Reports of Practical Oncology and Radiotherapy 25 (2020) 951–955

Available online at www.sciencedirect.com

Reports of Practical Oncology and Radiotherapy

journal homepage: http://www.elsevier.com/locate/rpor

Original research article

Prospective validation of a core curriculum progress assimilation instrument for radiation oncology residentship

Geovanne Pedro Mauro[a,b,*], Gabriel Faria Najas[a], Heloisa de Andrade Carvalho[a,c], Rosangela Correa Villar[a,d]

[a] Department of Radiology and Oncology – Faculdade de Medicina FMUSP, Universidade de Sao Paulo, São Paulo, Brazil
[b] School of Medicine, Universidade Nove de Julho (UNINOVE), São Paulo, Brazil
[c] Hospital Sirio – Libanês, São Paulo, Brazil
[d] Radiotherapy Department of Boldrini Childrens Center, Sao Paulo, Brazil

ARTICLE INFO

Article history:
Received 22 June 2020
Received in revised form 28 July 2020
Accepted 16 September 2020
Available online 29 September 2020

Keywords:
Graduate medical education
Residency
Radiation oncology
Educational measurement

ABSTRACT

Objectives: To develop a tool that could assess residents' knowledge beyond simple information gathering and evaluate its reliability.
Methods: An assessment tool of 40 objective questions and at least one essay-based question was developed to assess residents' comprehension of general radiation oncology accordingly to validated training curricula beyond level 2 in the Bloom scale. With randomized content, questions were developed such as to be classified as at least 2 in the Bloom scale, so that reasoning and not only information gathering could be assessed. Criteria validation was made using the Classical Test Theory to describe difficulty and discrimination of each item. Reliability was tested by internal consistency using the Cronbach alpha test.
Results: Between 2016 and 2019, 24 residents were assessed. Six different versions of the test were made with a total of 240 objective questions and 18 essay-based questions. Five of the six versions were deemed valid and reliable. Comparisons between 1st (PGY-1) and 3rd (PGY-3) year residents were made. Consistently, PGY-3 residents had scores 150% higher than PGY-1 residents. Only two different PGY-3 reached the most complex level of answers in the essay-based questions. The results demonstrated that the major part of the acquired knowledge and retention occurs in the first six months of training rather than in all the following period.
Conclusion: The instrument can be considered valid. This developed instrument also raised the hypothesis that residents may not assess and analyze their acquired knowledge beyond the application level.

1. Introduction

For the development of any intervention in education and subsequent evaluation of results, a measurement instrument is necessary. Since the work of Tyler (Tyler, 1942),[1] the development of educational sciences has consolidated the evaluation of each intervention as fundamental for its acceptance as scientific evidence, in a reproducible way by future researchers. However, these assessments are often made with non-calibrated, biased instruments, which could produce results that are not only harmful to educational and scientific development, but have consequences that are not assessed by the proposed instrument. Thus, it is necessary to develop a calibrated instrument.

Few countries have a competency-based curriculum developed for residency in radiation oncology. One of the first countries to develop such technology at the national level was the United Kingdom, and a unified European[2] curriculum already exists. Singapore, Australia, and New Zealand[3] have theirs published and validated internationally. In Brazil, the Brazilian Radiotherapy Society (*Sociedade Brasileira de Radioterapia* – SBRT) has developed a document towards this direction, with the first report containing basic recommendations for the construction of a residency program, but not a nationally structured competence matrix. Thus, such a matrix is needed so that efforts can be done in assessing whether it has been properly implanted and how to analyze and improve its results.

In the case of the development of curricula for medical residentship courses, few instruments are validated for the proper

* Corresponding author.
E-mail addresses: geovanne95@gmail.com, geovanne.mauro@usp.br (G.P. Mauro), gabriel.najas@gmail.com (G.F. Najas), heloisa.carvalho@hc.fm.usp.br (H.d.A. Carvalho), villardias@uol.com.br (R.C. Villar).

https://doi.org/10.1016/j.rpor.2020.09.003
1507-1367/

Table 1
Examples of objective multiple-choice question.

Type	Test
Objective multiple-choice	One of the major concerns about long-term treatment of left breast is cardiac toxicity. By the review of Darby et al.,[11] the cardiac toxicity threshold is 30 Gy in the chest. However, the findings of this study are debatable by the retrospective nature of the data employed. Regarding cardiotoxicity of treatment for breast cancer, check the CORRECT answer. a- Currently, all treatment-related cardiotoxicity is due to the drugs used, and there is no role for improving radiotherapy techniques b- The only cardiotoxic drug in the treatment of these patients and which has the greatest chance of causing long-term problems is doxorubicin c- Her-2 positive patients are more likely to have cardiotoxicity d- Hypofractionated treatments are potentially less cardiotoxic because of the total dose delivered e- The maximum dose in the anterior descending coronary artery is the most reliable constraint currently for cardiotoxicity A 75-year-old patient was diagnosed with a left cervical lymph node enlargement 1 month previously. She was biopsied and revealed metastatic squamous cell carcinoma (SCC). Upon staging, the patient underwent laryngoscopy, which biopsied the tonsils, vestibular recesses, nasopharynx and piriform sinuses. However, there was no diagnosis of a primary tumor. Staging PET was performed, without primary lesion location. The patient was then staged as a Tx N2b occult cervical SCC. She was referred by the head and neck surgeon for evaluation for radical treatment. Patient has good performance, ECOG 0, lives alone. There is emotional support from the family, two children dedicated to the mother, but she does not give up living alone. During the patient's staging, you find out that she has three amalgam crowns (restoration made of mercury-based metal alloy, used in the past) on posterior molars, which promote a great strike on the tomography image. The patient's son asks if something similar could happen during the patient's radiation treatment. Which of the answers below best reflects the situation and its physical explanation? a- Yes, because both the exam and the treatment use X-ray b- Yes, because in both situations the atomic number of the dental alloy is equal c- No, because the density of the alloy will be a variable considered in the treatment planning d- No, because the atomic number of the alloy has little effect on radiotherapy e- No, since this area will be treated anyway

evaluation of interventions. As far as the researchers' knowledge is concerned, no quantitative teaching instrument in the form of progress test[4] in a competency-based curriculum has ever been developed for the assessment of residentship in radiation oncology. This work, therefore, aims at the development of this important measurement instrument and the results of its application.

2. Methods

Before the design of the assessment instrument, the objects that were to be evaluated with the proposed instrument were determined. Our institution has a curriculum that fits us well, but it was not built using the methodology and techniques necessary to classify it as a competency-based curriculum. For this, it had to be based on the CanMeds[5] methodology. The levels of proficiency and milestones are not written in a systematic fashion as proposed by the International CBME Collaborators[6] as it is still not internationally validated or published. Thus, a combination of different data to develop the competence core curriculum was used. The basis for it was the Royal Australia and New Zealand College of Radiologists' Radiation Oncology Training Program proposed in 2013. This competence matrix was used because it is very comprehensive and detailed and because it was also externally validated in another country. A few additions were made to update it. For the physics curriculum, the matrix was complemented with the American Society for Radiation Oncology (ASTRO) recommendations.[7] The imaging and radiology core competences were changed to reflect new data.[8] Thus, based on this core curriculum, a specific instrument was created to assess the development of residents during their training.

The basic content validation was made based on Bloom's taxonomy model.[9] This model classifies educational interventions based on a hierarchy of cognitive processes used by the student to solve a given task based upon a taught concept. To assess levels 2 and 3 in the taxonomy (lower levels on the cognitive process, named comprehension and application of given concept), 40 objective multiple choice questions were designed per instrument, all of them with 5 alternatives. All items with level 1 (knowledge) were excluded. Levels 4–6 (analysis to evaluation) were assessed by an essay-based question, also present in every version of the instrument. Table 1 shows examples of both types of questions. Content was based on the core curriculum and randomly chosen by a computer-based randomization program at random.org operated by Randomness and Integrity Services Ltd. at ¨https://www.random.org/company¨to assure that every competence described on the core curriculum could be assessed with equal odds. A different version was made and applied every six months for the development of this instrument. Each version was designed to be administered to all residents in different years of their residentship. The instrument was designed to assess the development and implementation of core curricula, so that all residents were submitted to the same test.

Criteria and construction validations were made based on the Classical Test Theory (CTT). The CTT is a psychometric theory that is the basis for the application of the statistical tests used to validate an assessment instrument in different areas of humanities and sociological sciences, such as psychology and educational sciences. Validation of criteria was reached if the item had positive difficulty with no negative or null values and no null variance within evaluated tercile. Discrimination was assessed by correlating the item with the total of items via t Student test. Reliability was based on internal consistency and was tested with Cronbach's alpha test.

3. Results

In our institution, there could be up to six residents of each residentship year (in Brazil, Radiation Oncology programs took three years until 2019). From 2016–2019, a total of 24 residents were assessed in this study. A different version of the test was applied every six months from June 2016 to July 2019. Every version of the instrument had 40 multiple choice objective questions (total 240 questions in the period) and 1–4 essay-based questions (total 18 in the period). The latter 2 versions had only one essay-based question so the instrument would not be very time consuming. This decision was based on the data for the first four test versions, where the variation among scores of each resident was limited among the various essay-based questions.

Except for version #2, every version was considered valid after validation of criteria and construction for the multiple-choice questions. Version #2 had very few items validated, with some items with statement design problems and recurrent negative discrimination values and was considered difficult by residents. The criteria validation results for the validated items in every version of the test can be seen on Table 2 as the results on the reliability tests. Item Response Theory (IRT) was not used for validation due to the limited amount of radiation oncology residents each year.

Once again, every version of the test was considered reliable, except version #2, due to its validation results. The comparison

Table 2
Validation and reliability data (objective multiple-choice questions).

Parameters	Test						
	#1	#2	#3	#4	#5	#6	Expected values
Median total score	0.36	0.32	0.42	0.4	0.36	0.4	–
Median difficulty of validated items	0.6	0.6	0.35	0.5	0.64	0.6	0.6 was considered a hard test
Median discrimination	0.4	0.27	0.21	0.3	0.16	0.15	*
Median D index	20	−10	33	20	25	20	20**
Number of validated items	15	3	14	22	18	17	–
Internal consistency (alpha Cronbach)	0.816	0.833	0.824	0.846	0.795	0.860	>0.75
Number of difficult validated items	3	0	2	4	9	7	Every test was designed to have 10 difficult items
Subject of most difficult item	General oncology (quality of life)	Head and Neck (larynx) and Pediatrics	Breast and Hematology (Hodgkin's lymphoma)	Hematology (non-Hodgkin lymphoma)	Radiobiology and GI (rectum)	GU (prostate)	

Legend: GI = gastrointestinal; GU = genitourinary * According to CTT, discrimination value should be around the threshold settled by the test developer as optimum. In this case, since it is a sufficiency progress test and median values were reached around the 0.3 threshold and most discrimination values were around the same values, we considered them optimal. No normal values are provided. ** D index should have values around the possibility of an individual to correctly guess the item based on probability only. Since all items are five option multiple-choice questions, D index should be around the 20% possibility value.

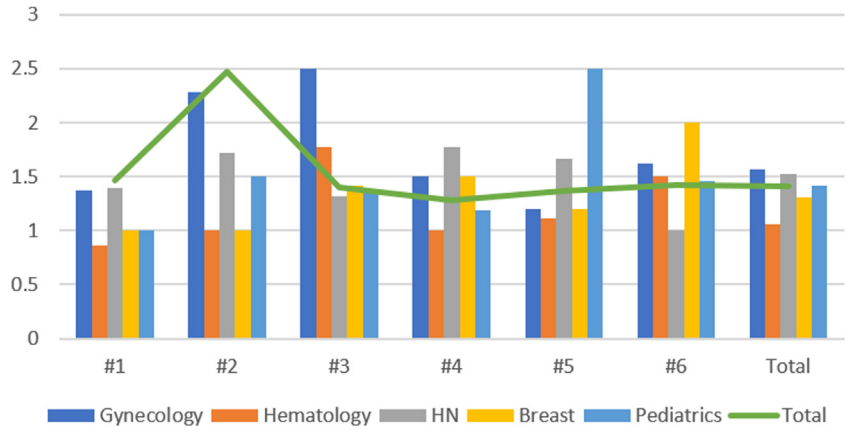

Fig. 1. Variation of the R3/R1 scores by subject at each test version. The subjects selected were the ones with the most aberrant results. Subjects not represented in this figure: physics, radiobiology, general oncology and radiology had results close to breast; gastrointestinal, genitourinary and central nervous system reflected the same curve as head and neck (HN); thorax, skin, soft tissue sarcomas and benign conditions also had their results close to HN, but less reliable due to the limited number of items.

between post-graduation year 1 student (PGY-1) and post-graduation year 3 (PGY-3) scores can be seen in Fig. 1. Consistently, PGY-3 had scores that were 150% higher than PGY-1 residents, except for version #2, when the difficulty was so high that scores for PGY-3 reached 250% of the scores of PGY-1 residents.

More advanced cognitive processes were evaluated via essay-based questions. These were used to assess not only the acquisition of knowledge but also how it could be applied. The variation within items, nevertheless, was very low. For the first four versions of the test, there were 4 essay-based questions. The results were assessed at this point, since those essays were extremely time consuming. No resident had answered any essay-based question to the level 5 (synthesis) and variations within groups and within the whole sample were so low that no statistical test could be applied. After that, the number of essay-based questions was reduced to one per version. In versions #5 and #6, a single resident on each version reached level 5 answers (not the same resident, and both PGY-3 at the time).

Based upon this finding, the test was considered valid and reliable and its results could be used for further assessment.

The difference between first and last year's residents' scores were analyzed. Most subjects kept a difference between PGY-1 and PGY-3 scores between 1 and 2, with PGY-3 scoring among the same and doubled the ones from PGY-1. When scores from PGY-1 were superior to those of PGY-3, the items were reanalyzed. Fig. 1 graphically represents these findings. The subjects selected were the ones with the most divergent results. A few aspects must be considered. Gynecology results were selected because all residents during the study had access to the translated core curriculum on the subject and all residents had standardized discussions on gynecology in the PGY-2. Hematology was selected because it had the most negative results in the comparisons. Breast was selected because most of the curriculum is discussed in the first year of residentship. Physics, radiobiology, general oncology and radiology also had results close to breast and were not represented. Those subjects are also extensively discussed in PGY-1. Pediatrics is only discussed in PGY-3. Subjects such as gastrointestinal (GI), genitourinary (GU) and central nervous system (CNS) reflected the same curve as head and neck (HN), hence the selection. Other subjects such as thorax, skin, soft tissue sarcomas and benign conditions also had their results close to HN, but the limited number of items made their results less reliable.

No differences were observed when comparing PGY-1 with PGY-2, probably due to the low number of subjects. In addition, a few presented exceptional results that would confound the final analysis (Table 3).

Table 3
Example for essay-based question.

Type	Test
Essay-based (statement)	You are at the radiotherapy department's quality control meeting when the following patient is introduced: 47-year-old female patient diagnosed with squamous cell carcinoma (SCC) of the esophagus extending from 17 to 21 cm from the upper dental arch using endoscopy. PET-CT and transesophageal USG staging determined T3 N0 M0. Based on this diagnosis, neoadjuvant concomitant chemotherapy and radiotherapy were used as by the CROSS clinical trial. The contour displayed at the meeting contained elective irradiation of the cervical level IV (supraclavicular fossa - SC), as it is a tumor at the carina level. The summary appendix published with the article defines GTV and PTV as follows: "The Gross Tumor Volume (GTV) is defined by the primary tumor and any enlarged regional lymph nodes, and will be drawn on each relevant CT slice. The GTV will be determined using all available information (physical examination, endoscopy, EUS, CT-thorax / abdomen). The Planning Target Volume (PTV) will provide a proximal and distal margin of 4 cm, in case of tumor extension into the stomach, a distal margin of 3 cm will be chosen. A 1.5 cm radial margin around the GTV will be provided to include the area of subclinical involvement around the GTV and to compensate for tumor motion and set-up variations". On the same occasion, the study's toxicities were also published. The frequency of grade III or higher non-haematological toxicities was up to 13% according to the publication. In 2014, Oppendik et al. published the CROSS study[12] recurrence patterns. The results indicated that the recurrence in the supraclavicular was the same in the surgery only arm (4.2%) and the neoadjuvant arm (4.3%), with a decline in every other site in the neoadjuvant arm except this one. Taking the above data into account, would you advocate elective supraclavicular irradiation in this case? Argue your point of view.
Essay-based (answer key)	Is the argument consistent and cohesive? Items with knowledge level (Bloom 1) Do they identify that SC irradiation is elective due to N0 disease? Does it adequately measure the order of magnitude of SC recurrence? Do they identify pulmonary toxicity as potentially the most critical in the irradiation of SC? Do they adequately measure lung toxicity rates in the study? Do they identify SC as the first drainage station outside the mediastinum for tumors above the carina? Items with comprehension level (Bloom 2) Do they identify that the proposed treatment is NOT a CROSS standard? Do they rate the order of magnitude of SC recurrences as relevant? Do they identify that the relapses in the surgery and neoadjuvant group are statistically similar? Do they rate the order of magnitude of perioperative pulmonary toxicities as relevant (13%)? Items with application level (Bloom 3) Do they properly organize the pros and cons of SC irradiation? Do they organize the toxicities expected with irradiation and judge them as acceptable or not? Do they foresee the surgical procedure and possible complications of SC irradiation in it? Items with analysis level (Bloom 4) Do they contrast the two alternatives in terms of expected toxicities? Do they conclude that the toxicity expected by SC irradiation is higher and stratifies it higher or lower than pulmonary and perioperative toxicity? Items with synthesis level (Bloom 5) Do they adequately synthesize the benefits of the chosen option and judge them as superior to those of the discarded option? Do they adequately synthesize the harm of the chosen option and judge it as inferior to the benefits of the chosen option? Do they adequately synthesize their conclusion based on the patient's real chances, bringing its conclusion closer to the patient in question and avoid generalization (emphasize the importance of the position of the primary tumor, chance of relapse in SC and percentage of lung treated with conformational technique for esophagus)? Items with evaluation level (Bloom 6) Do they adequately assess their chosen alternative as having the best cost / benefit ratio for the described patient at their conclusion? If opted for irradiation, does the resident correctly assess that the patient's gain is based on studies other than CROSS, that this topic was not addressed in CROSS and that the SC recurrence in this study is concerning? If opted for non-irradiation, does the resident correctly assess that the best randomized evidence does not predict SC irradiation, which prevents toxicities at the expense of an acceptable locoregional failure?

4. Discussion

To our knowledge, this is the first report on the evaluation of radiation oncology residency using a quantitative teaching instrument based on Bloom scale. The developed instrument reflects a process evaluation method as it is a progress test, not a sufficiency method. Thus, the residents' grades were not the endpoint of the evaluation. The studied cohort were training physicians from a public university hospital which is a reference center in the country. Basic and advanced technologies have been available in the facilities with a well-established radiotherapy training program for more than 50 years. The program was adapted and modified along the years, but since now, no specific quantitative evaluation has been performed in a pre-designed prospective study. Currently, there are six residents / year, in a three-year program when the study started (in 2020, a four-year program was implemented), performing 24 studied subjects.

Concerns about the ideal sample size were taken into account. To validate a test formed by 40 objective questions and an essay with 20 items, the estimated sample size would be of approximately 360 residents per version. Considering our entire country, there are 70 vacancies for radiation oncology residency / year distributed in 36 programs. Not all vacancies are opened for training or fulfilled every year. Even in similar or different environments worldwide, we believe that this ideal sample size would not be achievable. Statistically, we could validate both construction and reliability of the applied tests, but our results should be seen as hypothesis generators rather than concrete results. Nevertheless, we consider this an important information that may be used to improve the training approach and quality of the program.

Therefore, the following statements should be analyzed with these premises in mind. According to Fig. 1, the overall difference in knowledge between PGY1 and PGY3 was only about 50%. Residents tend to learn more and retain knowledge in their first six months of training rather than in all the following period. It seems that the first months have a greater impact and it is when they absorb and memorize data. Since then, the encouragement to process the content is small and they do not evolve in the same intensity. Radiation oncology is a specialty that is seldom approached in medical graduation curricula around the world.[10] Thus, residents start training with little to no previous knowledge on the subject. The first six months of training are crucial. It is the period when a large part of clinical reasoning is built and the low results on the essay-style questions show that during training little impact is made on residents concerning analysis and assessment of the cognitive content gathered. Resident training should be more than simple information acquirement. Since no items on the tests were made that would assess Bloom level 1 knowledge, those results do not show how much residents know but to what extent they can apply, analyze, and assess their current knowledge. Hence, reasoning should be trained, and efforts should be directed to make residents not only recognize and follow clinical protocols and guidelines, but also to understand and confront with scientific thinking the protocols and their application in daily practice.

To solve this matter, stimulated clinical research and project design could be helpful. Research projects give residents an opportunity to assess and critique the protocols they are called to

recognize and follow, and use this knowledge in the decision-making process of every day's challenges. However, care should be taken not to engage the resident in very complicated and time-consuming projects. This would turn the resident into a very good specialist in that field but still no clinical reasoning on other subjects is acquired. Also, research can consume an important part of resident's time that should be dedicated to patients and clinical practice. So, the ideal balance of the workload between practice and research is yet to be defined.

Resident training is not assessed enough or properly. Assessment is generally made on the knowledge and simple application levels of the Bloom scale. New approaches on stimulating resident knowledge on radiation oncology beyond the levels of simple application are warranted.

5. Conclusion

This developed instrument raised the hypothesis of a possible lack of application and analysis by residents of the acquired knowledge on clinical application. More input should be made on assessment tools that look beyond the clinical expert role and into the roles that require radiation oncologists to rethink and develop their knowledge. The first months of training are crucial for learning and retaining knowledge.

Authors contributions

Geovanne Mauro were responsible for study design and ethics committee approval. Geovanne Mauro were responsible for data collection. Geovanne Mauro were responsible for statistics analysis. Geovanne Mauro and Gabriel Najas were responsible for writing manuscript. Heloísa Carvalho and Rosangela Villar were responsible for overall orientation and manuscript review.

Declarations

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics approval: Ethics committee authorization was obtained in the local ethics committee according to Brazilian law and the Declaration of Helsinki.

Consent to participate: A waiver was granted upon ethics submission to request each patient consent due to it's a retrospective study by the ethics committee.

Availability of data and material: data will be available upon request to corresponding author

Data sharing statement: Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

Conflict of interest/ competing interest: The authors do not have any conflict of interest to declare.

Financial disclosure

None.

References

1. Tyler RW. *Basic principles of curriculum and instruction. University of Chicago press*; 2013.
2. Benstead K, Lara PC, Andreopoulos D, et al. Recommended ESTRO core curriculum for radiation oncology/radiotherapy 4th edition. *Radiother Oncol*. 2019;141:1–4.
3. Shakespeare TP, Back MF, Lu JJ, Wynne CJ, Bloomfield L. Design of an internationally accredited radiation oncology resident training program incorporating novel educational models. *Int J Radiat Oncol Biol Phys*. 2004;59(4):1157–1162, http://dx.doi.org/10.1016/j.ijrobp.2003.12.009.
4. McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. *Med Educ*. 2005;39(2):221–227, http://dx.doi.org/10.1111/j.1365-2929.2004.02060.x.
5. Frank JR, Danoff D. The CanMEDS initiative: Implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29(7):642–647, http://dx.doi.org/10.1080/01421590701746983.
6. Englander R, Frank JR, Carraccio C, et al. Toward a shared language for competency-based medical education. *Med Teach*. 2017;39(6):582–587, http://dx.doi.org/10.1080/0142159X.2017.1315066.
7. Burmeister J, Chen Z, Chetty IJ, et al. The American Society for Radiation Oncology's 2015 core physics curriculum for radiation oncology residents. *Int J Radiat Oncol Biol Phys*. 2016;95(4):1298–1303, http://dx.doi.org/10.1016/j.ijrobp.2016.03.012.
8. Giuliani ME, Gillan C, Milne RA, Uchino M, Millar BA, Catton P. Determining an imaging literacy curriculum for radiation oncologists: An international delphi study. *Int J Radiat Oncol Biol Phys*. 2014;88(4):961–966, http://dx.doi.org/10.1016/j.ijrobp.2013.12.009.
9. Krathwohl DR, Anderson LW. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman*; 2009.
10. Agarwal A, DeNunzio NJ, Ahuja D, Hirsch AE. Beyond the standard curriculum: a review of available opportunities for medical students to prepare for a career in radiation oncology. *Int J Radiat Oncol Biol Phys*. 2014;88(1):39–44, http://dx.doi.org/10.1016/j.ijrobp.2013.08.003.
11. Darby SC, Cutter DJ, Boerma M, et al. Radiation-related heart disease: Current knowledge and future prospects. *Int J Radiat Oncol Biol Phys*. 2010;76(3):656–665, http://dx.doi.org/10.1016/j.ijrobp.2009.09.064.
12. van Hagen P, Hulshof MC, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. 2012;366(22):2074–2084, http://dx.doi.org/10.1056/NEJMoa1112088.