit of radical resection and risk of neurological deficits and other sequelae from surgery.”

We thank the reviewer and the editor for your patience and feedback. We believe that these additional changes have significantly strengthened and clarified our manuscript and we hope the manuscript and its important conclusions will be now suitable for publication in your distinguished journal.

Sincerely,

Alfredo Quinones-Hinojosa on behalf of all the authors