An FMEA evaluation of intensity modulated radiation therapy dose delivery failures at tolerance criteria levels

Jacqueline Tonigan Faught^{a)}

*Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
IROC Houston QA Center, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
Department of Radiation Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA*

Peter A. Balter and Jennifer L. Johnson

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Stephen F. Kry

IROC Houston QA Center, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Laurence E. Court

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Francesco C. Stingo

*Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
Department of Statistics, Computing, Applications, University of Florence, 50121 Florence, Italy*

David S. Followill

IROC Houston QA Center, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

(Received 16 March 2017; revised 25 July 2017; accepted for publication 6 August 2017; published 19 October 2017)

Purpose: The objective of this work was to assess both the perception of failure modes in Intensity Modulated Radiation Therapy (IMRT) when the linac is operated at the edge of tolerances given in AAPM TG-40 (Kutcher et al.) and TG-142 (Klein et al.) as well as the application of FMEA to this specific section of the IMRT process.

Methods: An online survey was distributed to approximately 2000 physicists worldwide that participate in quality services provided by the Imaging and Radiation Oncology Core - Houston (IROC-H). The survey briefly described eleven different failure modes covered by basic quality assurance in step-and-shoot IMRT at or near TG-40 (Kutcher et al.) and TG-142 (Klein et al.) tolerance criteria levels. Respondents were asked to estimate the worst case scenario percent dose error that could be caused by each of these failure modes in a head and neck patient as well as the FMEA scores: Occurrence, Detectability, and Severity. Risk probability number (RPN) scores were calculated as the product of these scores. Demographic data were also collected.

Results: A total of 181 individual and three group responses were submitted. 84% were from North America. Most (76%) individual respondents performed at least 80% clinical work and 92% were nationally certified. Respondent medical physics experience ranged from 2.5 to 45 yr (average 18 yr). A total of 52% of individual respondents were at least somewhat familiar with FMEA, while 17% were not familiar. Several IMRT techniques, treatment planning systems, and linear accelerator manufacturers were represented. All failure modes received widely varying scores ranging from 1 to 10 for occurrence, at least 1–9 for detectability, and at least 1–7 for severity. Ranking failure modes by RPN scores also resulted in large variability, with each failure mode being ranked both most risky (1st) and least risky (11th) by different respondents. On average MLC modeling had the highest RPN scores. Individual estimated percent dose errors and severity scores positively correlated ($P < 0.01$) for each FM as expected. No universal correlations were found between the demographic information collected and scoring, percent dose errors or ranking.

Conclusions: Failure modes investigated overall were evaluated as low to medium risk, with average RPNs less than 110. The ranking of 11 failure modes was not agreed upon by the community. Large variability in FMEA scoring may be caused by individual interpretation and/or experience, reflecting the subjective nature of the FMEA tool. © 2017 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.12551>]

Key words: FMEA, IMRT, quality assurance, radiation therapy, risk

1. INTRODUCTION

Prospective risk mitigation techniques borrowed from other industries have become increasingly popular for radiation oncology processes in recent years. Failure Modes and Effects Analysis (FMEA) is currently the most prominent of these techniques and is a primary focus in the recently published AAPM TG-100 report^{1–6}. FMEA involves detailed process mapping, potential failure mode identification and failure mode ranking by means of the risk probability number (RPN). The RPN is equal to the product of three scores assigned for each failure mode by a multidisciplinary team of individuals experienced in the specific process being analyzed. These scores are: the likelihood of occurrence (O), severity (S) and lack of detectability (D). Each of these scores has a predetermined scale, defined by the team of experts, generally with values increasing in risk from 1 to 10. The most risky failure modes, commonly identified as those with an RPN above an arbitrary threshold, also defined by the institution's team of experts, are then examined and process improvements implemented to reduce these risks. This threshold is commonly based on other industrial applications of FMEA, however, the AAPM TG-100 has recommended investigation into appropriate threshold levels.¹ In addition, it is common to examine failure modes with high severity scores (≥ 8) because the perceived patient safety concerns are unacceptably large.

The widespread use of FMEA and related tools was in part initiated as a response to several catastrophic radiotherapy errors made public in the United States.^{7–10} One of the primary goals of FMEA implementation in the medical field is to better account for large scale risks and to increase process-wide comprehension. In addition to these goals, application of FMEA specifically to radiotherapy seeks to address the need for and number of ever-increasing quality assurance measures required for the advanced technologies in use in today's radiotherapy clinics. As radiotherapy treatments and associated procedures including imaging become more complex, the demand on busy clinics is only increasing and the efficiency and effectiveness of quality management programs become critical components to any practice. Basic physics quality practices have been called into question and the continuation of current practices with modern technologies needs to be evaluated.¹¹ Overall, FMEA aims to achieve a better understanding of the process and helps put into place more effective quality management practices at the points that matter the most to the overall objective of the process.

While the potential benefits of FMEA are obvious, a few limitations and weaknesses must be addressed before its widespread acceptance and use by individual clinics. Primarily, an FMEA uses subjective, ordinal scoring to obtain risk information. The subjective nature of these scores leads to questionable reliability and validity of the results, as well as variability in the scores.^{12,13} The assigned scores are often biased and can be highly variable between experts and teams of experts, with separate groups shown to identify different failure modes and RPNs for the same process in other

healthcare fields.¹⁴ In addition, in many FMEA studies, physics components are grouped together, leaving out valuable process detail information important to physics quality management. Finally, although the overall scoring in FMEA is subjective in nature, it is common for both Occurrence and Detectability scoring scales to contain quantitative information to guide the scoring process. The quantitative components of the Severity scoring scale, however, require a degree of clinical judgement. While the severity of a radiotherapy error consequence is a complex concept, more directly quantitative guidance could be a valuable tool for better understanding clinical impact on patients and the relationship to discrepancies in physics QA measurements, particularly for physics-specific failure modes.

In this study, we sought to understand the aforementioned limitations of FMEA in radiotherapy and expand its application to those physics-specific processes of IMRT dose delivery called into question, which have not been directly addressed for this process to date. To do this, a survey version of FMEA was performed of basic physics failure modes for the intensity modulated radiotherapy (IMRT) process. Responses to the survey serve to evaluate the range of possible responses in a radiotherapy application as well as the potential variability in interpretation. Overall, knowledge of the risk presented and the error magnitude in dose delivery from key failure modes is required to provide better guidance so that FMEA can be used to establish a comprehensive quality management program for radiotherapy processes.

2. MATERIALS AND METHODS

The survey for this study was distributed via email to approximately 2000 medical physicists worldwide that participate in quality services provided by the Imaging and Radiation Oncology Core - Houston QA Center (IROC-H). The survey was accessed online and was completed anonymously. A panel of medical physicists identified 11 potential critical IMRT failure modes. The eleven different identified step-and-shoot IMRT dose delivery failure modes at or just outside commonly accepted tolerance criteria levels, specified in the AAPM's TG-40¹⁵ and TG-142¹⁶ reports, are listed in Table I. Based on their current clinical practices, respondents were asked to estimate the worst case scenario percent dose error in a step-and-shoot IMRT treatment of a head and neck patient for each of the failure modes keeping in mind both target structures and organs at risk. Respondents were also asked to assign the three FMEA scores (O, D, and S) for each failure mode. The scales used in this survey for FMEA scores were based on conventional scales used in other radiotherapy FMEA studies¹ and were color coordinated for ease of use. The scoring scales are shown in Table II. As described in the introduction, the severity score scales require input from medical clinicians. The following respondent demographic data were also collected: (a) current IMRT clinical practices used including linear accelerator manufacturer, (b) treatment planning system, and delivery technique, (c) years of experience in medical physics, (d) percent of time dedicated to

TABLE I. Physics-specific failure modes and magnitude of failure evaluated in survey.

Failure mode	Magnitude of failure
1. Beam energy	1%
2. Beam symmetry	2%
3. MLC systematically in one bank	2 mm
4. Gantry angle systematically	2.0°
5. Collimator angle systematically	2.0°
6. Couch angle systematically	2.0°
7. MU linearity for < 5 MU systematically	6%
8. MLC transmission and leakage modeling	0.5%
9. MLC tongue-and-groove modeling	0.5%
10. MLC leaf end modeling	0.5%
11. CT number to electron density table systematically	2%

clinical work, (e) familiarity with FMEA, (f) clinic location, and (g) board certification. FMEA was not mentioned in the survey before the demographic questions to avoid intimidating those not familiar with the FMEA risk mitigation technique.

In addition, in the literature, both consensus and averages have been used to evaluate FMEA scores, so both methodologies were examined in this study by also collecting responses from groups of medical physicists.¹⁷ Ten clinics were emailed the survey with the request to complete it as a group. In addition to those questions included on the individual survey, groups were asked to assign quantitative meaning to the severity scoring scale before beginning the scoring process.

TABLE II. FMEA scoring scales used in survey.

Rank	Occurrence (O)		Detectability (D)		Severity (S)	
	Qualitative	Frequency	Qualitative	Estimated probability of failure going undetected	Qualitative	Categorization
1	Failure unlikely	0.01%	Never undetected	0.01%	No effect	
2		0.02%	Very low likelihood undetected	0.2%	Inconvenience	Inconvenience
3	Relatively few failures	0.05%		0.5%		
4		0.1%	Low likelihood undetected	1%	Minor dosimetric error	Suboptimal plan or treatment
5		< 0.2%		2%	Limited toxicity or tumor underdose	Wrong dose, dose distribution, location or volume
6	Occasional failures	< 0.5%		5%		
7		< 1%	Moderate likelihood undetected	10%	Recordable event, potentially serious toxicity or tumor underdose	
8	Repeated failures	< 2%		15%		
9		< 5%	High likelihood undetected	20%	Reportable event, possible very serious toxicity or tumor underdose	Very wrong dose, dose distribution, location or volume
10	Failures inevitable	> 5%	Always undetected	> 20%	Catastrophic	

Post survey collection analysis included RPN calculation and failure mode ranking according to RPN as well as by each of the individual scores; O, D, and S. Ranking was done to assess the priority assigned to each failure mode in addition to magnitude of responses. In addition to visual inspection and direct comparisons, data were coded and statistically analyzed using IBM SPSS. To investigate the relationship between percent dose error and severity scores, which in theory should be directly related, we used Spearman Rank Order Correlation (Spearman's rho) to look for a correlation between the two values over all the failure modes ($P < 0.05$). The Chi Squared Test of Association was used to identify significant relationships between scoring and categorical demographic data (treatment planning system, linear accelerator manufacturer, familiarity with FMEA, IMRT technique, continent, and certification) with Bonferroni adjustments applied to the alpha levels ($P < 0.001$). Effect size is reported with Cramer's V. Spearman's rho was used to investigate the relationship between scoring and percent of time dedicated to clinical work or years of experience. The Chi Squared Test of Association was used to look for differences between group and individual scoring. Extreme outliers (as conventionally defined on Tukey box-and-whisker plots, greater than three times the interquartile range from the bottom of the first quartile or top of the third) were excluded from statistical analysis.

3. RESULTS

A total of 184 complete responses (9.2% response rate) were received for the individual survey. Three responses were

discounted because respondent comments displayed lack of comprehension or lack of following instructions. Three medical physics groups also completed the group survey.

3.A. Demographics

The respondent demographic data collected is shown in Figs. 1–3. Fourteen countries on four continents were represented in the responses, with 84% of all responses coming from North America and 95% of those from the United States. Most individual respondents dedicated at least 80% of their time to clinical work (76%), with a large portion dedicated entirely to the clinic (36%). Almost all respondents were certified by a national organization (92%), with the majority having American Board of Radiology certification (72%). Respondent experience in medical physics ranged from 2.5 to 45 yr, with an average of 18 yr. Just over half of the individual respondents were at least somewhat familiar with FMEA (52%), while 17% reported no familiarity. Several IMRT techniques, linear accelerators, and treatment planning systems were listed as those primarily used for head and neck IMRT treatments at respondent clinics as shown in Fig. 3. Current practice of step-and-shoot and sliding window IMRT techniques were almost equally represented, with 41% and 47% of responses, respectively. Most respondents primarily used Varian linear accelerators (73%) and the most popular treatment planning system amongst respondents was Varian Eclipse (54%). The most represented combination amongst respondents was sliding window using Eclipse treatment planning on Varian linear accelerators (43%).

3.B. Scores and ranking

The values and variability in O, D, and S scores assigned for each failure mode and resultant RPNs are presented as box plots in Figs. 4 and 5, respectively. In these figures, it can be seen that all failure modes were scored in the mid- to low-risk ranges with median O, D, and S score values between 2 and 6 and median RPN values between 15 and 85. Interestingly, each of the eleven failure modes received scores ranging from 1 to 10 for occurrence, at least 1–9 for detectability, and at least 1–7 for severity. The largest mean

severity score was 5 for failure mode 3 (MLC position shifted systematically in one bank by 2 mm), which also had the largest range in values excluding outliers. A large spread was also seen in the RPN scores, with the smallest spread in scores by different respondents for any one failure mode being 251 for failure mode 4 (gantry angle offset of 2 degrees). These large spread in the data were indicative of the wide variation and subjectivity of assigning FMEA scores for the treatment technique identified in the survey. However, despite the large variability between physicists in the scoring of a given failure mode, the median scores showed relatively little variability across different failure modes – the median severity score for all 11 failure modes only varied between 3 and 5.

The ranking of failure modes according to RPN also resulted in large variability, with each failure mode being ranked both most risky (1st) and least risky (11th) by different respondents. Identical rankings were given by two respondents only once, in which case both respondents assigned scores of “1” across the board for O, D, and S, giving all failure modes equal ranking. The distribution of respondent RPN rankings is demonstrated in Fig. 6. Based on the distribution of RPN scores, the respondents indicated more concern for failure modes 7–10 and less concern for failure modes 1, 4, 5, and 6. Failure modes 7–10 represent MU linearity and three MLC modeling parameters, respectively.

In addition to the FMEA scores and rankings, the respondent estimated percent dose error, keeping in mind both the target and OARs, for the worst case scenario of each failure mode was also collected. All failure modes had median estimated errors of 1%–3% with the exception of failure mode 3 (MLC position shifted systematically in one bank by 2 mm) which had a median estimated error of 5%. The lowest mean percent dose errors are shown to be less than 2% for failure modes 1 and 11, corresponding to beam energy and CT number vs. electron density table, respectively. These values are shown in the box plots in Fig. 7. Maximum extreme outliers are not displayed in the figure to preserve visibility of the data, noted by purple crosses. These maximum extreme outliers include 50% error for failure modes 3–6 and 8 assigned by various respondents. Maximum outliers also include

FIG. 1. Demographics of the respondents for (left) years of experience and (right) time dedicated to clinical work. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 2. Demographics of the respondents for their current continent, certification, and familiarity with FMEA. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 3. Distribution of linear accelerator manufacturer, treatment planning system, and IMRT technique listed as primarily used by respondents for head and neck cases. [Color figure can be viewed at wileyonlinelibrary.com]

failure modes 3–6 assigned 100% error by a single respondent and failure mode 7 was assigned the maximum error of 105% by one other respondent. As observed with the FMEA O, D, S, and RPN scores, the variability in the estimated percent dose errors associated with the worst case failure modes was very large. Perceived severity is typically directly associated with estimates of dose error. The variability in this data from our survey shows the need for more accurate and consistent estimates of dose error associated with specific failure modes. Having better quantitative estimates would provide improved guidance for the medical physics community to estimate severity scores with less variability.

3.C. Relationships

As expected, a positive correlation of all respondent data was found between estimated percent dose error and severity scores for each failure mode (data not shown). The strength of the correlation varied between failure modes, with the Spearman correlation coefficients falling between 0.3 and 0.6 and with each result being statistically significant at the

$P < 0.01$ level. This indicates that estimated percent dose error and severity scores monotonically increase together. Another way to assess the relationship between percent dose error and severity scores is to look at the dose errors represented by each severity score, indirectly answering the question: what do the severity scores mean quantitatively? This is shown in Table III. The groups that were surveyed directly assigned quantitative values to the severity scores and these are shown in Table IV. Although the ranges are not entirely consistent, the average percent dose error assigned for each severity score flows nicely from least to greatest with the exception of $S = 10$ in the individual survey responses. This can be attributed to the fact that only one respondent assigned a catastrophic severity score of 10 for any failure mode and therefore the average does not represent the cohort distribution.

None of the various demographic data universally resulted in significant associations (Chi Squared) or correlations (Spearman's Rho) with any of the scores or estimated percent dose errors. Although none of the demographics appeared to significantly relate to any score overall, several statistically

FIG. 4. Occurrence, lack of detectability, and severity scores for $N = 184$ responses (including three groups) for eleven failure modes (as numbered in Table I). Box plots are shown with red representing the second quartile and blue representing the third quartile. Red circles represent the mean score, open circles represent outliers, and stars represent extreme outliers. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 5. Risk Probability Number (RPN) calculated for $N = 184$ responses (including three groups) for eleven failure modes (as numbered in Table I). Box plots are shown with red representing the second quartile and blue representing the third quartile. Red circles represent the mean score, open circles represent outliers, and stars represent extreme outliers. [Color figure can be viewed at wileyonlinelibrary.com]

significant differences and correlations were noted throughout the data. All of the significant results are presented in Table V, according to demographic. The most interesting of these significant results include negative correlations between detectability scores (D) and percent of time dedicated to clinical work for the four failure modes regarding beam modeling (8–11). These correlations indicated that those physicists with more time dedicated to clinical work believe that such modeling errors would be more likely to be detected (lower lack of detectability score). Although these correlations were significant, it is important to note that the strength of the relationship was very weak, with a ρ of ± 1 indicative of a

perfectly monotonous relationship. This was the case for all significant correlations found, with the strongest correlation found being between failure mode 1 (beam energy off by 1%) severity score and respondent years of experience with $\rho = -0.281$. This relationship indicated that respondents with more experience tended to think that the consequences of a 1% beam energy error would be less severe than those with less experience, but again, though this relationship was statistically significant, the relationship was weak. It is also interesting to note that all correlations were negative, indicating those with more years of experience or those with more time dedicated to clinical work generally assign lower scores for the failure modes and scores specified.

4. DISCUSSION

4.A. Limitations and biases

This study provides a look into the variability present in opinions on basic IMRT dose delivery quality assurance failures at or just exceeding the AAPM's recommended tolerance limits. While the expectation for variability in opinions can be extended to FMEA scoring of other radiotherapy processes, the results of this survey are very specific to an IMRT process. The process evaluated by an FMEA must be very specific, which limits this study in two ways. The first of these limitations is that the specific scores obtained were based specifically on step-and-shoot head and neck IMRT. While this may not directly affect the scoring of these failure modes, this specification in the survey and any potential for bias should not be negated. The second limitation is that

FIG. 6. Ranking of failure modes in order of the risk they present using the RPN. The most risky failure mode would have the highest RPN and would be ranked “1”. The size of the bubbles in the chart indicates the frequency at which each failure mode was assigned each rank according to the RPNs calculated. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 7. Estimated percent error assigned for $N = 184$ responses (including three groups) for eleven failure modes (as numbered in Table I). Box plots are shown with red representing the second quartile and blue representing the third quartile. Red circles represent the mean score, open circles represent outliers, and stars represent extreme outliers. Purple crosses indicate that the maximum extreme outliers not shown. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE III. Percent dose errors estimated by survey respondents and corresponding severity scores that were assigned.

Severity score	Estimated percent dose error			
	Min	Max	Average	Median
1	0%	5%	0.62%	0.60%
2	0%	20%	1.08%	0.99%
3	0%	20%	1.74%	1.77%
4	0%	50%	2.77%	2.87%
5	0%	50%	4.00%	4.07%
6	0.01%	105%	6.13%	5.98%
7	0%	100%	12.06%	9.68%
8	1%	50%	16.80%	11.18%
9	3%	100%	20.22%	13.75%
10	15%	15%	15%	15%

although the process was described specifically to the respondents, the interpretation by each individual of the process, the failure modes, and the scoring could be different, and this

TABLE IV. Quantitative values assigned to severity scoring scale by groups ($N = 3$).

Severity score	Quantitative value	
	Min	Max
1	0%	3%
2	0%	4%
3	—	—
4	2%	5%
5	—	—
6	—	—
7	10%	> 20%
8	—	—
9	20%	> 30%
10	50%	> 50%

likely influenced the results. Of course, these same limiting factors can also play a role in a conventional FMEA, but one would expect it to influence the results to a lesser degree with an in-person team working in the same environment.

Another factor that may have added to the variability in the scoring could be that for low level failures such as those near tolerance criteria levels, which are more often what is seen in the clinic, it may be more difficult to estimate the consequences and impact on a complex treatment such as IMRT than large scale, catastrophic failures. This was corroborated by the wide spread in estimated percent dose errors and severity scores from the survey respondents for each failure mode.

4.B. Scoring and ranking

Overall, the eleven failure modes were evaluated to be low to medium risk with average RPNs under the commonly cited industrial threshold value of 125.^{2,4} Since the magnitude of these failures was just outside the currently established action levels, it makes sense that the perception of their risk is not high; otherwise the tolerance levels would be tighter.

TABLE V. Statistically significant relationships between demographics and FMEA scores or estimated percent dose errors for each failure mode. Rho value is given for Spearman's Rho correlations, Cramer's V (V) is given for Chi Squared Test for Association.

Demographic	Response	Statistical relationship
Failure mode 1: energy		
Years of experience	O	$P = 0.008$, rho = -0.194
Years of experience	S	$P = 0.000$, rho = -0.281
Years of experience	RPN	$P = 0.000$, rho = -0.267
Failure mode 2: symmetry		
Years of experience	S	$P = 0.018$, rho = -0.175
Failure mode 3: MLC position		
Linac manufacturer	% error	$P = 0.000$, V = 0.000
Failure mode 4: gantry angle		
Years of experience	S	$P = 0.049$, rho = -0.149
Continent	RPN	$P = 0.000$, V = 0.000
Failure mode 8: MLC tongue-and-groove modeling		
Clinical time	D	$P = 0.011$, rho = -0.187
Failure mode 9: MLC leakage and transmission modeling		
Clinical time	D	$P = 0.024$, rho = -0.166
Linac manufacturer	% error	$P = 0.000$, V = 0.000
Continent	D	$P = 0.000$, V = 0.000
Failure mode 10: MLC leaf end modeling		
Clinical time	D	$P = 0.006$, rho = -0.201
Certification	D	$P = 0.000$, V = 0.000
Failure mode 11: CT table		
Clinical time	D	$P = 0.025$, rho = -0.165

Linac = linear accelerator.

Variability in the responses of the individuals assigning scores was expected to a degree, especially in this situation of a survey because of the large range of professional experiences possible in different clinics and individual career specialization. Differing clinical practices also likely contributed to the variability in scores, particularly for the detectability score which may reflect the confidence respondents have in their QM program or more objectively the experience they have had catching failures. However, the variability demonstrated in these survey responses was extreme. This variability existed in both the scores themselves as well as the resultant ranking of the failure modes. The magnitude of the scores themselves is important, with respect to both predetermined thresholds that may be put in place as well as when assessing quantitative features of the scores and failures. There was no clear ranking of the failure modes agreed upon because of the very large variability in the rankings. This is especially important to note since the overall results of an FMEA rely primarily on the ranking of the failure modes, not just the magnitude of scores assigned. The fact that each failure mode was ranked both most risky and least risky by different respondents emphasizes the fact that the FMEA process is extremely subjective, results are likely institution-specific, and the FMEA results may be of questionable reliability. While specific scoring and rankings could not be agreed upon, general groupings can be made when examining

Figure 6 such as failure modes 8–10 are high risk, and 1, 4, and 5 are low risk. This would indicate that some of the basic periodic linac QA tests (beam energy, collimator angle, gantry angle) are generally perceived to be less critical to quality IMRT than specific beam modeling components. This could potentially be attributed to widely successful QA practices for basic linacs tests, whereas beam modeling is less well understood by the general medical physics community. It is clear, however, from the overwhelming variability in responses, that these ranked groupings would not be the specific conclusion drawn by each independent respondent and as a result, we cannot draw definitive implications for widespread QA practices from this survey. However, these groupings may highlight a potentially beneficial simplification of the FMEA process for application to radiotherapy quality assurance where grouping of failure modes into high/medium/low-risk groups as opposed to specific rankings may be more appropriate and useful. These results additionally demonstrate some of the potential limitations of the results of FMEA used in radiotherapy. FMEA is not an over-arching solution for reducing the amount of QA required, but rather an evolving means of evaluating and optimizing institution-specific processes.

4.C. Relationships

Surprisingly, there were no strong universal correlations between any of the scores, rankings, or dose errors and any of the demographics. This indicates that the demographic information collected was not sufficient to categorize the individual experiences or interpretations, both of which would be expected causes of variability in the scoring. The negative correlations seen between time dedicated to clinical work and detectability for modeling errors (failure modes 8–11) could indicate a difference in understanding of the modeling parameters between those who practice mostly clinically and those who do not, or it could be a reflection of one group having more or less experience with beam modeling. Since the relationships are very weak, these conclusions cannot be reliably drawn. The lack of strong relationships supports the notion that the scoring process is very complex and that individual clinics may need to assess their processes on their own to find the scores that are most appropriate for their specific situation. However, we can note that the FMEA findings between clinics may be very different if the variability observed with our survey is any indication of the opinions of the medical physics community.

5. CONCLUSION

The risk presented by eleven step-and-shoot head and neck IMRT dose delivery failure modes was not agreed upon by the medical physics community. While the common perception of QA tolerance criteria (AAPM TG-40¹⁵ and TG-142¹⁶) level failures in IMRT delivery tends to suggest low risk, the large variability in FMEA scores and estimated percent dose errors found in this study indicate otherwise. The QM

strategies introduced in TG-100, such as FMEA, are an important step in radiation oncology physics to optimize QM efforts and address the constantly evolving technology and treatment techniques introduced in already resource-limited clinics. These strategies represent a different approach to QM in radiation oncology and it will take time to adjust to this new way of thinking and gain the experience to most effectively implement changes. As FMEA becomes more widely implemented, it is important to grasp the potential for variations in results between users and settings and to obtain more quantitative based scoring scales. Large variability in subjective FMEA scoring may be caused by individual interpretation and/or experience, thus reflecting the subjective nature of the FMEA tool. This is of particular importance when FMEA may be used to eliminate routine QA procedures in the interest of time or in the situation of a solo physicist. Recently introduced AAPM Medical Physics Practice Guidelines (MPPG) are aimed at providing guidance on the minimum medical physics support that should be included in various clinical practice settings and can serve as a base level when optimizing a QM program using these relatively new TG-100 strategies.

ACKNOWLEDGMENTS

This work was supported by Public Health Services Grant No. CA 180803.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to report.

^{a)} Author to whom correspondence should be addressed. Electronic mail: jacqueline.faught@stjude.org; Telephone: (901)595-4280.

[This article was corrected after original online publication on 24 October 2017: Corresponding author email address was changed.]

REFERENCES

1. Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. *Med Phys.* 2016;43:4209–4262.
2. Ciocca M, Cantone MC, Veronese I, et al. Application of failure mode and effects analysis to intraoperative radiation therapy using mobile electron linear accelerators. *Int J Radiat Oncol Biol Phys.* 2012;82:e305–e311.
3. Ford EC, Gaudette R, Myers L, et al. Evaluation of safety in a radiation oncology setting using failure mode and effects analysis. *Int J Radiat Oncol Biol Phys.* 2009;74:852–858.
4. Huq MS, Fraass BA, Dunscombe PB, et al. A method for evaluating quality assurance needs in radiation therapy. *Int J Radiat Oncol Biol Phys.* 2008;71:S170–S173.
5. Perks JR, Stanic S, Stern RL, et al. Failure mode and effect analysis for delivery of lung stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys.* 2012;83:1324–1329.
6. Sawant A, Dieterich S, Svatos M, Keall P. Failure mode and effect analysis-based quality assurance for dynamic MLC tracking systems. *Med Phys.* 2010;37:6466–6479.
7. Bogdanich W. Radiation Offers New Cures, and Ways to Do Harm. *The New York Times*; 2010.
8. Bogdanich W. As Technology Surges, Radiation Safeguards Lag. *The New York Times*; 2010.
9. Bogdanich W. A Pinpoint Beam Strays Invisibly, Harming Instead of Healing. *The New York Times*; 2010.
10. Bogdanich W. Radiation Errors Reported in Missouri. *The New York Times*; 2010.
11. Amols HI, Klein E., Orton C. (Moderator). Point/Counterpoint: QA procedures in radiation therapy are outdated and negatively impact the reduction of errors. *Med Phys.* 2011;38:5835–5837.
12. Franklin BD, Shebl NA, Barber N. Failure mode and effects analysis: too little for too much? *BMJ Qual Saf.* 2012;21:607–611.
13. Shebl NA, Franklin BD, Barber N. Failure mode and effects analysis outputs: are they valid?. *BMC Health Serv Res.* 2012;12:150.
14. Shebl NA, Franklin BD, Barber N. Is failure mode and effect analysis reliable?. *J Patient Saf.* 2009;5:86–94.
15. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40.
16. Klein EE, Hanley J, Bayouth J, et al. Task Group 142 report: Quality assurance of medical accelerators.
17. Ashley LaA G. Failure mode and effects analysis: an empirical comparison of failure mode scoring procedures. *J Patient Saf.* 2010;6:210–215.