On behalf of our multi-disciplinary team, we thank the editors and anonymous reviewers for their time and thoughtful remarks on our manuscript "Optimizing Clinical Outcome and Toxicity in Lung Cancer Using a Genomic Marker of Radiosensitivity". We were particularly encouraged by the strong support all reviewers gave as to the topic and general approach, and the soundness of the underlying gene signature. While all reviewers were generally positive, there were significant concerns regarding the derivation of the key advance, the RxRSI. Upon consideration of the reviewers' clear comments and concerns, we are happy to report that the error lies in our exposition, not in the formulation itself, and we are happy to present a clarified exposition which we believe should clear up the concerns.

In the following sections, we address each comment, from each reviewer, in turn. We include, for ease of re-review, relevant additional figures and exposition in this document, and have also made relevant edits in blue text, in the main manuscript document. Further, we have significantly expanded and revised the supplemental data section to include clarification of the mathematical exposition, a walk through of our code (which is also available on github in a linked repository), and we have appended .csv files of the relevant tables and requested data.

We appreciate the opportunity to present this, now clarified and expanded manuscript for your re-consideration. We are further eager to have the opportunity to allow our work to be vetted by the expert reviewers you have assembled, and hope to help push our field forward together.

Best,

Jake and Javier, on behalf of our co-authors.

Reviewer 1

First and foremost, we would like to thank reviewer #1 for her/his careful assessment of our manuscript. There are a number of very legitimate concerns brought up, all of which, when addressed, will make the manuscript stronger. Upon first reading, we were most concerned by the issues raised surrounding the mathematical soundness. All other issues raised seemed addressable, and indeed would make the article stronger. No changes, however, would matter, if the mathematical analysis was indeed flawed – so we first endeavoured to ascertain where either the error, or the misunderstanding, lay.

Thankfully, the reviewer laid out his/her concerns clearly, and with appropriate references to the text. This allowed us to quickly ascertain that the perceived flaws actually arise from ambiguities in our notation rather than our analysis. For continuity of thought, we will first address what we feel are the most concerning statements about the mathematics, in turn, and afterward address the reviewer's remaining comments.

Mathematical issues

1. **"I also have severe problems with the mathematical development in the paper. The equation defining α_g cannot be correct, except when $n = 1$, and $d = 2$, as RSI is defined based on SF2. If $n = 1$ and $d = 2$ then it is clear that RSI is the predicted SF2, and α_g is the resulting alpha parameter to be used in the LQ model, assuming all the variability shows up as variability in α, and none in β. Why not state this?"**

This is indeed worth stating, as we did in our Lancet Oncology paper [1]. We have yet to make any attempt at capturing changes in β, and therefore set it as a constant, and equal to a mean for a given disease site pulled from the literature. This will be amended in future iterations of our work. Further, α_g is absolutely defined with $n = 1$ and $d = 2$ as the reviewer intuits. This is the crux of the analysis: this parameter is derived from the genomics and the LQ-model (as a surrogate for SF2), and then is subsequently used to determine overall effect of repeated doses. In particular, we derive α_g from the LQ model, yielding:

$$\alpha_g = \frac{\ln \mathrm{RSI}}{-nd} - \beta d, \tag{1}$$

which, when we substitute 2Gy for dose (in one fraction), yields:

$$\alpha_g = \frac{\ln \mathrm{RSI}}{-(1)2\mathrm{Gy}} - \beta\, 2\mathrm{Gy}. \tag{2}$$

This transformation of RSI to α_g preserves the rich heterogeneity, but yields a useful quantity. We show the cohort transformed here (along with a code snippet from the original analysis to assure the reviewer that this was a purely expositional issue and not a *post-hoc* change), and will include this in the supplements in the main text.

Code Excerpt:

```
n_e = 1
d_e = 2
beta = 0.05
df <- mutate(df, alpha_g = log(RSI)/(-n_e*d_e) - beta*d_e)
```

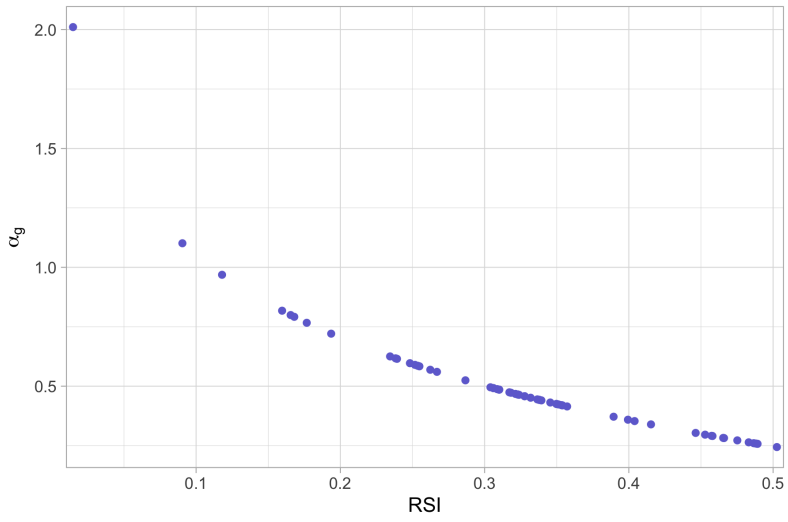


Figure 1: Each data point represents the calculated α_g as a function of RSI for each patient, demonstrating the original computation.

2. **"Another statement that simply cannot be true is this one (bottom, p. 3): "It is worth noting that in the case when a patient receives 2Gy fractions, the $\beta * d$ terms drop out, and GARD reduces to $-\ln(RSI)$." Yet RSI is a number that only depends on gene expression, and has nothing to do**

with dose, according to the first, un-numbered equation. Dose can be anything, GARD presumably changes, yet must equal $-\ln(RSI)$...?"

We thank the reviewer for pointing out this error in our exposition. We intended to define a term "GARD_{2Gy}" here. This is defined as the *effect* of a 2Gy dose of radiation in terms of GARD.

$$\text{GARD}_{2Gy} = nd(\alpha_g + \beta d), \tag{3}$$

$$\text{GARD}_{2Gy} = (1) \cdot 2\text{Gy}(\alpha_g + \beta \cdot 2\text{Gy}), \tag{4}$$

$$\text{GARD}_{2Gy} = 2\text{Gy}\left[\frac{-\ln \text{RSI}}{2\text{Gy}} - \beta\, 2\text{Gy} + \beta\, 2\text{Gy}\right] = -\ln \text{RSI}, \tag{5}$$

For a clinical GARD (which we will from here on denote GARD_c, one would have to scale this by the number of fractions given, so (assuming 2Gy fractions) this would yield

$$\text{GARD}_c = \frac{n_c}{n_e}(-\ln \text{RSI}), \tag{6}$$

where n_i where $i \in \{c, e\}$ is the number of doses given in the clinical scenario, and the experimental (by defintion 1) conditions. By failing to define n_c here, as we intended, we started what turned out to be a confusing chain of notational ambiguities which we believe led to misunderstanding our entire analysis.

For convenience, we show the calculation of GARD_c for the entire cohort here, and will include this in the supplemental material (it is also included in the github repository).

Code Excerpt: GARD_c is calculated computationally as a function of treatment dose and individual α_g.

```
df <- mutate(df, GARD_c = n_c*d*(alpha_g + beta*d))
```

3. **"For other reasons as well, the definition of GARD itself is confusing. First of all, it does not have units of dose. As the authors note that GARD has been presented before, I looked up that reference and got different definition of GARD: "The calculation for GARD is similar to biologically effective dose, except that patient-specific alpha is derived by substituting the radiosensitivity index for survival (S) in equation 1, where dose (d) is 2 Gy, n=1, and β is a constant ($0.05/\text{Gy}^2$)."**

The reviewer is correct, GARD is unitless. We apologize for any confusion about the defintion of GARD. In the previous paper, from where the quote is derived, we hoped to capture the spirit of GARD, which is indeed *similar* to BED, but **not mathematically equivalent**. Indeed, in the appendices, which we reference in the paragraph from where that quote is derived, the definition of GARD takes the same form we use here, which is the same form as *biological Effect*: $E = nd(\alpha + \beta d$, and subsequently $\text{GARD} = nd(\alpha_g + \beta d)$. We do NOT use the formalism for BED, and, as the reviewer points out in the subsequent paragraph, this would yield 'a very different kind of GARD'. So, utilizing our new, hopefully clearer, notation:

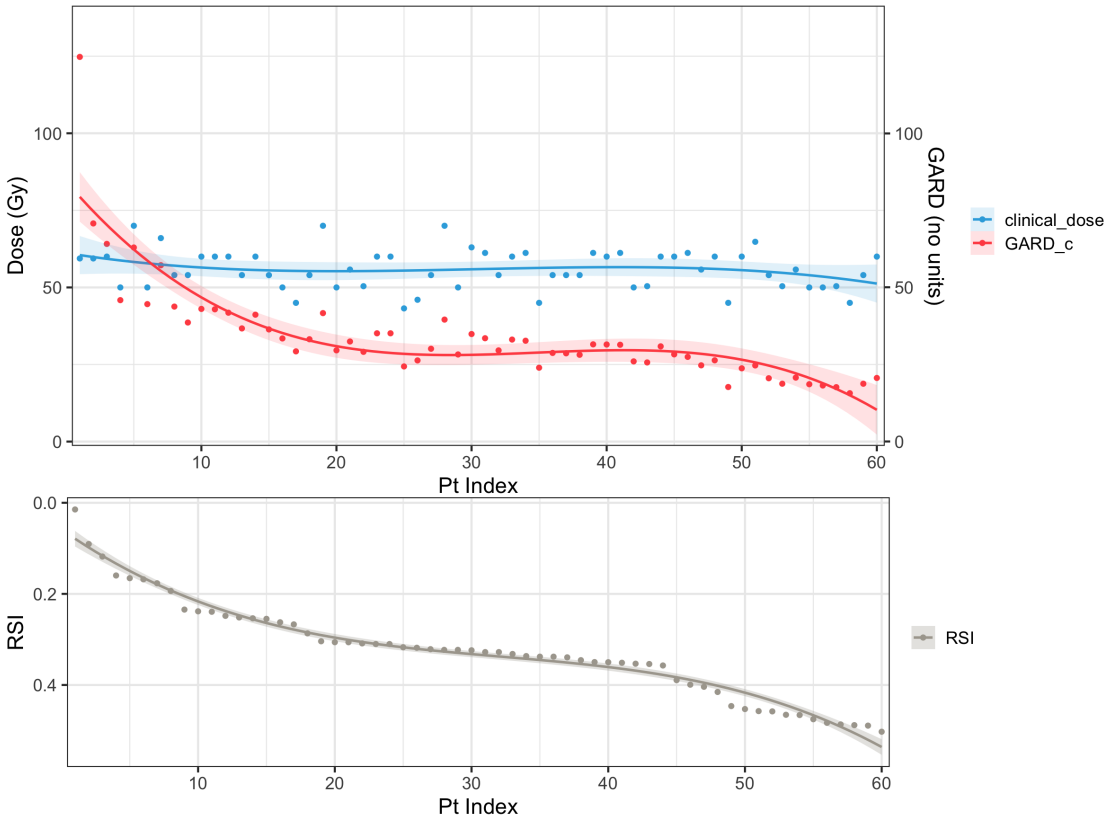


Figure 2: Each point along the x-axis represents an individual patient. RSI for each patient (matched vertically between the plots) is plotted underneath in order to demonstrate the relationship between between the two parameters.

$$\alpha_g = \frac{-\ln RSI}{n_e d} + \beta d \quad \text{and,} \tag{7}$$

$$\text{GARD}_c = n_c d(\alpha_g + \beta d). \tag{8}$$

4. **"Confusion extends to RxRSI, which is repeatedly called a biologically optimized dose, yet the definition in the paper of RxRSI is given as $\frac{\text{GARD}}{(\alpha+\beta d)}$, which by definition apparently always reduces to nd. This equation obviously does not involve biology."**

The reviewer has caught the final, and most important error in our exposition. As written, the reviewer is entirely correct, we have essentially 'cancelled out' any biology from the formalism. Luckily, our mathematical exposition here does not match our computation, which does NOT cancel the biology, and the solution to clarifying the exposition is as simple as introducing some new notation to clear up an ambiguity. The crux of the matter is that the GARD in the original equation for RxRSI is a static number. Once we derived GARD_c for every individual patient in the cohort (i.e. using each individually derived α_g and the actual dose of radiation received: $n_c d$), we created a distribution. This distribution was then used to find an

Table 1: **Cox Regression on linear GARD$_c$ term.**

coef	exp(coef)	se	z	chi-sq	p
-0.04	0.96	0.02	-1.82	3.31	0.07

Table 2: **Cox Regression on stratified GARD$_c$.**

coef	exp(coef)	se	z	chi-sq	p
1.23	3.44	0.51	2.41	5.80	0.02

optimal GARD threshold (GARD$_T$), which minimized the $p-$value of the Kaplan-Meier statistic using the clinical outcomes.

For the cohort used in this manuscript, for which we identified a GARD$_T = 33$, we show this analysis here, and include it in the supplements of the manuscript, and include the code in the linked github repository.

Code Excerpt:

```
km <- survfit(Surv(time,event) ~ 1, data = df)
```

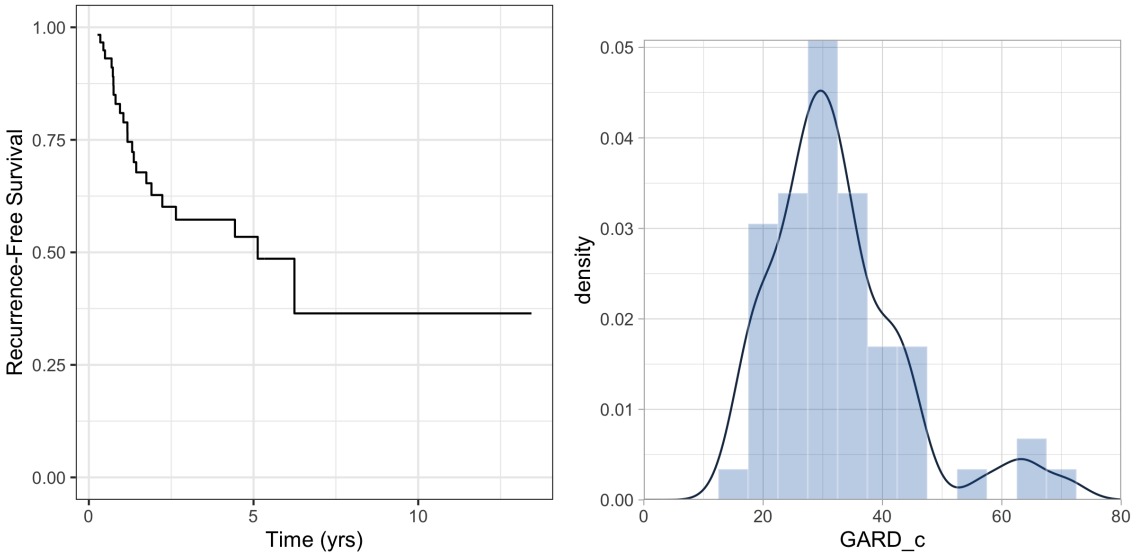


Figure 3: The left plot below shows the Kaplan-Meier plot of the 60 patient cohort. The right side shows the kernel density estimate of the calculated GARD$_c$. These are the same values shown in Figure: 2 (top panel).

Code Excerpt: Cox proportional hazard models are shown below in order to briefly illustrate the relationship between GARD$_c$ and outcomes, first as a linear variate (table 1) and then with the derived threshold value (table 2).

```
fit <- cph(Surv(time,event) ~ GARD_c, data = df,x = T, y= T )
```

```
fit <- cph(Surv(time,event) ~ GARD_c<33, data = df,x = T, y= T )
```

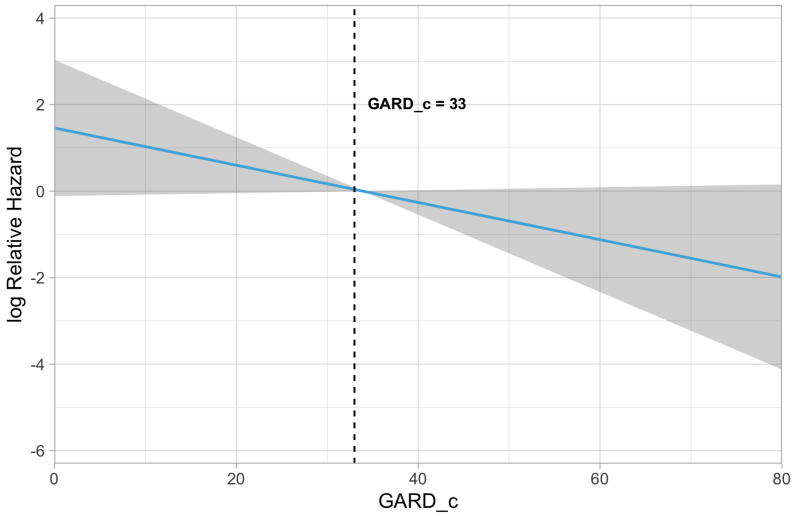


Figure 4: caption here

```
km <- survfit(Surv(time,event) ~ GARD_c<33, data = df)
```

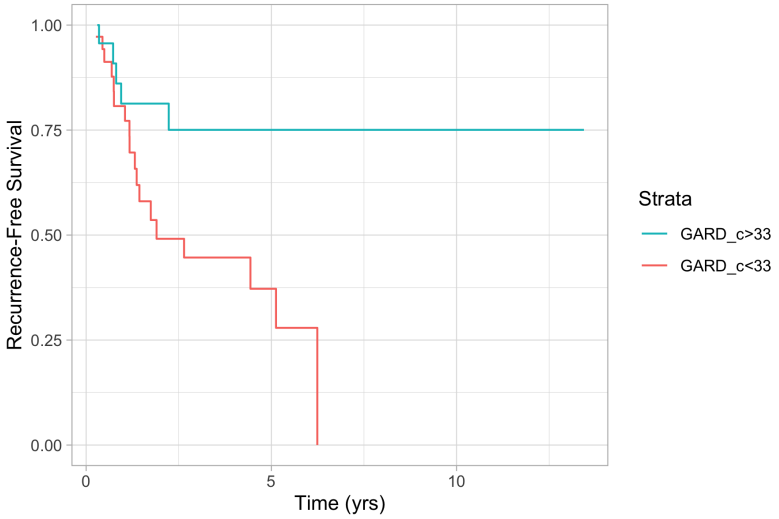


Figure 5: The Kaplan-Meier curves demonstrate the described stratification with the derived GARD_T value, as applied to this patient cohort.

At this point, GARD_T is a **set, static number: a target biological Effect to achieve through physical dosing**, it no longer has the α_g term in it. Therefore, when we define RxRSI, using GARD_T,

$$\text{RxRSI} = \frac{\text{GARD}_T}{\alpha_g + \beta d}, \tag{9}$$

we are NOT cancelling out the biology, but actually reinserting it (note here also that the units of RxRSI are now Gy). For each patient, we ask what physical dose (nd), modulated by their biology (through α_g) is needed to reach this **target effect (GARD_T)**? This yields another opportunity for clarifying notation, by introducing n_T, a target number of doses.

In summary, we immensely the appreciate the careful and detailed reading of our manuscript, and apologize for the confusion surrounding our mathematical analysis. We contend, however, that the issues lay only in exposition, not in computation or conception. The entirety of the preceding exposition has been added to a new supplemental section entitled “Derivation of RxRSI”.

Non-mathematical issues

5. **"The authors should be commended for tackling a problem that does not yet have a suitable treatment, namely, the incorporation of a validated genomic predictive index (RSI) into treatment planning. I find the previously published evidence that RSI is truly a prognostic and predictive index, having been shown to relate specifically to outcomes of patients treated with standard fractionated radiation, to be compelling."**

   We thank the reviewer for their confidence, and we are excited by the (long awaited) change in the tenor of the response of RSI in the community – something which has taken a decade of validation studies to create. We will be especially excited to share our new pooled analysis showing that both RSI and GARD continuously predict survival, they don't just function as dichotomizers of cohorts.

6. **"As an aside, I note that the statement "Importantly, the Lancet Oncology commission identified GARD as a research priority in the field of radiation oncology" is not accurate. That Lancet paper cited RSI/GARD approvingly as an example of a biomarker that seems to explain outcomes in terms of genetic variability. This really cannot be said to be the same thing as identifying future work specifically with GARD as a research priority."**

   This same point was brought up by reviewer 3 as well. On page 683 of the report, in the section entitled: **New technologies in radiotherapy: genomics, imaging, charged particles, and radionuclides**, they say "...candidate indicator is the genome-based model for adjusting radiotherapy dose, a composite metric involving a quantitative assessment of the expression of ten individual genes to determine a radiosensitivity index and an equation accounting for the radiation dose schedule", and reference our 2017 Lancet Oncology article [1]. We agree that it might have been an overcall to say that GARD is a research priority in this article. However, in a new publication from the EORTC [2], this has been stated more clearly. We now state this in the introduction and tone down the language about the Lancet Oncology commission.

7. **"Starting from a very high level, I do not find this manuscript to be a proposal to optimize treatments. that would imply that, for each patient, a dose was being chosen that balanced TCP and NTCP. Here, there is only a population based estimate of NTCP via a population based prediction of organ mean doses. This is a severe limitation."**

   We agree that we had not formally optimized the dose in the current work. We have now added verbiage in the discussion describing how one can use our new combined model for formal optimization (in short, solving the inverse problem to include GARD and OAR doses to maximize the survival function).

8. **"More importantly, there is a long history of thinking about how to do this in radiobiology and radiation oncology, almost none of which is referenced (e.g., Anders Brahme, Jack Fowler, Soeren Bentzen). This includes work on TCP models, little of which is referenced. The classic work of Zagars, Schultheiss, and Peters (Inter-tumor heterogeneity and radiation dose-control curves. Radiotherapy and Oncology. 1987 Apr 1;8(4):353-61.), would be applicable and adaptable, for example. If that work was used as a stepping of point, there would be a more clear attempt to fit available dose response data with reference to genomic variability. There are of**

course many more modern sources of dose response in lung cancer that are also not referenced. The general lack of scholarly references extends across TCP, NTCP, SF2, radiogenomics, and the LQ model."

The reviewer is correct that we are indeed standing on the shoulders of giants, and not innovating on our own. We have taken more care to cite a number of seminal papers from the listed authors to direct credit where it is due. An exhaustive review of NTCP and TCP modeling however, as it is outside the scope of this paper. It should be noted that the Zagars work [3] is NOT the stepping off point however, as it critically relies on the same Poisson statistics that most TCP modeling does. Our work does not require these stochastics (though when added to our work, they would likely produce noise that could better explain data – though at a resolution I would argue we can not measure). We present a first order (linear) estimate of inter-tumoral heterogeneity in the form of a personalized α. We hope that together with previous work, those luminaries in radiobiology mentioned can help us advance the field together.

9. **"Although RSI itself has been proven, beyond any doubt, to generate cut points that distribute likely outcomes on K-M analysis, it is farfetched to use the untransformed empiric distribution of RSI values as direct determinants of the LQ model α value. Indeed, given that it was derived based on *in vitro* data, it would be surprising if this was the case. More problematically, the extremely wide variance shows that it cannot be used to explain TCP curve slope directly, which has lower CVs."**

   We thank the reviewer for their confidence in RSI as a predictor of survival. We have just completed a meta-analysis (which should be posted to the medRxiv by the time of this rebuttal), which shows that RSI *AND* GARD are *continuous predictors of survival* (not just providing cutpoints). To derive the α value for the LQ model, we transform RSI through the equations described above – indeed this was the entire subject of our publication in Lancet Oncology [1]. The direct comparison of TCP curves to distributions of α (i.e. transformed RSI) is not the subject of this paper, but one that we are pursuing in another line of questioning (and which we presented at ASTRO 2019, abstract here: https://www.redjournal.org/article/S0360-3016(19)33320-6/abstract – poster here: https://figshare.com/s/314c26dd5108e61237bf).

10. **"Moving on, the selected clinical problem is not appropriate on several counts: the dataset is for post-op RT. There is no gross disease. Why would RSI apply to this case? The planning problem is certainly very different from RTOG 0617, which applied to gross disease.".**

    We agree that this is an imperfect comparison, a statement that we have now added to the discussion. Given the paucity of appropriate data however (disease matched, all data gathered), this is the best comparison we are capable of. We hope that our advancement of the model will encourage others to gather the types of data we need for future validation. Of note, we have a pending grant proposal (along with support from NTG) to analyze the tissue from RTOG 0617 directly, which will provide better validation for this specific cohort. We assert, however, that publication of this new model, with the (imperfect) comparison, is warranted to allow the theory to advance, and to allow others to complete their own analyses. Further, RSI and GARD have been applied to post-operative cases in many cohorts previously, and have been shown to correlate with both local control and survival. A forthcoming pooled analysis (watch for a preprint in the coming weeks, abstract accepted to ASTRO 2020), also shows that GARD predicts for local control and survival in every dataset we have

available – making the comparison between adjuvant and primary therapy somewhat more valid.

We are excited to help move the field toward true personalization, and realize that this can not be done by any one entity, but instead by a community of scientists standing on one anothers shoulders, and helping to advance one another's work.

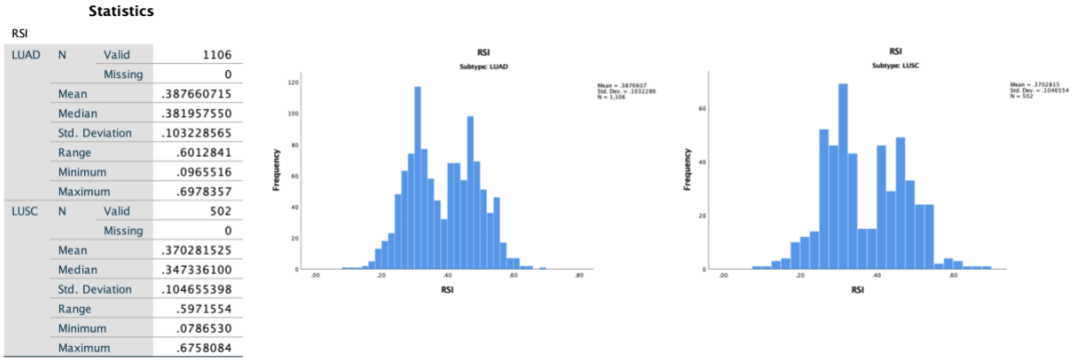
Reviewer 2 (Remarks to the Author):

We thank Reviewer 2 for their largely positive comments. We will address their comments one by one in the following. Many of the answers to the questions raised by the reviewer are NOT KNOWN, but are excellent avenues for future research. We are happy that Nature Communications provides an opportunity to publish peer review openly, so that these questions, and our responses, can be part of the record.

1. **"RSI and GARD are definitely the right direction for radiotherapy to advance from uniform dose escalation to a more biologically adapted approach. I am quite surprised that the 10 genes expression upfront correlate with outcome irrespective of all what we have learned about distinct biology of cancer entities. In NSCLC there is a large discrepancy between SCC vs. Adeno in numerous terms, also for radiosensitivity. I wonder, if these 10 Genes could separate SCC from Adeno? And is the GARD cut-point similar for both?"**

   We thank the reviewer for his/her support of RSI as an approach to biology-guided radiotherapy. Many people (including ourselves) have been surprised by the robustness that RSI has demonstrated as an agnostic predictor of radiosensitivity. However, perhaps it is critical to remind ourselves that RSI was developed using a systems biology approach and thus the 10 genes are only a sub-network within a much larger network that includes 474 genes (that is itself a sub-network of the entire signalling network). The fundamental hypothesis was that this approach would capture (at least a surrogate for) the central biology that regulates radiosensitivity across tissue types. Differences in radiosensitivity across tumor types are addressed in a separate manuscript that is currently under review but that has been deposited in the bioRxiv [4]. We include part of the first figure (showing RSI differences across tumor types) below in Figure 6. To address this concern, we looked at RSI differences between Squamous cell carcinomas and adenocarcinomas in our lung cohort. Median RSI for Squamous cell carcinomas (n = 502) was 0.347 vs 0.381 for adenocarcinomas (n = 1106) (Mann Whitney, p = 0.05), suggesting that indeed squamous cell carcinomas might be slightly more radiosensitive than adenocarcinomas as suggested by the reviewer. Regarding GARD cut-points we have not made that assessment yet, and we would reserve that for a disease specific manuscript, instead of this one where we seek to introduce several new conceptual advances.

2. **"The same apply to location, this was very relevant in past also RTOG 0617 trial, do you see differences as a function of tumor location on the RSI subgroups? In the RTOG 0617 20% of patients in both LD/HD arms suffered from pulmonary toxicity with grade >3 AEs. Therefore, despite the dose escalation the lung toxicity was not mainly different rather the impact of heart dose (V5) was considered to be pivotal for inferior outcome of the escalated dose arm. Your data show that this dose escalation was not sufficient for a large fraction of patients, and RSI would even prescribe higher doses to a fraction. While, the low RSI dose group would be spared**



Statistics

RSI

LUAD	N	Valid	1106
		Missing	0
	Mean		.387660715
	Median		.381957550
	Std. Deviation		.103228565
	Range		.6012841
	Minimum		.0965516
	Maximum		.6978357
LUSC	N	Valid	502
		Missing	0
	Mean		.370281525
	Median		.347336100
	Std. Deviation		.104655398
	Range		.5971554
	Minimum		.0786530
	Maximum		.6758084

Comparison of RSI between lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC).
(Median RSI, p = 0.05; Mann Whitney U test).

Figure 6: Breaking LUAD and SCC, we find similarities in the distribution (both bimodal suggesting two underlying groups), and differences. Median RSI 0.347 vs. 0.381, Mann-Whitney U p = 0.05.

from such a trial, how do you aim to not enhance the dose in the group requiring very high doses and what would be the V5 required to justify that they will benefit more because of their tumor biology."

We have not assessed for differences in RSI and location of the primary lung tumor (central vs. peripheral), so unfortunately we do not have the answer to this question. We have written about differences in RSI in metastatic lesions from different locations [5]. For example metastatic lesions to lung are more sensitive by RSI than metastatic lesions to liver. This analysis holds even when we limit it to metastatic lesions from the same individual. This suggests that the microenvironment influences the radiosensitivity of tumors – see Ahmed 2018 for more details [5]. However we don't know whether RSI for primary lesions in the lung vary because of location, although we have some evidence that suggest that nodal metastasis might be more sensitive. Regarding the optimization of the patients predicted to require more than 74 Gy and how to balance with heart V5, we think this will be a critical question that our model can help solve. For example, as the reviewer points out, many of these patients may need doses that may violate this safety parameter. However our model allows the quantification of the potential benefit of the dose optimization on the tumor vs. the potential harm for that individual patient. This benefit/harm ratio will be different for each patient and the "penalized model" is the first approach to calculate this. We have now included additional data, demonstrating how to use the model for each individual patient. For some patients dose escalation to the RxRSI will not be safe. These patients might be candidates for further treatment intensification with novel drugs. In addition, biological differences between RxRSI sub-groups can lead to novel understanding of how to target radioresistance in a clinically-meaningful way. Finally, the integration of the RxRSI into treatment planning will provide treatment planning software with an additional parameter to generate an optimal plan for each patient. Just like we currently include the heart V5 as a constraint to respect, the

RxRSI could be provided as the target dose. The ability to optimize treatment plans not just based on normal tissue constraints but tumor biological constraints would open a new dimension for treatment planning optimization - this is now formally mentioned in the discussion.

3. **"With advent of systemic therapy we are increasingly re-irradiating NSCLC patients, only if sufficient lung capacity is left, e.g. in RTOG 0617, almost 60% undergo salvage therapy (not counting immunotherapy), 40% were re-irradiated. How would be the strategy for your three groups in terms of residual lung capacity, cumulative dose etc. would RSI remain the same in salvage therapy. It would be great to know if data exist on dynamic switches of RSI?"**

   We don't know the answer to any of these questions but since RSI is a measure derived from gene expression, it is unlikely that it will have any bearing on residual lung capacity. This will remain an important clinical variable for determining whether a patient should be re-irradiated. In data that we will be presenting at ASTRO later this year, we have found that the RSI of metastatic lesions increases after chemotherapy treatment, suggesting that indeed there might be dynamic changes in RSI that occur as a result of intervention.

4. **"In RTOG 0617 trial IMRT was associated with substantial reduction of tox despite larger tumors being treated, possibly also biased as IMRT was given mainly at large volume centers. What are the NTCP/LC dose constraints considering RSI for larger V5 volume with IMRT and sparing of heart."**

   At this point, we are presenting the model for the first time and how this can impact dose constraints for specific populations is something we can only speculate about. However, the central point is that our findings will allow the field to evolve to a more personalized approach to dose constraints, where the toxicity cost of dose optimization can be quantified at an individual level rather than a population level.

5. **"It would be great if the authors could also provide a link to known molecular markers of NSCLC biology and resistance, it must be easily deducible from the Affy data, e.g. KEAP1, Kras, p53 status."**

   Mutational level data cannot be abstracted from the microarray. However in another study, we have looked at relationships between mutations and RSI and have not found any clear relationship between the mutational status of any gene and RSI phenotype [4]. These data are currently in revision in a peer reviewed journal.

Other comments

Introduction:

6. **"I could not find a direct link to GARD in the Lancet Oncology Paper Ref. 17 as proposed in the Introduction!"**

   Reviewer 1 brought this up as well, please see the detailed description there. In short, there is a direct link (and citation), but it is mentioned as a 'promising candidate' rather than a 'research priority' as we claimed. We have softened this language, and also added a new citation to an EORTC consensus statement about RSI/GARD [2].

Methods:

7. **"Please provide GEO/EBI public access to the Affy data, raw and processed, normalized, batch effect reduced data."**

   We did not generate the raw data used for this study, but instead it was generated as part of a large institutional repository/biobank. We have aggregated the data used in this study for all patients, and have included it as supplemental data as a .csv file.

8. **Would GARD cutpoint not differ between different NSCLC histologies, SCC vs. Adeno?"**

   See answer to question 1 (reviewer 2).

9. **"NTCP, seems to consider QUANTEC criteria for personalized RSI adapted dose, why not considering personalized NTCP? how does the RSI 10 gene reflect normal tissue response to RT?"**

   Indeed, a personalized approach to tumor and normal tissue optimization is what our combined model makes possible for clinicians. This first iteration only includes a personalized approach to tumor dose optimization while maintaining a population-level approach to normal tissue. We agree that a personalized approach to normal toxicity would be a future step, but currently lies outside the scope of this work. We are currently finalizing a proposal to NRG to assess RSI for patients in 0617. This would provide an opportunity to test the hypothesis of whether RSI can predict for toxicity as well - though it is by no means obvious that RSI (trained on tumor tissue) should predict normal tissue toxicity. We have added a statement to this effect in the discussion.

10. **"I would love to hear average and CI for absolute intensity counts and relative expression of the 10 RSI Gene across your cohort. Please provide data, how often the gene expression was close or below background in the samples. Have you imputed your data? The question is how robust the expression values of these genes within the signature was present in the samples."**

    There was no imputation. The requested data are now included in the supplemental data.

Results

11. **"Predicted and observed local failure concordance should be reported (Fig. 1C), not only on population level."**

    We are confused by the reviewers question. Figure 1C is a visualization of the predicted RxRSI for a large cohort of patients measured at Moffitt Cancer Center (the TCC cohort). There is no failure reported.

12. **"Uncertainties in NTCP estimates should be evaluated on pLC."**

    We thank the reviewer for this comment. This highlights an avenue for improvement going forward, and also strengthens the message of the work. Confidence intervals have been added to both the NTCP vs. dose plot as well as to pLC in the supplemental data. There is also a discussion here about issues regarding this error propagation, which is not trivial, and we leave for a separate work.

13. **"H0 for the Kolmogorov-Smirnov and Anderson-Darling tests are usually formulated as F0=F1, being in general the wrong metric (no testing for equivalence). Nevertheless, with the mentioned H0, p $<$ 0.001 would indicate different distributions."**

    The reviewer is correct, these metrics are not designed for discrete distributions. Thank you for pointing this out. We have removed this inappropriate statistical comparison, and instead report the mean and standard deviation of the two dominant modes of each distribution, which are extremely similar (for want of a better evaluation of comparison). The values were calculated using an expectation-maximization approach on the bimodal Gaussian distributions. This analysis can be found in Supplemental Figure 4.

14. **"PLC60 and PLC74 should be evaluated with tests for equivalence."**

    Answer: two-tailed tests for equivalence was calculated at 1, 2, and 5 years between PLC60 and PLC74.

15. **"Reported 74 Gy estimates are smaller than 60Gy estimates, overall estimates are inverted for PLC. Further, predictions do not lay within CIs of the RTOG estimates."**

    Please see the answer above, and in the new supplemental section "Error propagation".

16. **"2.4: first occurrence Fig4C should probably read 4A."**

    Answer: The reviewer is correct, and we have corrected this error.

17. **"Fig4 BC should include CIs (significant differences are stated)."**

    Done.

Reviewer 3

1. **"It can be of benefit to many readers if the authors can provide a brief discussion about the genomic marker of radiosensitivity among patient populations in general."**

    We agree with the reviewer. We have added additional discussion of RSI in the background and discussion, and included more references in the introduction to delineate other work to demarcate radiation effect from genomics.

2. **"GARD is calculated using the individual RSI and the LQ model. The latter includes the radiosensitivity of a tumor type, i.e., alpha/beta, which can also be patient specific. How does the uncertainty of alpha/beta play out for the accuracy of GARD calculation?"**

    We agree that an estimate of uncertainty is critical and can have an important impact on our calculations. To be clear, we do not use the "radiosensitivity of a tumor type" (i.e. low $\frac{\alpha}{\beta}$ for prostate cancer, ≈ 3), but rather always calculate a patient specific α_g using the RSI calculated for each patient based on gene expression (see clarification in comments to reviewer 1 concerns). Since β is always a constant, the main uncertainty would be introduced by α. In laboratory studies, done as part of CLIA validation, we have estimated that 90% of the technical variability for RSI is within 0.1 (absolute value).

3. **"Several patients had the calculated RxRSI (the physical dose required to achieve an optimized biological outcome) as low as 15, 30 or less than 40 Gy shown in Figure 1 (B). What were the actual local control and treatment dose for those cases? Please elaborate ..."**

   The actual doses delivered for each patient are presented in figure 2 in the response to reviewer 1. Patients that achieve a GARD above 33 have an improved local control as can be seen in the accompanying Kaplan-Meier curve in figure 5.

4. **"Interesting to see the red region in Figure 1 (C) where 42% patients needed RxRSI greater than 74 Gy and most of them had the failure of local control. Did the authors look at other factors of failure, such as, tumor size and PTV margin size, etc.?"**

   To clarify, in this figure we are presenting the distribution of calculated RxRSI for a representative population of lung cancer patients. This distribution suggests that a significant proportion of patients with lung cancer require more than 74 Gy to achieve biological optimization. The proportion of patients requiring 74 Gy is similar to the proportion of patients with local failure after RT in RTOG 0617, suggesting that this distribution is consistent with known clinical data – and indeed the distributions are very similar. We agree with the reviewer that there are other factors that may impact failure, however at this point we are only including genomic features in our model. We have added a statement pointing out this caveat in the discussion addressing this concern.

5. **"The authors suggested adjusting the dosing based on RxRSI calculation to reduce the toxicity to the normal tissues for sensitive and intermediate patients, but not for resistant patients. Do we know if resistant patients would have the similar level of radiosensitivity in normal tissues to the tumor?"**

   The competing hazards model presented in the paper assumes that there are no differences in radiosensitivity in normal tissues between patients. But indeed the reviewer addresses a critical question that at this point remains unanswered. The main rationale for our proposal to adjust dose for sensitive and intermediate patients (but not resistant) is because this adjustment can be done within standard of care for a significant proportion of these patients. However resistant patients require doses above 74 Gy in the setting of primary RT, so we think it is easier (and safer) to start with sensitive and intermediate patients.

6. **"Hypofractionated SBRT can also be an option to achieve higher TCP and lower NTCP. How would the authors address tumor radiosensitivity for much higher dose per fraction?"**

   Indeed this is an important clinical point. We recently published a paper in JTO addressing this issue [5]. As we stated in response to reviewer 1, we are limited with this formalism to standard fractionation without more information about β, a shortcoming we are eager to address!

7. **"Another approach to personalized RT is to use machine learning (AI algorithms) for the outcomes analysis, which presumably would include more variables. What are the authors views on the future development?"**

   We agree there is great promise from AI and have added a reference and statement to this effect in the introduction.

8. **"Suggest rounding the dose values by two decimals in the tables as the third decimal has little clinical relevance."**

   While we agree that this level of precision is likely not clinically relevant, we would prefer to leave it to the reader/re-analyzer to choose where to truncate these data. We have further moved these data into a .csv for convenience, and so the reader can truncate easily to whatever level of precision they desire.

9. **"The paper stated radiation therapy serves as the backbone for nearly 40% of all cancer cures, and is received by nearly 70% of all cancer patients at some point in their cancer journey in the abstract, and Approximately, 50% of all cancer patients receive RT. What was the latest and accurate numbers?"**

   We have made these statements consistent.

References

[1] Jacob G Scott, Anders Berglund, Michael J Schell, Ivaylo Mihaylov, William J Fulp, Binglin Yue, Eric Welsh, Jimmy J Caudell, Kamran Ahmed, Tobin S Strom, et al. A genome-based model for adjusting radiotherapy dose (gard): a retrospective, cohort-based study. *The Lancet Oncology*, 18(2):202–211, 2017.

[2] Gillian Thomas, Elizabeth Eisenhauer, Robert G Bristow, Cai Grau, Coen Hurkmans, Piet Ost, Matthias Guckenberger, Eric Deutsch, Denis Lacombe, Damien C Weber, et al. The european organisation for research and treatment of cancer, state of science in radiation oncology and priorities for clinical trials meeting report. *European Journal of Cancer*, 131:76–88, 2020.

[3] Gunar K Zagars, Timothy E Schultheiss, and Lester J Peters. Inter-tumor heterogeneity and radiation dose-control curves. *Radiotherapy and Oncology*, 8(4):353–361, 1987.

[4] George Daniel Grass, Juan Carlos Lopez Alfonso, Eric A Welsh, Kamran Ahmed, Jamie Teer, Louis Benjamin Harrison, John Cleveland, James Mule, Steven Eschrich, Heiko Enderling, et al. Harnessing tumor immune ecosystem dynamics to personalize radiotherapy. *bioRxiv*, 2020.

[5] Kamran A Ahmed, Jacob G Scott, John A Arrington, Arash O Naghavi, G Daniel Grass, Bradford A Perez, Jimmy J Caudell, Anders E Berglund, Eric A Welsh, Steven A Eschrich, et al. Radiosensitivity of lung metastases by primary histology and implications for stereotactic body radiation therapy using the genomically adjusted radiation dose. *Journal of Thoracic Oncology*, 13(8):1121–1127, 2018.