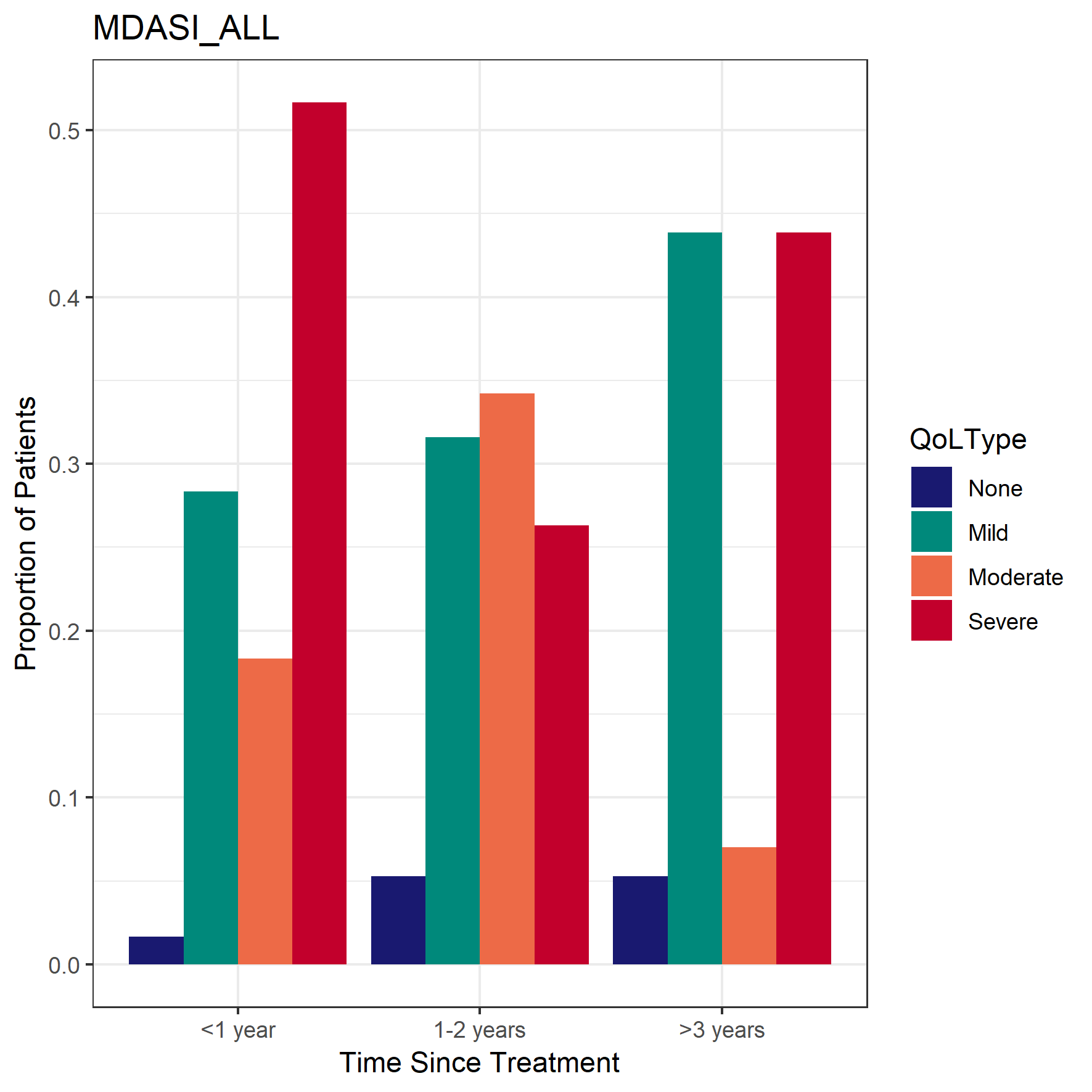
Towne: Assessing patient-reported symptom burden of long-term head and neck cancer survivors at annual surveillance in survivorship clinic

MDASI-H&N

Previously published and conventional thresholds were used to define symptom severity: symptom free (all ratings zero), mild (all ratings s zero), mild (all ratings < 5 with at least one rating ≥1) moderate (all ratings <7 with at least one rating >=5) and severe (at least 1 rating >=7)

SW: Repeat figure 1 and figure 2

Figures 1 has been reproduced with our data below – following similar 1-year intervals as in Townes, and plots MDASI total score. Similar figures specific to each PRO item have been saved to a folder (Townes Figure 1).



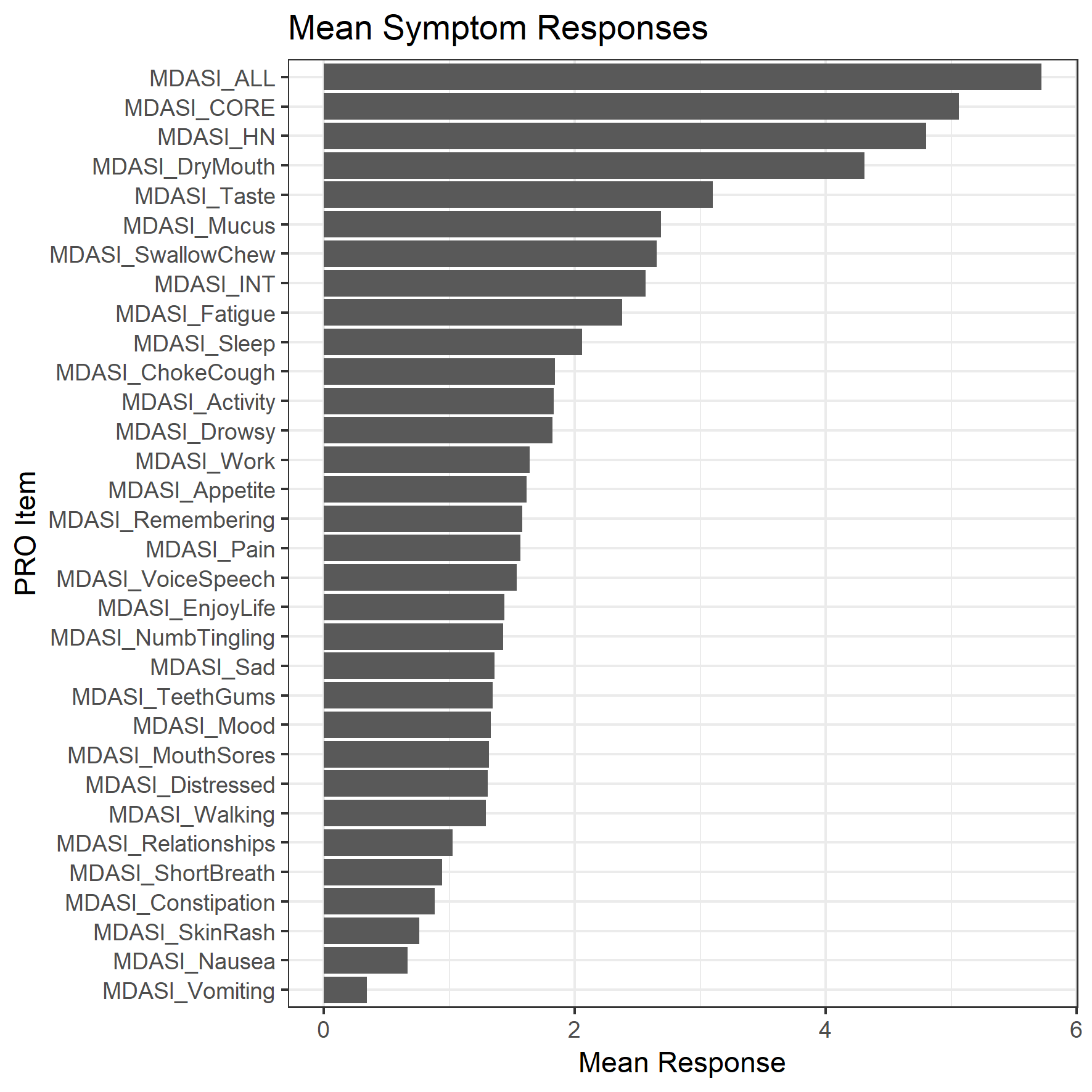


Figure also saved to Townes Figure 2 folder.

McDowell

Patient-reported quality of life and toxicity in unilateral and bilateral radiotherapy for early-stage human papillomavirus associated tonsillar carcinoma

e EORTC QLQ-C30 module and the MDASI-HN

43 patients only early tonsillar carcinoma half bilateral rt – test benefit to QOL of unilateral vs bilat RT

The MDASI-HN assesses cancer symptoms (22 items) and their interference on daily activities (six items) [17]. The symptom items include 13 general and nine head and neck cancer symptoms. The mean symptom severity score is the average of the 22 symptom items and the mean symptom interference score is the average of the six interference items. The minimally important difference (MID) is estimated at 0.98–1.21 [17]

Report mean and SD, and estimate difference with CI and p-value in a table.

We noted medium-sized differences in global health status/QoL (84 v 69, p = 0.0005), social functioning (93 vs 78, p = 0.033) and cognitive function (86 v 75, p = 0.129), in favour of URT [16]. The differences in physical and role functioning were considered trivial.

SW – repeat table 2 high dose/low dose

Note that they difference they saw in dose was 26Gy vs 10 Gy for the contralateral PG and 49 v 39 Gy for pharyngeal constrictor. Our doses differences are way smaller.

SW – repeat table 3 – ranking of difference between techniques (here differences between doses)

SW – in theory, the range of differences in doses achieved by ART is captured in your data set. If you can show those differences (which ones?) matter to QOL, then you indicate ART goals.

I like this as a potential way to summarize the effect of dose differences and have mapped out a possible example below. Tables reporting both the impact of planned dose and delivered dose may help to compare. (entries = planned (#)/delivered (#)). We have 4 OAR of interest, parotid glands (previously condensed to min PG dose), pharyngeal constrictor, brainstem, spinal cord. Perhaps a column to specify OAR to combine results into one table (if used in the paper then helps with table limit). For MDASI, supplement CI with proportions/odds ratio of moderate/severe? Need to set reporting periods, paired measures, comparison approach then table will be much more streamlined. Tables take me the most time because they’re so manual, so will circle back to this one..

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measure | OAR | Radiotherapy Dose | | Estimate Difference (95% CI) | Clinical Impact of Difference | p-Value |
| Met Planning Objective | Exceeded Planning Objective |
| E.g., CORE | Parotid glands | X (x)/ Y (y) | X (x) /Y (y) | X (x)/ Y (y) | E.g., minimal/moderate |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Hutcheson

transoral robotic surgery 257 patients, (75 underwent TORS and 182 received radiotherapy)

MDASI-HN

MDASI swallowing symptom severity item scores were significantly worse in the post-TORS group compared with postinduction (mean [SD] change, 2.6 [1.1]) and treatment-naive (mean [SD] change, 1.7 [0.3]) patients. This result inverted at radiotherapy end, and all groups converged at 3 to 6 months.

SW - They did the converse(?) study as well – patients with and without moderate severe dysphagia – what is the difference in baseline, clinical stage, etc.

Table of numeric/factor tests, repeated for individual none/mild/moderate/severe and non/mild vs. moderate/severe. Ran previously with non/mild/moderate vs. severe and found that PROMs as a dependent factor variable were most promising (in line with Hutcheson *et al*.) when accounting for follow-up time (to start, raw univariate data is included below).

|  |  |  |  |
| --- | --- | --- | --- |
| Independent Variable | Dependent Variable | Statistical Test | Significance |
| Chart Parameters: Factor - gender, cancer site, T stage, N stage, chemo agent, HPV status, alcohol intake, smoking history, PEG/NG tube; Numeric – months since treatment, age, ECOG, Charlson comorbidity score, initial BMI, change BMI | MDASI subgroups (CORE, HN, INT, ALL); looking at none vs. mild vs. moderate vs. severe and none/mild vs. moderate/severe | Kruskal-Wallis for numerical indep. variables, Fisher’s Exact for factor indep. variables | MDASI INT: none vs. mild vs. moderate vs. severe significantly different based on months since treatment (p = 0.002), ECOG (p = 0.004), and initial BMI (p = 0.007); none/mild vs. moderate/severe significantly different based on initial BMI only (p < 0.001) |
| Geometric Parameters: CBCT measurements, initial volume, change in volume, final volume, centre of mass shifts | MDASI subgroups (CORE, HN, INT, ALL); looking at none vs. mild vs. moderate vs. severe and none/mild vs. moderate/severe | Kruskal-Wallis for numerical indep. variables, | MDASI INT: none/mild vs. moderate/severe significantly different based on final BMI only (p < 0.001); some parameters correlated with BMI (face diameter, neck diameter had p<0.05 before multiple testing corrections but not after) |
| Dose Parameters: Numeric – planned, delivered, and violation | MDASI subgroups (CORE, HN, INT, ALL) and individual items; raw numerical responses | Kendall correlation | No significant results |
| Dose Parameters: Numeric – planned, delivered, and violation | MDASI subgroups (CORE, HN, INT, ALL) and individual items; looking at none vs. mild vs. moderate vs. severe and none/mild vs. moderate/severe | Kruskal-Wallis | No significant results |

Tyler

114 patients

One hundred fourteen patients with sinonasal and nasopharyngeal malignancies received surgery, radiation, systemic chemotherapy, or a combination thereof, with curative intent. Validated global ([EuroQol-5D] Visual Analogue Scale [EQ-5D VAS]) and disease-specific instruments (MD Anderson Symptom Inventory–Head and Neck [MDASI-HN], Anterior Skull Base Questionnaire [ASBQ]) were administered to patients who were both free of disease and had completed treatment at least 12 months previously. Associations between instruments, instrument domains, and specific clinical parameters were analyzed.

SW – were our patients free of disease? May consider for subgroup analysis. Also re-irradiation

“ Subjects included were ≥ 18 years of age; alive; and completed a treatment regimen consisting of either surgery, radiation, systemic chemotherapy, or a combination thereof delivered with curative intent.” Best part of my day so far.

Adam and Demetra – Tyler, Figure 1 is a reasonable starting point to consider where to go with this study as well.

Our patients were free from loco-regional disease at the time of survey completion; no patients had repeat irradiation of the head and neck. A few patients had been treated with neck dissection for residual disease, and some had distant metastasis (e.g., lung). Fortunately, all of ours were also alive when they completed the survey.

Kamal

SW – ranking of most severe items seems to be a result as well (e.g. differ between low/high PG dose groups)

SW table 2 has mean, sd and univariate, multivariate regression with FOIS (functional oral intake scale). Also P value

Suggest the same for PG, PC doses.

SW Figure 1 – I really like this as a way of presenting general results. Possibly order symptoms by severity occurance in them – or group.

SW Fig 3 – have equivalent, I think. Review how they write this up. % patients in each cluster.

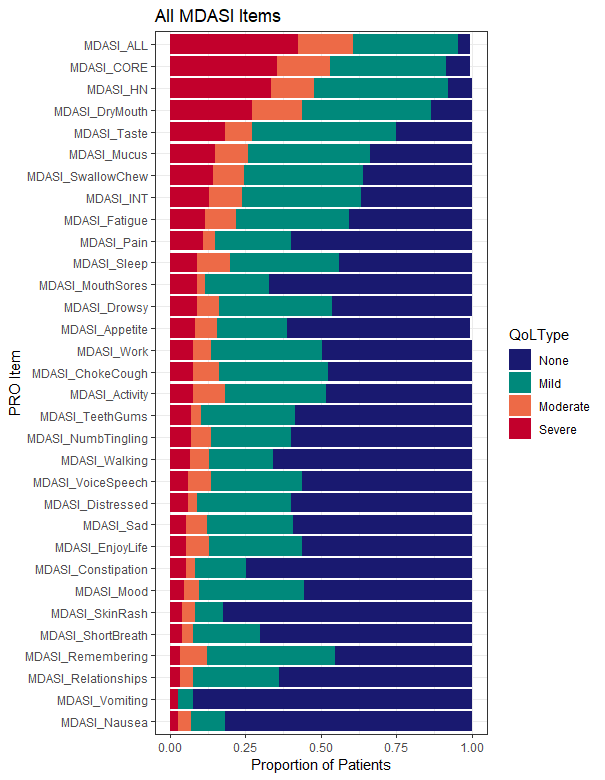
SW fig 4 – this is what I had in mind for a first review of dose vs severity of score.

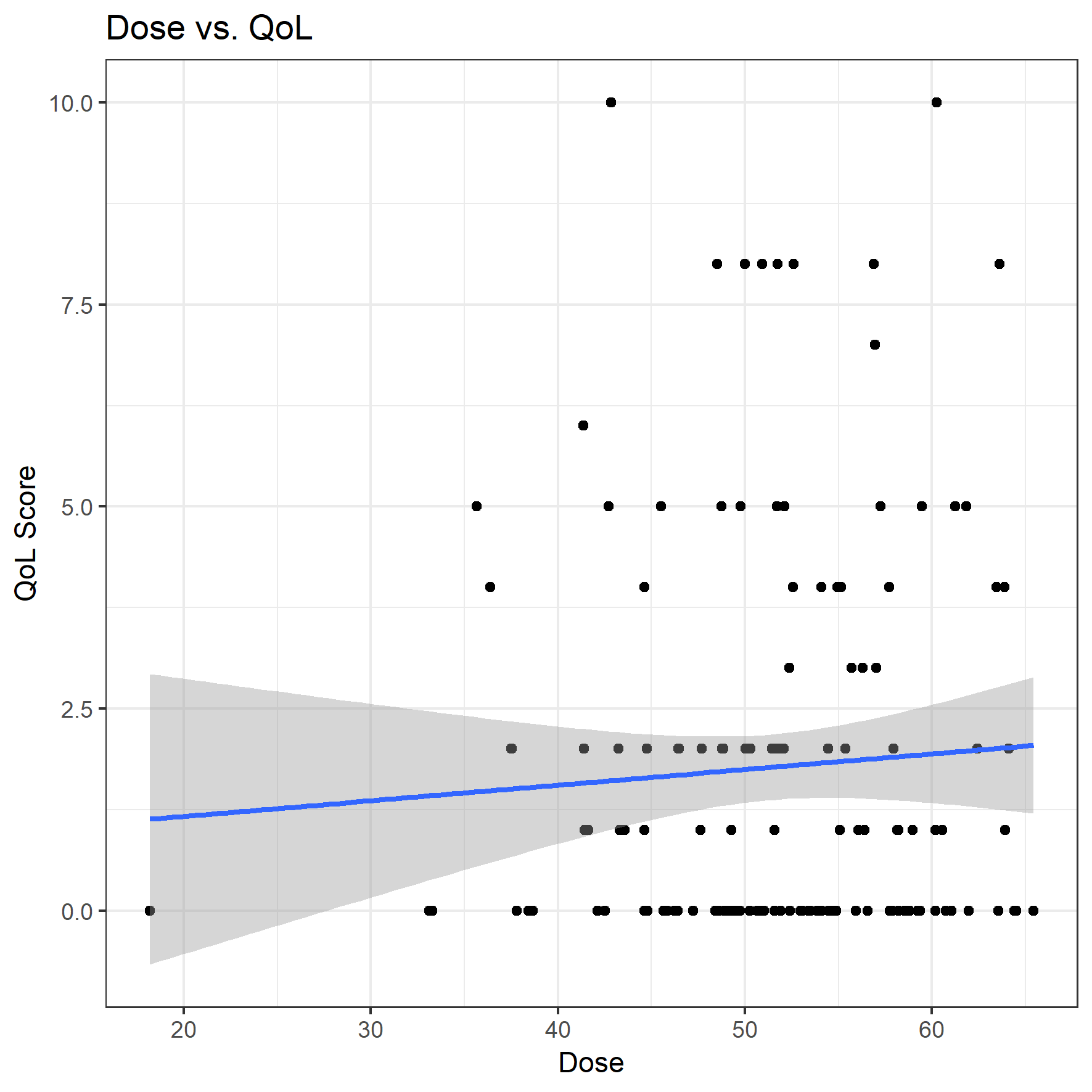
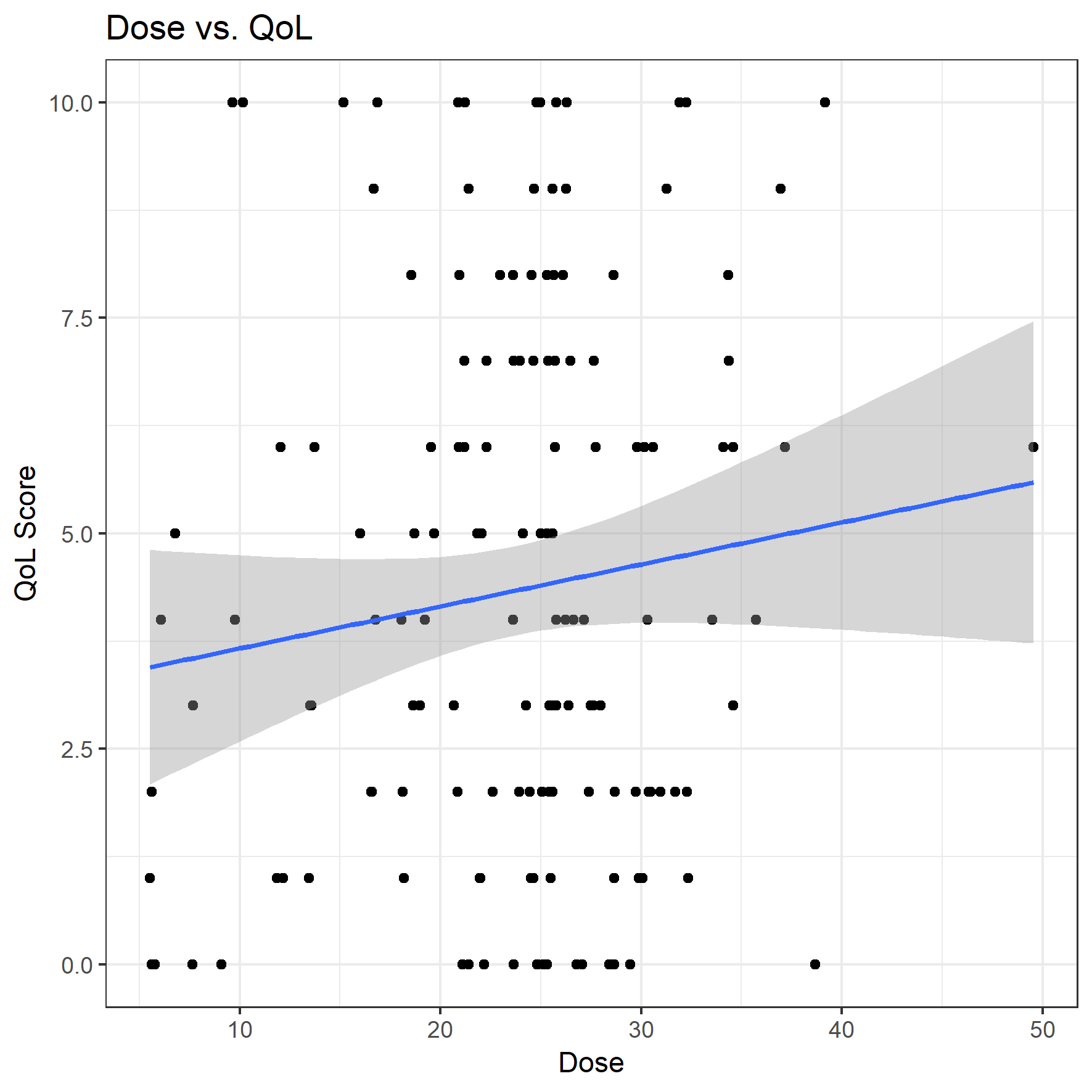
Reproducing Figure 1 (next page), with rows sorted by decreasing proportion of “severe” results, also with dose comparisons and dose/time comparisons saved to the Kamal Figure 1 folder.

Table 2 similar to McDowell but with regression (confounded by noise for our data). Regression figures (similar to Figure 4) included below and in graphs folder (Kamal Figure 2\Planned Dose, Kamal Figure 2\Delivered Dose, Kamal Figure 2\Violation).

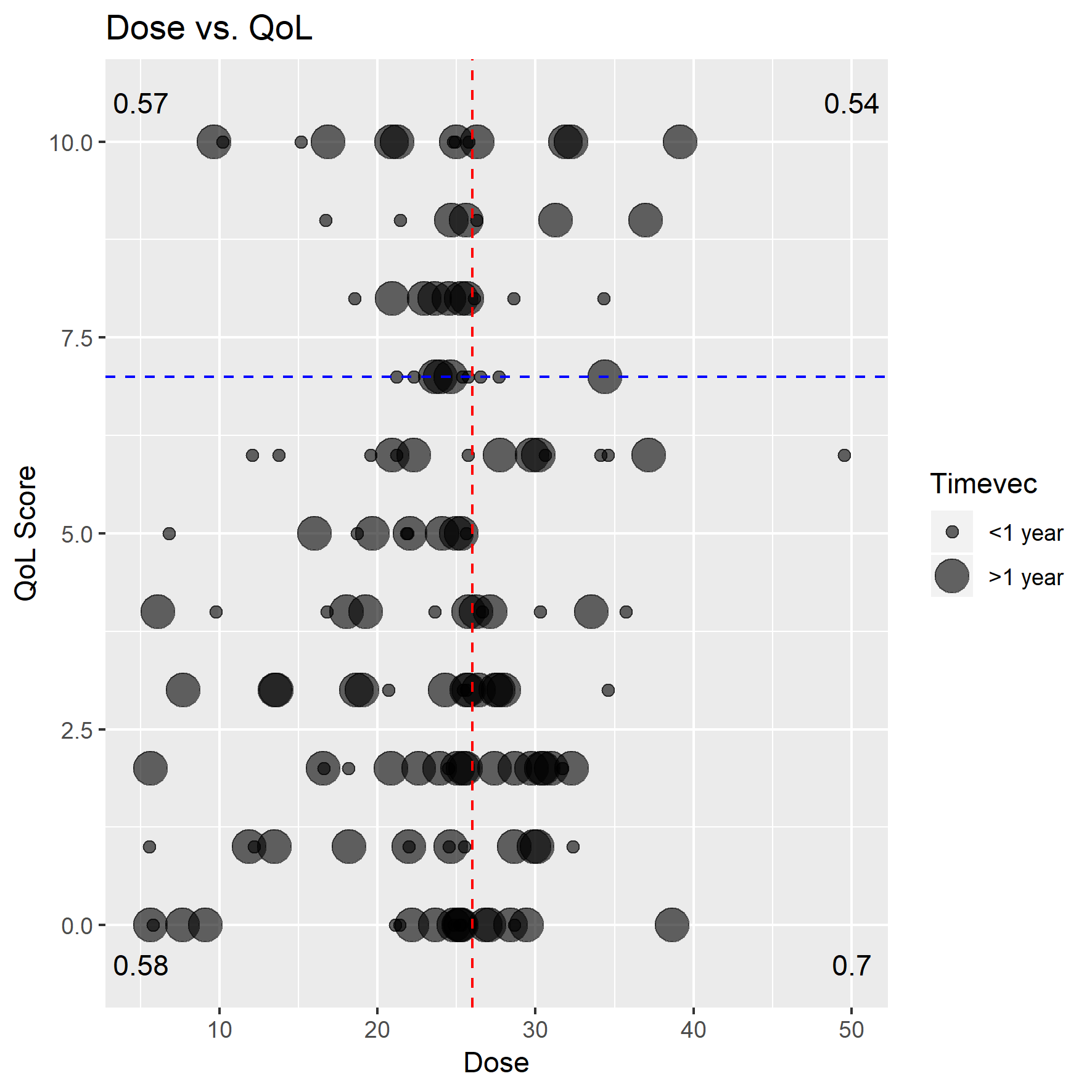
Scatter plots without regression lines, but with relative proportion patients in each quadrant have been saved to the Kamal Figure 2\Categorical – Planned Dose, Kamal Figure 2\Categorical – Delivered Dose, and Kamal Figure 2\Categorical – Violation folders.

Figure similar to Kamal’s Figure 3 has been included below (Eraj section).

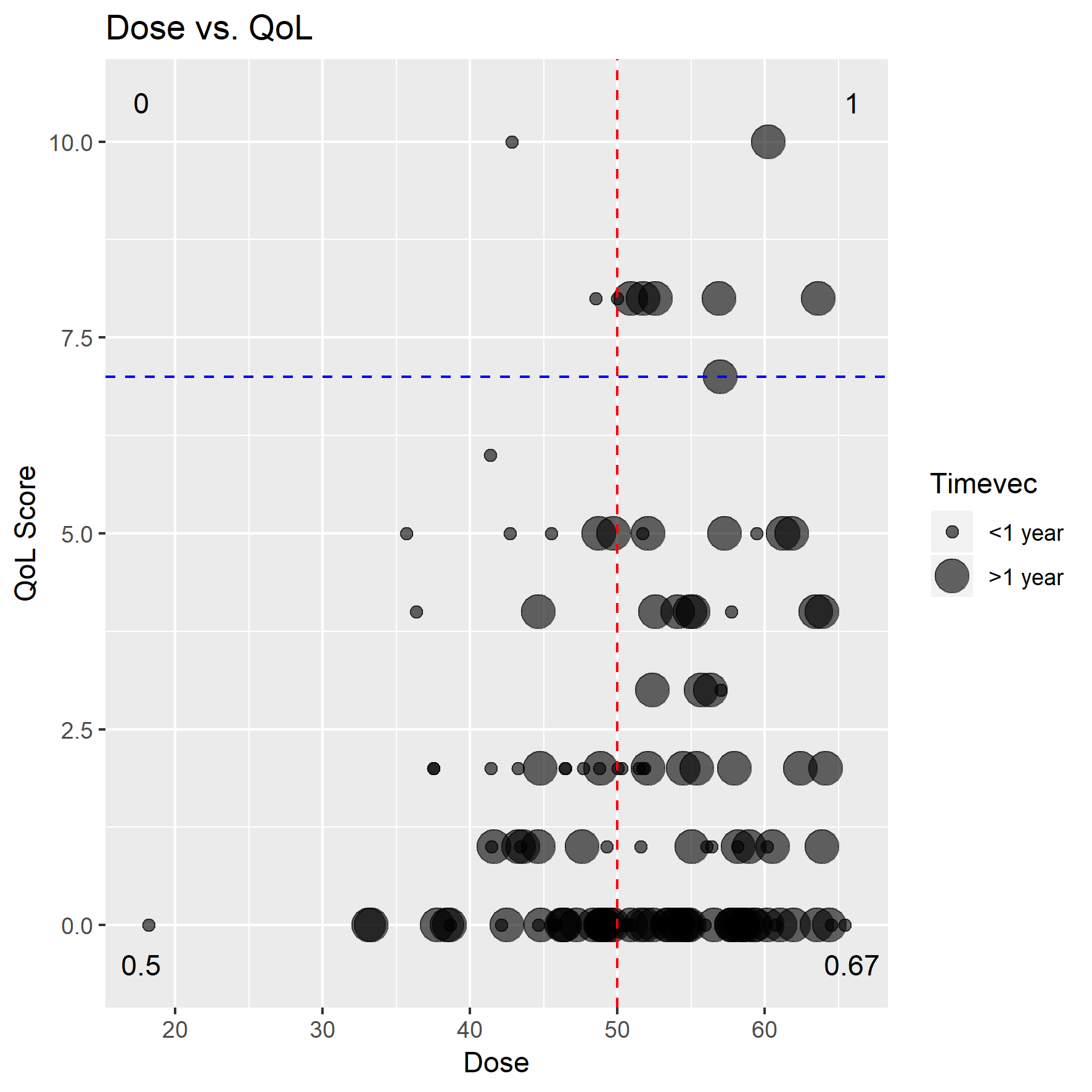




Left: planned parotid gland dose vs. MDASI dry mouth item. Right: planned pharyngeal constrictor dose vs. MDASI choking/coughing item. Not strong enough for thesis or paper, but logistic regression is a bit better (included in Sapir section below).



Dose/QoL association not clear



Dose/QoL association is clearer

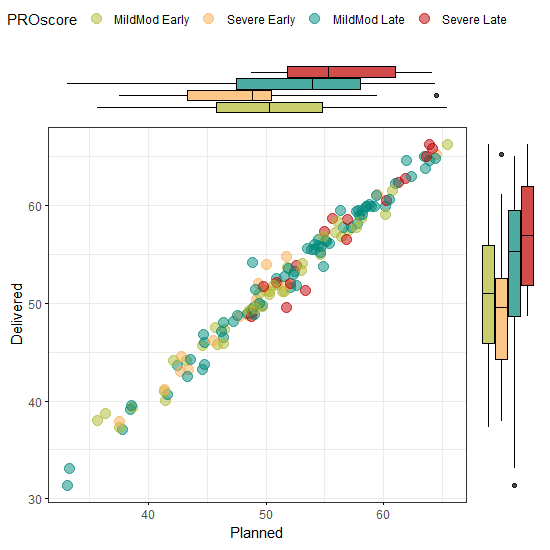
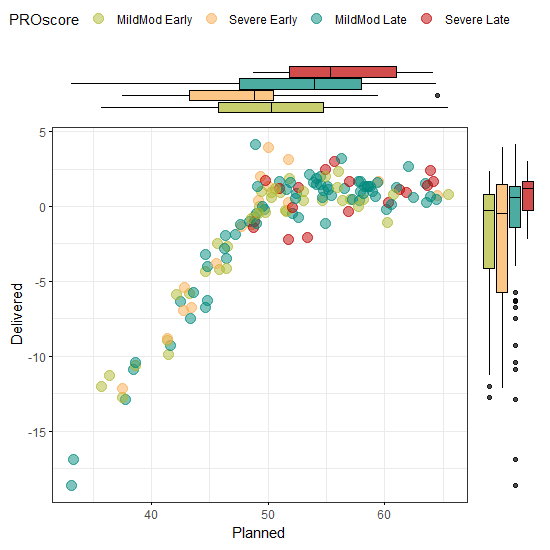
Same items as above but with markers indicating follow-up time, and text indicating the proportion of patients with lasting symptoms in each quadrant (categorical PRO responses, none/mild/moderate/severe to help clarify the data). Interesting that for some structures like PC (right), the majority of patients with long-term severe symptoms are those exceeding the planning objective (e.g., upper right quadrant vs. upper left quadrant)

Attempting to combine some of the data (below) but may make it less accessible/intuitive

* Severe Late (red) -> large circles in the above graphs with QoL scores 7
* Severe Early (orange) -> small circles in the above graphs with QoL scores 7
* MildMod Late (turquoise) -> large circles in the above graphs with QoL scores <7
* MildMod Early (orange) -> small circles in the above graphs with QoL scores <7

I liked these because they were a first sign that lasting severe symptoms were linked with dose. But many other ways to interpret the above scatterplots. If something like this below were to be included, I think they would need a matching scatterplot (above) to be more intuitive and transparent.

Pharyngeal constrictor, left: MDASI swallowing chewing planned (x-axis) vs. delivered (y-axis), right: MDASI swallowing chewing planned dose (x-axis) vs. potentially correctable amount (y-axis). Planned dose is significantly different among groups, violation is not. Violation values above zero are potentially correctable by replanning; amounts express delivered dose relative to the planning objective of 50 Gy, or increases in dose for patients with dose planned above the planning objective.



Eraj

PRO for >65y old

79 patients oropharyngeal carcinoma

SW – subgroup: 65 and older

A previously utilized method of patient grouping was used: symptom free (all ratings 0), no more than mild (all ratings… symptom free (all ratings 0), no more than mild (all ratings <5) no more than moderate (all ratings <7) severe at least 1 >7

Grouped and individual MDASI-HN items were tabulated and the proportions of patients reporting each level of symptom severity were presented graphically as heat maps for the entire cohort, as well as for clinical subgroups of interest, hypothesized to have different levels of symptom severity (tumor subsite, T-category, and receipt of concurrent systemic therapy).

Univariate and multivariate regression analysis were performed with the aggregate MDASI-HN symptom items and with composite of the top 5 symptom items as continuous variables against the following variables: sex, race, cancer subsite, receipt of systemic therapy, RT dose, T stage 1–2 vs. 3–4, N stage 0–1 vs. 2+, unilateral radiotherapy, neck dissection, smoking status, and CCI.

SW – honestly, I’m not finding these heat maps helpful in reviewing. I’ve seen them 3 times now, so we should generate them and might need to include as standard, but I don’t see the point. Especially since you can’t read the axis. I like reporting by severity better. Is there a way to merge severity and heat map?

SW - To explore symptom differentials by age, we compared the proportions of patients reporting moderate-severe level symptoms (≥5), comparing those ≥75 versus < 75 years old, and using this cut-point, there were no statistically significant differences detected.

SW – note write up on cluster

*“Hierarchical cluster analysis results are presented in Fig. 3. Cluster A comprised the majority (64%), with a subset symptom free and the majority with no more than moderate ratings for a limited number of items. The distribution of these items fell into two sub-clusters, one centered around more general, constitutional symptoms of fatigue, memory, drowsiness, and sadness, and another centered around more classical RT-related toxicities, such as choking/coughing, dry mouth, problem with mucus in mouth/throat, difficulty swallowing/chewing, and problem tasting food. Cluster B (33%) patients had a more moderate-severe symptom burden with a heterogeneous distribution of several severely rated items. There again was a sub-cluster centered around moderate global symptoms, yet more broadly spread than cluster A. Similar to cluster A were distinct bands, yet more severely rated, for the same classic RT-related toxicities observed in cluster A. Cluster C formed the small minority (~2%), with essentially severe ratings for the majority of all 22 items.”*

Do univariate and multivariate results.

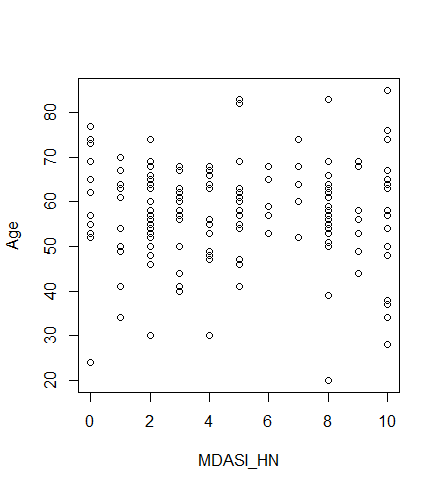
Some data on patient age in our cohort is included below, including sample size of elderly patients (e.g., > 75 years) vs. younger HPV+ patients. Age was not found to be significantly associated with QoL according to Fisher’s exact test.

I find the raw QoL heatmaps hard to interpret and compare as well. Reproduced hierarchical clustering with symptom levels as another option below.

Univariate analyses reproduced for Hutcheson et al. above.. I coded in MANOVA tests earlier, but normality assumption doesn’t hold so results are harder to trust. Multivariate linear /logistic regression is another option…

A screenshot of a cell phone

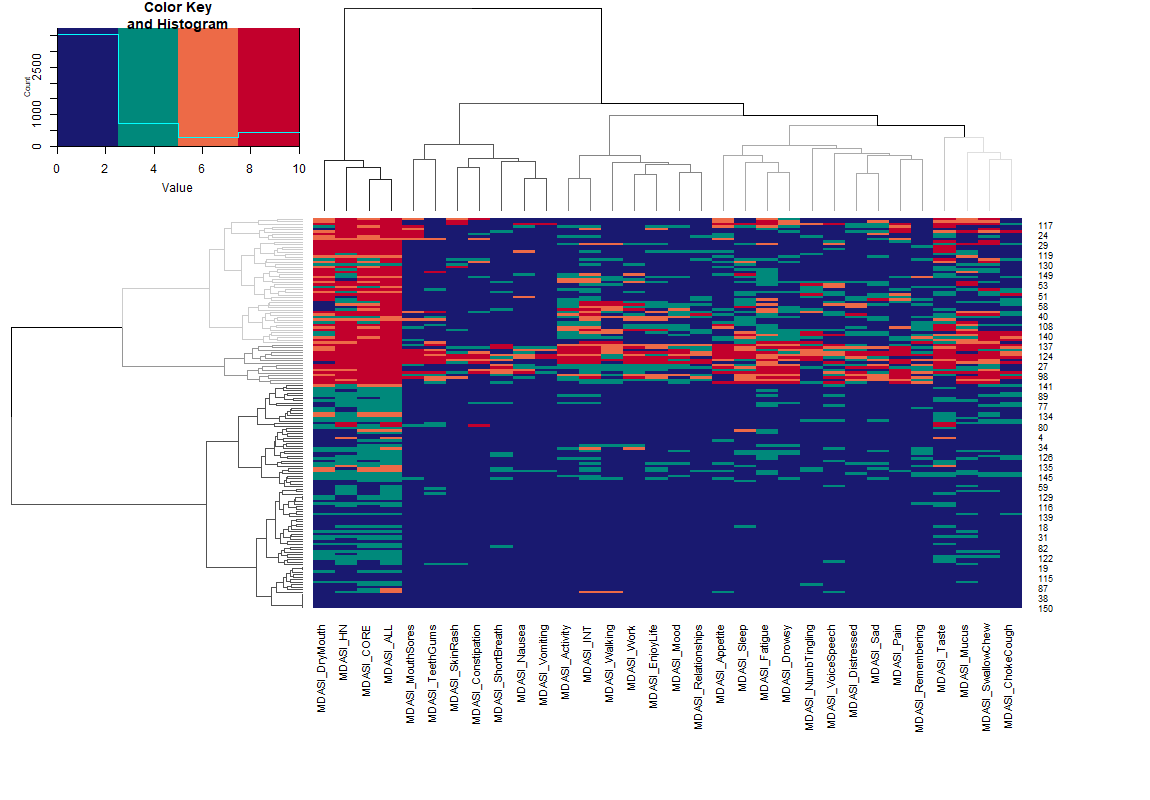
Description automatically generated

Histogram of patient age across the cohort. 6 patients were >75 years; 34 patients were >65 years. Average age was 57.4 years. Surprisingly, average age of HPV+ patients was 58.3 years; generally these patients are younger. There were 100 HPV+ patients.

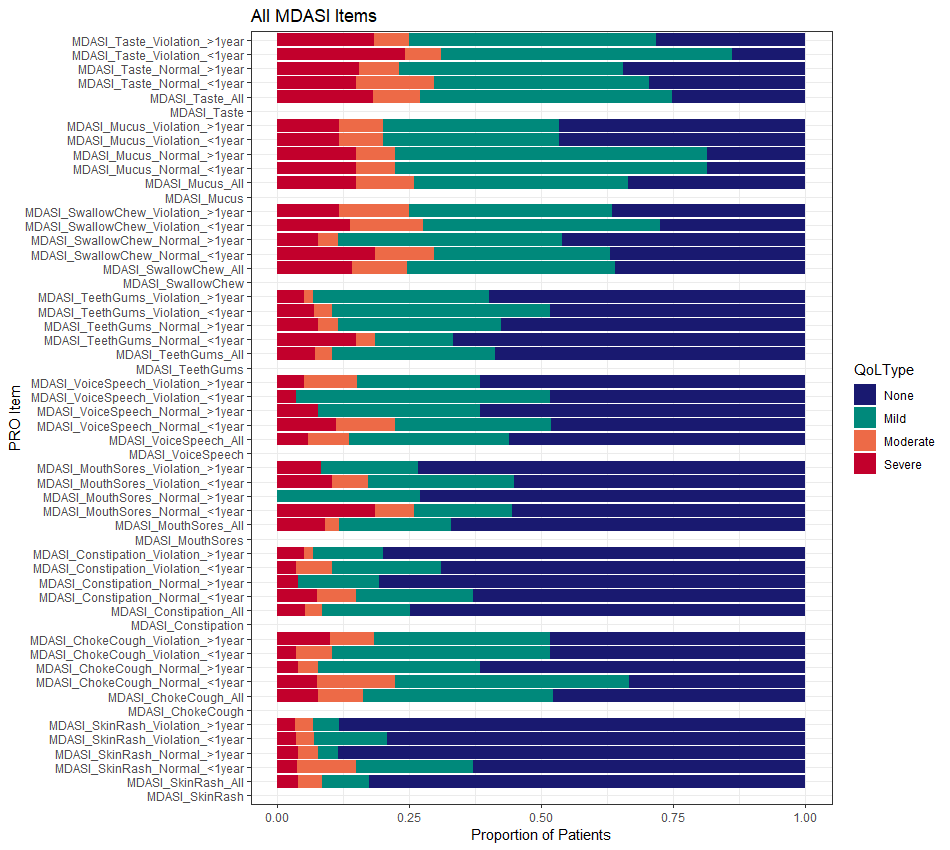
Age didn’t seem to have a strong association with the MDASI subgroups. In fact, earlier statistical tests indicated that PRO results were slightly better in older patients (a very weak trend perhaps seen here).

Text

Heatmaps w/ and w/o subgroup scores saved in Heat Maps folder; some of the saved maps have been extended (850x4000 pixels) to see full patient IDs (n = 155), here trimmed at right



Wong

SW Figure 3 – think about – simpler is better

Delivered pharyngeal constrictor dose, incorporating transient (<1 year) vs. lasting (>1 year) symptoms for the MDASI HN subgroup. E.g., swallowing/chewing and choking/coughing have more severe cases (severe or moderate/severe) for lasting symptoms based on whether pharyngeal constrictor planning objectives were met or not (dose < 50 Gy -> “normal”, dose > 50 Gy -> “violation). 1 year time division consistent with Memtsa 2017 and Petkar 2017.

Gunn

I like write up of results in abstract: Of 139 participants analyzed, 51% had received ipsilateral neck IMRT, and 62% single modality IMRT alone (no systemic therapy). There were no differences in mean individual symptom and interference ratings for those treated with bilateral versus ipsilateral neck IMRT alone. However, 40% of those treated with bilateral versus 25% of those treated with ipsilateral neck RT alone reported moderate-to-severe levels of dry mouth (p=0.03). Fatigue, numbness/ tingling, and constipation were rated more severe for those who had received systemic therapy (p

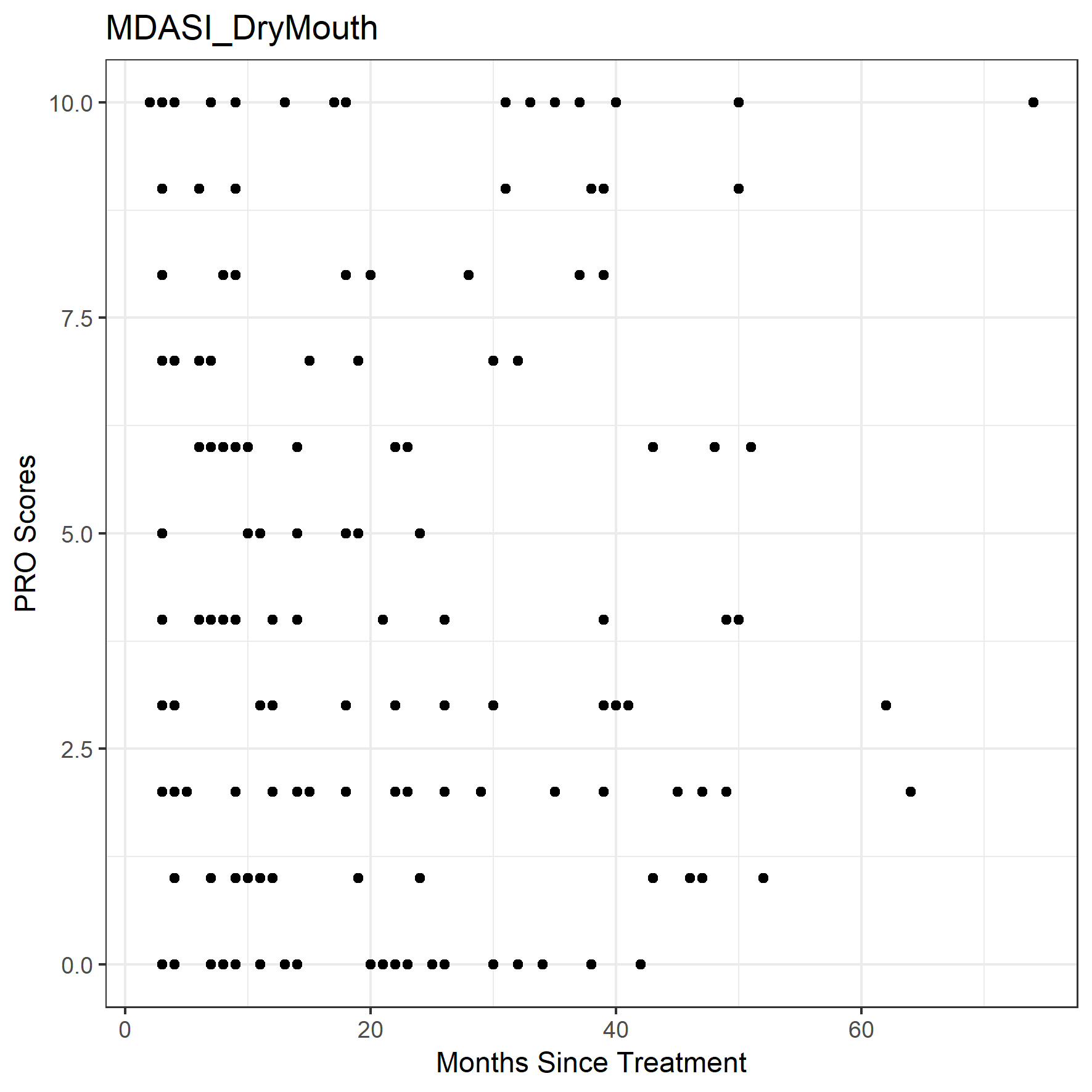
The actual graphs of results in the paper are less illuminating

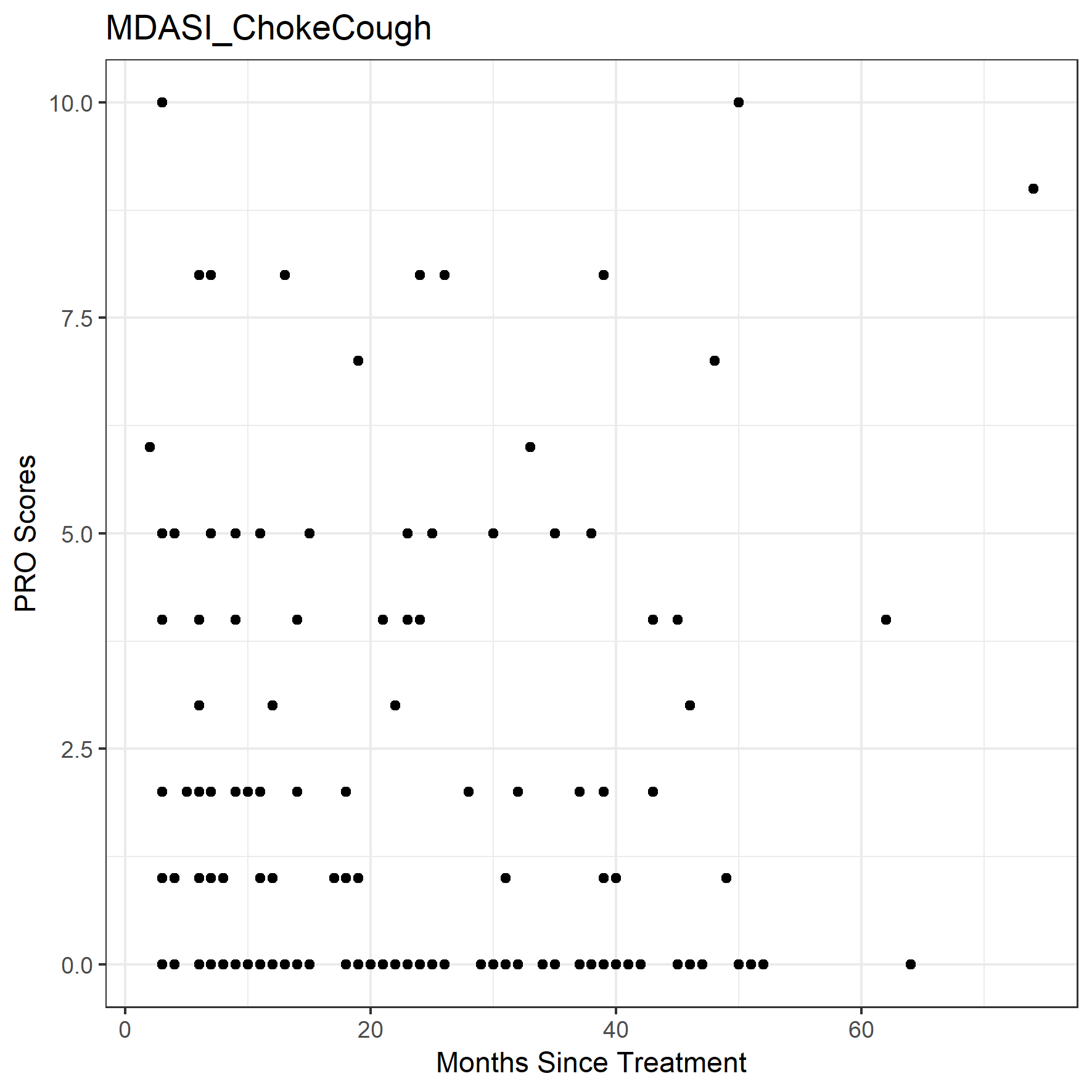
Graphs look more standard to me too, similar to a few reproduced above.

Meng –

Acupuncture to prevent xerostomia

The time-series data is interesting. If the TBCC moves forward with an ART pilot in the future based on our group’s work, it would be interesting to have weekly/routine PROs.





Time vs. PRO scores. E.g., dry mouth responses don’t have a clear time dependence, but item severity seems to fall off with time for choking/coughing, with clearer outliers. Plotted for each item and saved in PRO Results Vs Time folder. Right: Below diagonal has more points than above -> possibly stronger time effects, clearer outliers

Gogineni

retrospective analysis of the efficacy, toxicity, and quality of life (QoL) of patients treated with OARExtreme-sparing stereotactic body radiotherapy (SBRT) in previously-irradiated head and neck cancer

SW – figure 3 and 4 – I find the conclusions on the time course unconvincing when I look at the date for this study, but I felt that the way this data was presented allowed me to fairly evaluate it. They show data points for each patient and link the patients so you can see trends over time.

Looks like quadratic function is being forced but I like the symptom trends/lines. For full transparency of our analysis, I agree it’s important have simple things like this included – maybe scatter plot of dose vs. QoL score (like above with categories/proportions superimposed). To include planned and delivered dose show a trajectory between the two or plot on two axes, may help show that dose generally increases and that proportions of patients with dose > planning obj. and scores > 5 or 7 increase..

Ortigara

MDADI paper ! – report total score and what domain was most affected

They did a sample size calculation. I am super impressed

SW - Table 3 is a good way of collecting and presenting a large number of tests. I think it’s univariate, and you’d need to follow up with multivariate test. Can you do this? I also like table 4 which presents the odds ratio with a 95% CI.

Sample size results are rare! If their calculations scale linearly, then indicates our sample size should be enough to detect a 10 point difference in MDADI scores considered clinically significant on group comparisons (important for Adam and Demetra?). But would likely need to repeat..

Peng

363 patients completed overlapping of 3 questionnaires

SW – missing data? Blank questions? Exclude if > 25% missing

MDADI avg 80.67 15.16 sd

Adam and Demetra – Table 3 present correlation coefficient (strong/moderate/weak) with one of the 3 as the reference (FACT). Also: “From k-means clustering analysis, three separate clusters of patient toxicity outcomes were consistently identified in the vast majority of pairwise and three-way comparisons of the various FACT measures with MDADI and/or SSQ (Table 4, Supplemental figures S1–S48).” I’m not sure this is well described in this paper, because I don’t have the suppliementary figures, but I like the concept and the conclusion (or lack of it) would also be important for us to assess. see also figs 2 and 3.

Almost all patients who completed the consent form completed all 3 surveys. Only patient completed the MDASI but not MDADI or XQ. Some patients declined the initial consent, though, due to language barriers, feeling unwell, or difficulty reading the surveys (forgetting glasses).

Carmignami

60 advanced HNSCC patients. good paper

Divide patients into two groups and consider if there are sig differences between them in total/functional/physical/emotional MDAADI score.

MEMTSA

60 H&N adult patients

associations among QoL, xerostomia and quantity of saliva in a sample of H&N cancer patients who had received conventional radiotherapy

simple graphs

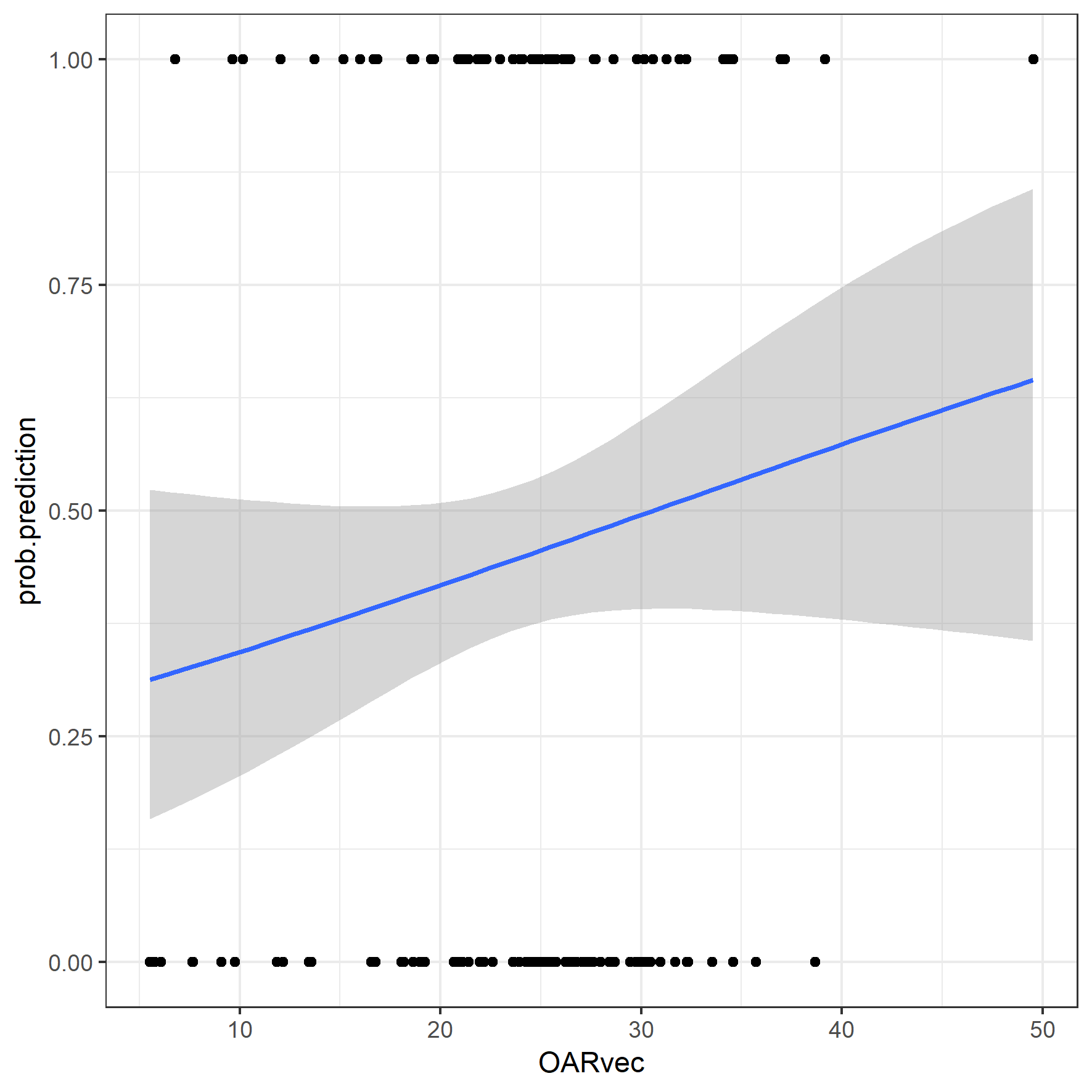
Good reference to clarify some of the variability we see in the data prior to 1 year. Similar graphs to Townes et al., above.

Sapir

We hypothesized that dosimetric parameters to the oral cavity and tongue after organ-sparing IMRT may be correlated with dysgeusia after radiation… assessed the effect of oral cavity doses on patient-reported dysgeusia and explored relationships between patient-reported dysgeusia and salivary flow rates or patient-reported xerostomia, recorded during a prospective longitudinal study of QOL in patients with stage III to IV oropharyngeal cancer receiving IMRT concurrently with chemotherapy

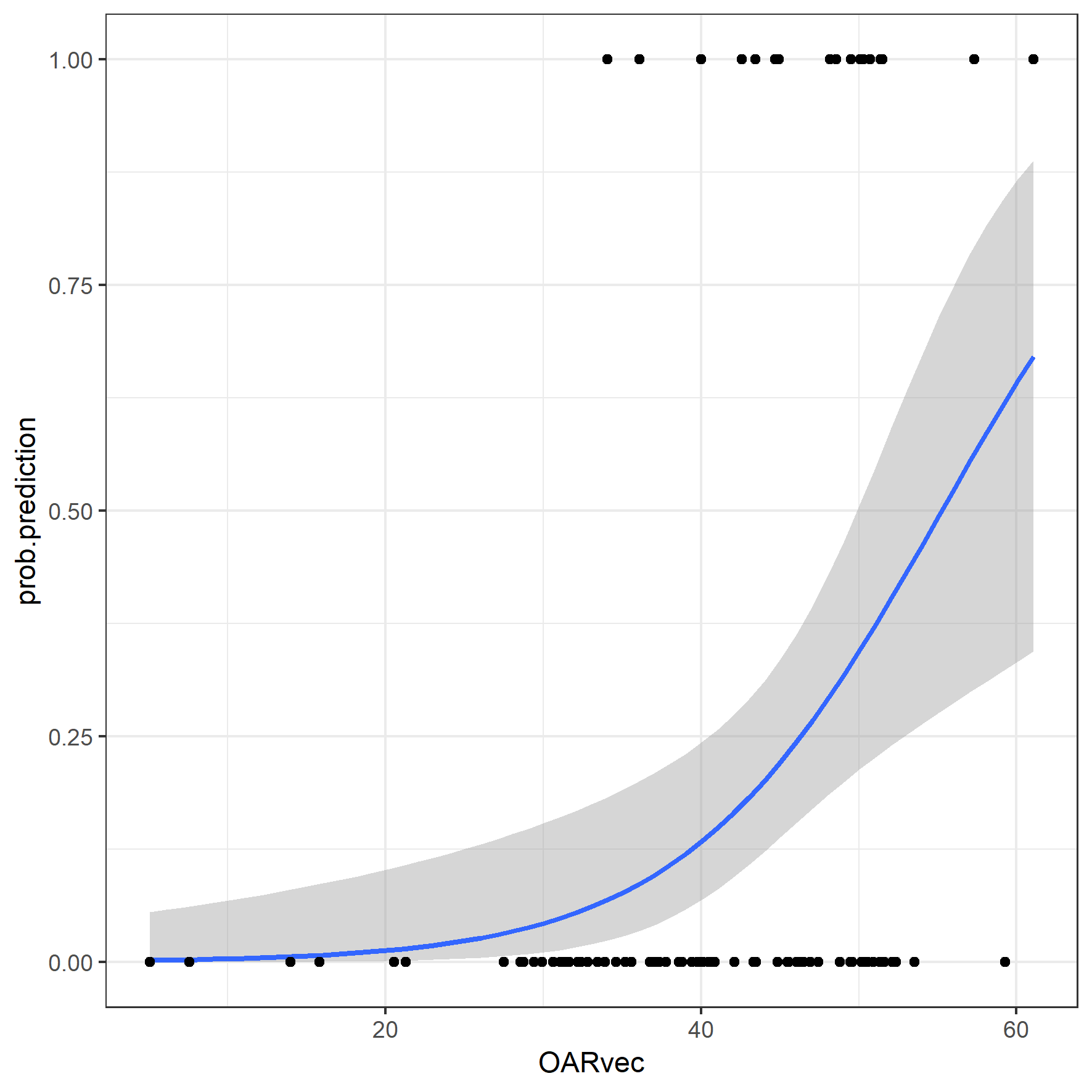
**SW – look at figure 1 – this is pretty cool and I think may be helpful way of framing the question.**

Motivates a look a linking PG dose to dysgeusia. A different reference indicated that the nerves associated with swallowing are also those associated with taste – examine PC dose too. UWQOL and XQ surveys both used, logistic regression models made for each. Focussing only relevant questions. Similar approach to McDowell et al., Hutcheson et al..

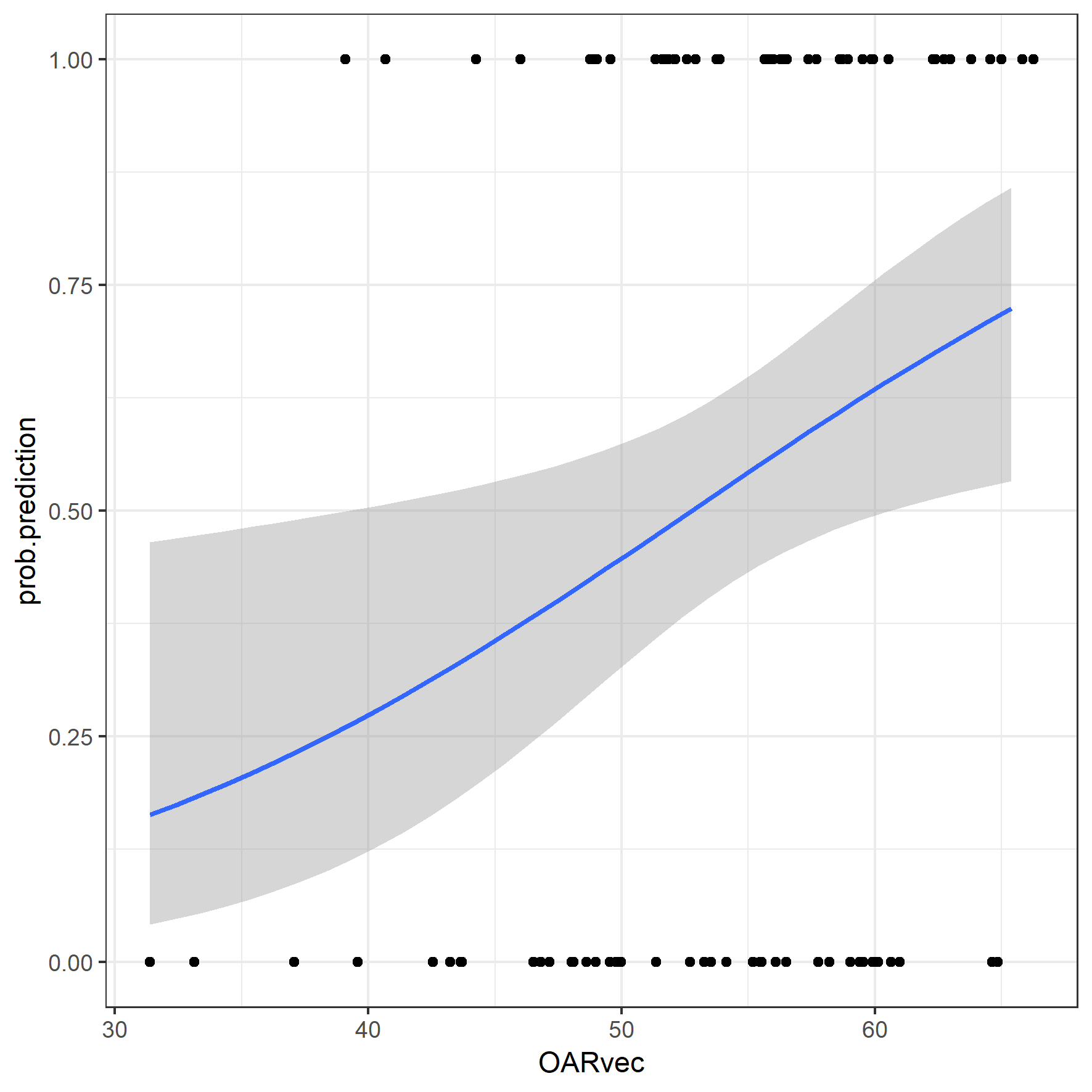
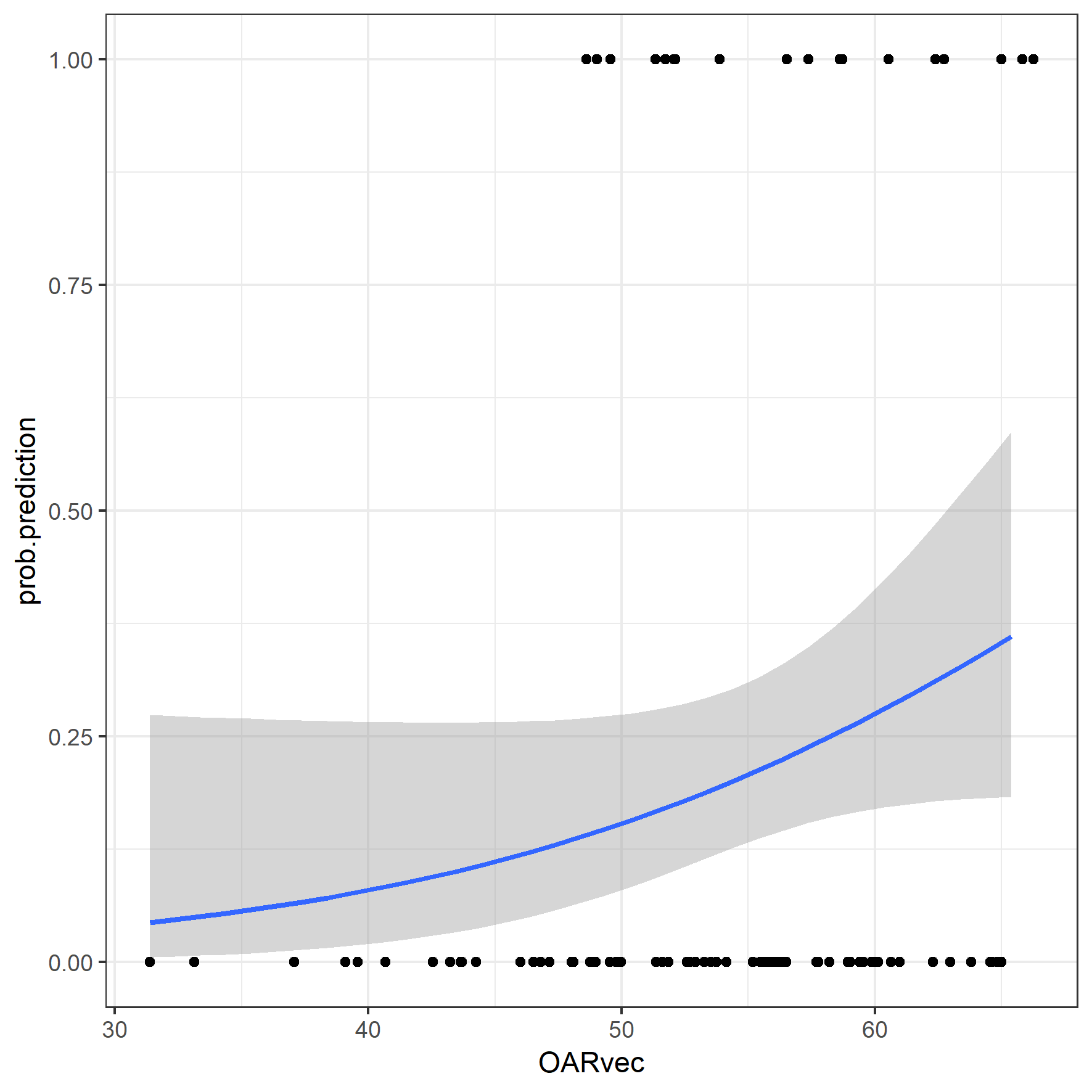


This is the most reasonable MDASI logistic regression plot produced for our parotid gland and pharyngeal constrictor data (here minimum parotid gland dose vs. probability of moderate/severe MDASI dry mouth response), but it looks like the upper right point is saving our bacon on this one and driving the model. Other models on planned, delivered, and violation data are saved to the Logistic Regression folder. Results seem strongly driven by min/max points.

Interestingly, brainstem came through very clear with these models, though.



Above: delivered brainstem dose vs. MDASI fatigue. Also link to numbness/tingling and interference subgroup (data not shown here). I’d expect that changes in brainstem dose are more likely due to setup uncertainties.



Motivation for including the other surveys. Left is a model fit to the MDADI-composite results, right is the single chewing/swallowing MDASI field based on delivered pharyngeal constrictor dose for lasting symptoms.

List domains

total/functional/physical/emotional MDADI

a bunch of MDASI stuff

XQ

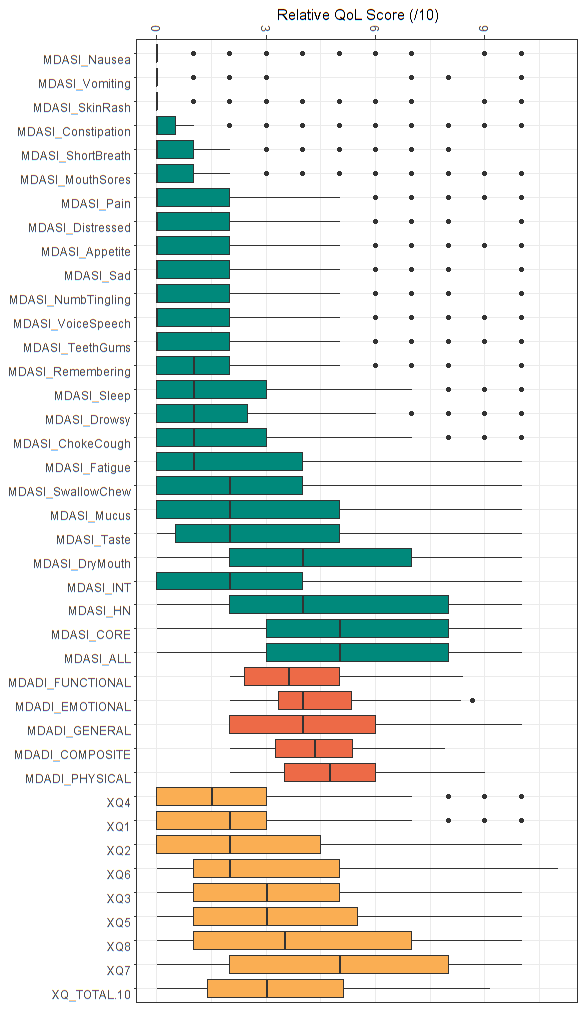
I suggest we initially pick just one to review. E.g. XD. Can repeat the analysis with other surveys, but is too complicated for this purpose to plan to include these all.

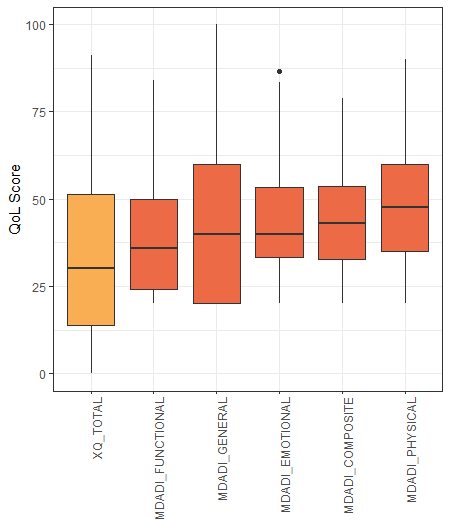
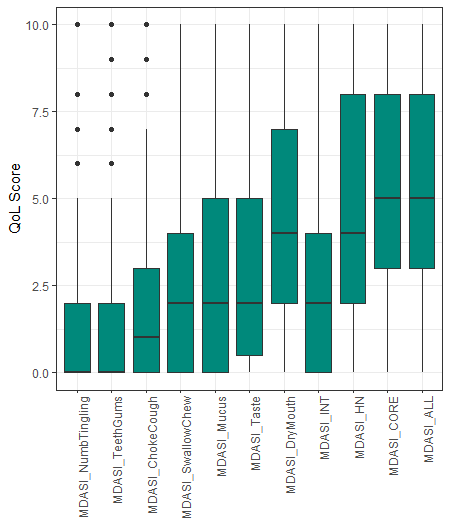
Study Question: Does dose to OARs influence QOL? I like this focus too, from what I’ve seen of the literature, this is very novel, and leads in well to whether ART is worth the resource costs or not.

Figures and Tables to review

Overall –

1. I like both of these, now,





7.1%

7.1%

7.7%

14.1%

14.7%

17.9%

26.9%

12.8%

33.3%

35.3%

42.3%

1. Figure 1 and 2 from Towne for each survey (Reproduced above )
2. Test 1: split patients into severe & moderate vs mild/none on dysphagia and run univariate/multivariate analysis to find out if there are sig differences in patient characteristics and dose PG/PC parameters. (e.g Eraj, Mcdowell Report mean and SD, and estimate difference with CI and p-value in a table.) \*
3. Repeat by splitting by different tests. \*

\*Examined none vs. mild vs. moderate vs. severe; none/mild vs. moderate/severe; and none/mild/moderate vs. severe giving no significant results on univariate analysis. Details on the tests included above. Separating patients by <1 year vs. >1 year consistent with Memtsa 2017 and Petkar 2017 helped to clarify results.

The above tests were repeated using factor-formatted quality of life (“Mild/Moderate” vs. “Severe”), combined with “Early” (12 months post-treatment) and “Late” (12 months post-treatment) designations. Corresponding factor levels included: “Mild/Moderate + Early”, ”Mild/Moderate + Late”, “Severe + Early”, ”Severe + Late”.

**Planned pharyngeal constrictor dose was significantly different among these factor levels for: MDASI core, dry mouth, fatigue, HN, taste, and mucus, swallowing/chewing and interference (p < 0.05 for each).**

**Delivered pharyngeal constrictor dose was significantly different among these factor levels for: MDASI core, dry mouth, fatigue, HN, taste, and mucus items, swallowing/chewing and interference items (p < 0.03 for each).**

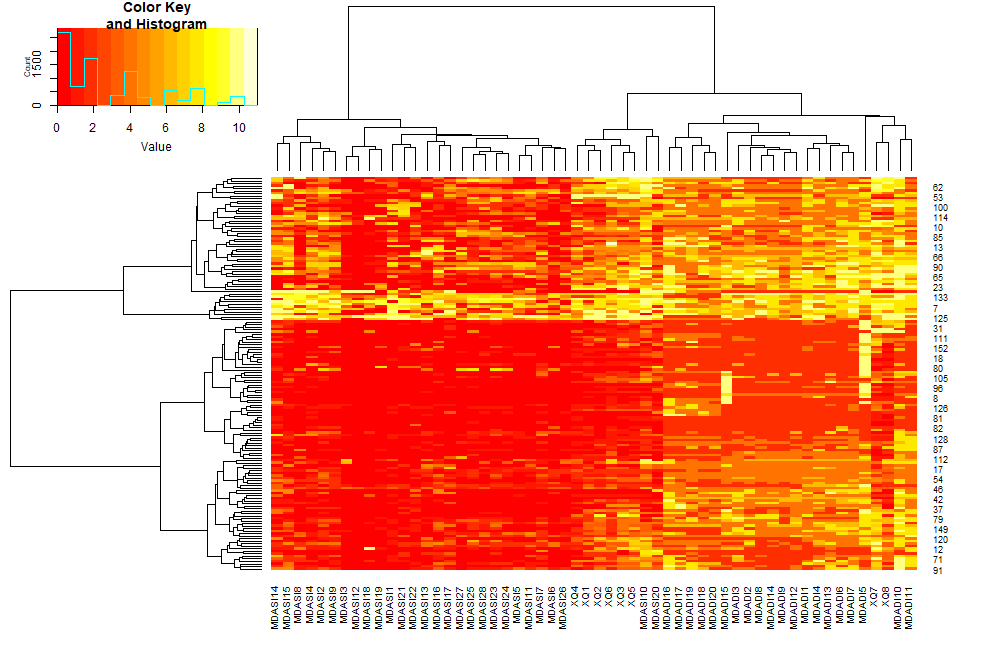
Statistical significance did not persist for pharyngeal constrictor dose formatted as a factor variable; factor levels indicated whether delivered dose met planning objectives (“normal”) or exceeded planning objectives (“violation”).

Potentially-correctable pharyngeal constrictor dose increases were associated with MDASI core, swallowing/chewing, choking/coughing items, but statistical significance did not persist after multiple testing corrections.

1. Plot QOL versus relevant doses to PG, repeat for PC, (Scatter plot with regression line)

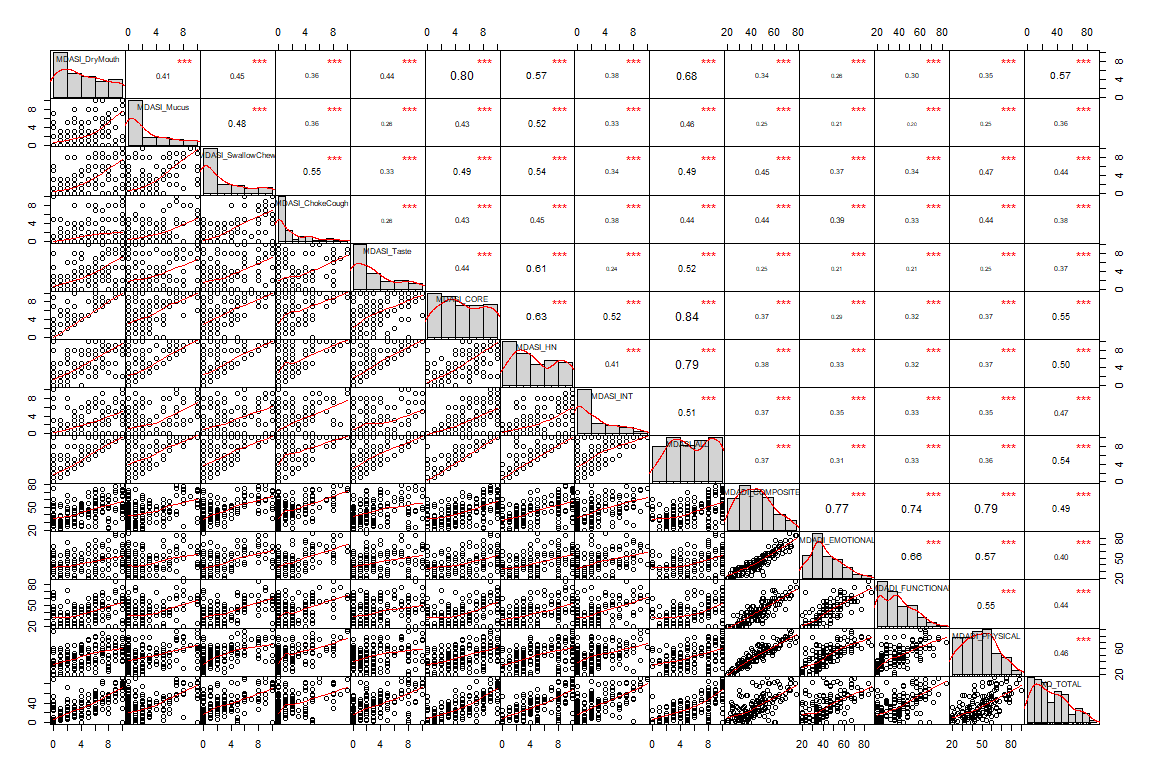
The MDASI data is quite noisy. Looking at the data I find it hard to trust the regression lines. These graphs were reproduced using categorical interpretations (e.g., none/mild/moderate/severe) and meeting/exceeding planning objectives. Categories + time seem to be key for finding any significance and filtering through the noise. Logistic regression looks a bit better..

1. Response heat map (bleh) I’ve left this one for now. I find these are hard to interpret and compare.. I like the hierarchical clustering heatmaps much better too (with an updated one included above in the Eraj section)
2. Cluster – I like this and think it is in line with the literature. Can you compare the patient characteristics between the clusters (may need to exclude B) in terms of doses?



Compared the simple clusters for the blue/grenn/orange/red heatmap reproduced above (Eraj section), but nothing statistically significantly different among groups with respect to individual dose and clinical parameter comparisons. Results may be clarified with <1 year >1 year but reduces cohort size. I think the symptom clusters above are believable, and would fit be an interesting result for a non-dose-focussed paper.

I think this one is going to go into Adam and Demetra’s work



Hopefully will be of some help to their work, even for hypothesis generation or to check their data. I find it helps to see everything at once (R correlation chart with MS word shapes overlaid)..