Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access

Author manuscript

Int J Radiat Oncol Biol Phys. Author manuscript; available in PMC 2021 July 07.

Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2020 July 15; 107(4): 827–835. doi:10.1016/j.ijrobp.2020.04.011.

A Systematic Review of Contouring Guidelines in Radiation Oncology: Analysis of Frequency, Methodology, and Delivery of Consensus Recommendations

Diana Lin, BS[#*], Kaitlyn Lapen, BS[#*], Michael V. Sherer, MD[†], Jolie Kantor[*], Zhigang Zhang, PhD[‡], Lindsay M. Boyce, MLIS[§], Walter Bosch, DSc[‖], Deborah Korenstein, MD[¶], Erin F. Gillespie, MD[*,¶,*]

[*]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York

[†]Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California

[‡]Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York

[§]Memorial Sloan Kettering Library, Memorial Sloan Kettering Cancer Center, New York, New York

[‖]Department of Radiation Oncology, Washington University in St Louis, St Louis, Missouri

[¶]Center for Health Policy and Outcomes, Memorial Sloan Kettering Cancer Center, New York, New York

[#] These authors contributed equally to this work.

Abstract

Purpose—Clinical trials have described variation in radiation therapy plan quality, of which contour delineation is a key component, and linked this to inferior patient outcomes. In response, consensus guidelines have been developed to standardize contour delineation. This investigation assesses trends in contouring guidelines and examines the methodologies used to generate and deliver recommendations.

Methods and Materials—We conducted a literature search for contouring guidelines published after 1995. Of 11,124 citations, 332 were identified for full-text review to determine inclusion. We abstracted articles for the intent of the consensus process, key elements of the methodology, and mode of information delivery. A Fisher exact test was used to identify elements that differed among the guidelines generated for clinical trials and routine care.

Results—Overall, 142 guidelines were included, of which 16 (11%) were developed for a clinical trial. There was an increase in guideline publication over time (0 from 1995-1999 vs 65 from 2015- 2019; $P = .03$), particularly among recommendations for stereotactic radiation and

[*]Corresponding author: Erin F. Gillespie, MD, gillespe@mskcc.org.

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

Supplementary Data
Supplementary material for this article can be found at https://doi.org/10.1016/j.ijrobp.2020.04.011.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

brachytherapy. The most common disease sites were head and neck (24%), gastrointestinal (12%), and gynecologic (12%). Methods used to develop recommendations included literature review (50%) and image-based methods (45%). Panels included a median of 10 physicians (interquartile range, 7-16); 70% of panels represented multidisciplinary expertise. Guidelines developed for a clinical trial were more likely to include an image-based approach, with quantitative analysis of contours submitted by the panel members and to publish a full set of image-based recommendations ($P < .005$).

Conclusions—This review highlights an increase in consensus contouring recommendations over time. Guidelines focus on disease sites, such as head and neck, with evidence supporting a correlation between treatment planning and patient outcomes, although variation exists in the approach to the consensus process. Elements that may improve guideline acceptance (ie, image-based consensus contour analysis) and usability (ie, inclusion of a full image set) are more common in guidelines developed for clinical trials.

Introduction

Contour delineation is a critical process in treatment planning because it involves outlining tumor (or areas at risk of microscopic disease) as well as nearby organs at risk (OARs) to guide radiation therapy plans that optimize tumor control and reduce radiation toxicity. However, variation in contour delineation among providers is common and can affect the resulting plan quality and patient outcomes.[1–3] Reviews of prospective clinical trials for radiation therapy quality assurance (QA) have shown that variations in target volume delineation can result in increased treatment toxicity and decreased survival.[4–6] The incidence of variation in contour delineation has been evaluated most rigorously in the setting of clinical trials, in which radiation therapy QA processes document protocol deviations; a recent review found that *major* deviations in target delineation occurred in up to 13% of radiation therapy plans across 5 different trials.[7]

Consensus guidelines with recommendations for contour delineation have emerged in an effort to reduce contour variation.[8] Several studies have demonstrated that the use of guidelines and contouring atlases can reduce variation in delineation of both target volumes and OARs,[9–12] with additional evidence that these improvements in contour delineation can improve predicted tumor control and normal tissue complication probability.[11]

Despite their ability to increase the consistency of contour delineation and improve predicted clinical outcomes, consensus guidelines are underused.[13] Known barriers to their use include unfamiliarity with their existence and difficulty accessing the information when needed.[13] Additionally, prior studies have shown that recommendations are inconsistent across guidelines developed for the same disease site, thus complicating guideline selection and subsequent use.[1,14–16] Although standards for the development of clinical practice guidelines exist to ensure guideline quality and usability,[17] no such standards exist regarding guidelines for contour delineation, a uniquely image-based clinical skill. To our knowledge, characteristics of consensus contouring guidelines and methodologies have not been described. Thus, we conducted a comprehensive literature review to investigate trends in published recommendations, characterize and review methods used to develop consensus

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contouring guidelines, and explore how recommendations are distributed and displayed. This will allow us to identify potential barriers to consistent guideline development and dissemination to inform future implementation efforts.

Methods and Materials

Data sources

We conducted a comprehensive literature search of the PubMed, EMBASE, Cochrane Library, Web of Science, and Scopus databases to identify relevant consensus contouring guidelines published between January 1, 1995 and September 3, 2019 (Search Strategy; Text E1).

Data extraction and synthesis

A professional librarian (L.M.B.) performed the initial search, which produced 14,551 results; an additional 20 articles were identified by hand-searching the websites of professional organizations. After duplicates were removed, we identified 11,124 unique citations. One reviewer (D.L.) screened the abstracts of these citations for relevance and identified 332 potentially relevant publications. Three reviewers (D.L., K.L., M.V.S.) independently examined and identified articles for full-text inclusion. The inclusion criteria were as follows: development by a consensus group (defined as ≥2 authors), recommendations regarding contour delineation for radiation therapy included, and full-text publication available in English language after January 1, 1995. We selected this date because it represents the approximate time when intensity modulated radiation therapy (IMRT) became commercially available.[18] Abstracts and guidelines that provided only recommendations for aspects of radiation therapy planning and did not include volume delineation (eg, patient selection, dose, and fractionation) were excluded. Publications were included if 2 reviewers independently agreed that they met the outlined inclusion criteria. Any discrepancies were discussed with the research team, and, if warranted, an additional reviewer determined final inclusion. Of the 332 full-text articles reviewed, 142 met the inclusion criteria. We selected datapoints of interest based on standards for clinical practice guidelines from the Guidelines International Network[17]; datapoints collected included year of publication, disease site, endorsing organization, purpose of guideline development (eg, clinical trial QA), inclusion of multidisciplinary panel members (eg, radiation oncologists, medical oncologists, radiologists, and surgeons), type of radiation therapy, methods used to develop recommendations (eg, literature review), and methods used to display recommendations.

Statistical analysis

We used descriptive statistics to analyze guideline characteristics, a 2-sided Mann-Kendall test to assess trends in guideline publication over time, and a Fisher exact test to evaluate associations between guideline components and purpose of guideline development. We performed all statistical calculations using R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). A significance level of $\alpha = 0.05$ was set for statistical testing. This was adjusted for multiple comparisons using the Bonferroni correction,[19] which

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

resulted in *P* values <.005 (.05 / 9) being considered significant for the analysis on guideline components.

We performed this systematic review in compliance with the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) literature selection protocol (Fig. 1),[20] and it has been registered with the International Prospective Register of Systematic Reviews (PROSPERO).[21] This study did not require institutional review board approval.

Results

Trends over time

A total of 142 guidelines published between January 1, 1995 and September 3, 2019 met the inclusion criteria (Fig. 1). A comprehensive list of included articles is provided in Table E1. The complete list of guidelines is also available with interactive hyperlinks on eContour.org. An increase in the publication of consensus guidelines since the advent of 3-dimensional (3D) treatment planning was observed (*P* = .03); 0 articles were published from 1995 to 1999, 10 from 2000 to 2004, 22 from 2005 to 2009, 45 from 2010 to 2014, and 65 from 2015 to 2019 (Fig. 2). Most guidelines (82%) gave recommendations on contours for conventional external beam treatment (including both 3D conformal [3DCRT] and IMRT techniques); 11% gave recommendations on volume delineation for brachytherapy, 8% on stereotactic radiation, and 2% on proton therapy (Table 1). There was also an increase over time in the number of guidelines with recommendations for brachytherapy and stereotactic body radiation therapy/stereotactic radiosurgery (Fig. 2).

Guideline characteristics

The most common disease sites addressed by recommendations were head and neck (24%), gastrointestinal (12%), gynecologic (12%), and genitourinary (11%) (Table 1, Fig. 3). One guideline (1%) gave recommendations for more than one disease site: head and neck and genitourinary cancer.[22] The majority of guidelines addressed delineation for only radiation therapy target volumes (62%), whereas guidelines for both target volumes and OARs (24%) and OARs alone (14%) were less frequent. Five (4%) of the guidelines identified were developed for delineation of normal tissue or OARs without specifying a disease site.[23–27] Most guidelines (90%) gave recommendations for the delineation of volumes on a computed tomography (CT) scan; only 15 guidelines (11%) gave recommendations for magnetic resonance imaging (MRI)-based contouring exclusively (Table 1).

Consensus group and purpose

Overall, 62% (n = 87) of the guidelines were endorsed by a cooperative group or professional organization, the largest proportion of which were endorsed by national organizations based in the United States (30%). Following those endorsed by US-based organizations, guidelines were most commonly endorsed by organizations from Europe (22%), international organizations (20%), and national organizations from Spain (6%) (Fig. E1). The number of participants involved in the consensus was available in 49% of the guidelines and ranged from 2 to 129, with a median of 10 (interquartile range, 7–16). The majority of participants were radiation oncologists. Most (70%) of the consensus groups

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

were composed of multidisciplinary panelists; they involved a radiologist in 56 (39%) and a surgeon in 41 (29%) of the guidelines. Only 15 (11%) of the guidelines were developed specifically to ensure the quality of contours for a clinical trial (Table 1). Guidelines developed as part of a clinical trial were more likely to use image-based methods ($P < .005$), use tools to estimate a consensus contour and assess contour variation ($P < .005$), and provide a complete contoured image set ($P < .005$) (Table 2).

Consensus process

Most (92%) of the guidelines described the methods used to reach a consensus on contour recommendations. Without being mutually exclusive, the most common elements of the consensus process were a literature review (50% of guidelines) and the use of image-based methods (45%), in which members of the group contoured on patient imaging to reach a consensus (Fig. 4). Additional consensus methods included group discussions, surveys, and the use of cadavers and other gross specimens. The use of quantitative contour assessment was variable. Reported quantitative metrics are defined and summarized in Figure 4. Simultaneous truth and performance level estimation (STAPLE), an approach to calculating the degree of volume overlap on an image, was used in 13% of guidelines, whereas tools used to assess variation among multiple contours were used in 18% of guidelines.

Display of consensus recommendations

The presentation and display of contour recommendations within the publications were variable. Less than half of the guidelines (42%) were publicly accessible (ie, the full text could be accessed without a paid journal subscription). Guidelines often included representative axial imaging with a median of 11 images (range, 0-94), whereas 53% of the guidelines displayed a table summarizing definitions and contouring instructions. A complete atlas or image set for a contoured case outlining the provided recommendations was available for only 16% of the guidelines.

Discussion

This investigation represents the first systematic review of consensus contouring guidelines in radiation oncology. Although guideline use has been shown to enhance the accuracy of radiation treatment, improve clinical outcomes, and reduce toxicity,[10–12] inconsistent recommendations and poor dissemination challenge the utilization of contouring guidelines.[13] Although standards exist for the development of clinical practice guidelines in other medical fields,[17] contour delineation for radiation therapy lacks such structure. By analyzing the methods used to develop and disseminate available consensus guidelines, with an emphasis on processes undertaken in the setting of clinical trials, this study could potentially serve as a basis for the development of formalized standards for contouring guidelines.

With advances in highly conformal radiation therapy techniques, such as IMRT, stereotactic radiation (stereotactic body radiation therapy/stereotactic radiosurgery), and image guided brachytherapy, increasing the precision of radiation treatment,[8,18] this investigation confirms the concomitant increase in publication of contouring guidelines, as hypothesized in prior studies.[8,13] Only 2% of guidelines addressed proton therapy, and all were published after

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2018. This may be explained by proton therapy accounting for less than 1% of radiation treatments delivered in the United States,[28] as well as contouring guidelines defining the clinical target volume, which is thought to change minimally in the setting of proton treatment planning. Similarly, historically limited availability of magnetic resonance (MR) simulators and hybrid MR- linear accelerator (linac) systems[29] likely resulted in few published MR-based guidelines to date. However, with an emergence of MR-linac use in routine care (and with adaptive planning that can generate an even higher contouring burden), the need for consensus guidelines for MRI-based planning is clear and reportedly ongoing by various MR-linac consortia.

Despite the rapid growth in the publication of contouring guidelines, there are no widely accepted standards for contouring guideline development. Methods used to reach consensus were diverse in form and rigor. Among essential components of high-quality and reproducible guideline development standards established by the Guidelines International Network,[17] only half of the identified guidelines performed a literature review, 70% involved a multidisciplinary panel, and, although the median number of panelists was 10, only 22% of guidelines included the Guidelines International Network recommended number of 10 to 20 panel members.

Contouring guidelines developed as part of a clinical trial protocol often follow a more standardized and democratic process, including routine use of image-based approaches that incorporate quantitative contour analysis. STAPLE has been adopted by the Radiation Therapy Oncology Group, now part of NRG, and is used to create a single contour from multiple expert contours; contours created via STAPLE often are provided in atlases and are used as reference volumes for trials.[30,31] This standardization process may reflect the use of platforms and frameworks, such as the Global Clinical Trials Radiation Therapy Quality Assurance Harmonization Group, to discuss and endorse guidelines developed for clinical trials.[8,32]

Guideline delivery correlates with the likelihood of their use.[33] Overall, a minority (42%) of guidelines identified could be accessed without paid journal subscriptions. Moreover, although radiation oncology is increasingly a 3D image-based treatment, only a small subset (16%) of the guidelines included a full case image set, and most of these were in the context of clinical trial QA. These findings highlight real-world barriers to guideline dissemination and may play a role in their poor utilization.[2,13] There is a preference and a need identified among practicing clinicians for accessible, easy-to-use, image-based contouring resources, which are hypothesized to enhance accessibility at the point of care.[13,34]

Although the most common disease sites treated with radiation are breast, lung, and prostate cancer (based on recent data publicly released by the Center for Medicare and Medicaid Services, analysis under submission), contouring guidelines were identified most typically for head and neck, gynecologic, and gastrointestinal cancers. For these disease sites, however, more data exist that support a correlation between poor quality radiation and clinical outcomes,[4,5,35] and a recent analysis of an online contouring decision-support reference showed their guidelines to be the more frequently reviewed, demonstrating the real-world demand for contouring guidance in these areas.[34] We did find multiple guidelines

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

were sometimes available for the same disease type. This has been investigated recently in breast and anal cancer.[14–16] Methodological inconsistencies, as well as institutional differences, likely contribute to this phenomenon,[36] emphasizing the need to (1) optimize multi-institutional, and often multinational, expert involvement, and (2) reduce bias in the consensus process to ensure buy-in from diverse stakeholders.[7,37]

Although limited data exist to define a preferred approach to contouring guideline development, those conducted in the context of clinical trial QA are often subjected to formal review by multiple institutions and organizations and may be considered best practice. As such, the following elements should be considered to optimize the quality of consensus guidelines for volume delineation:

- involvement of a multidisciplinary panel (including a radiologist) with formal voting from members, either via survey or actively contouring an example case
- inclusion of a literature review with consideration of publications on patterns of recurrence as well as clinical trials with acceptable outcomes (and subsequent review of protocol specifications)
- quantitative contour analysis, such as STAPLE, the kappa statistic, and the Dice similarity coefficient, to assess contour variability and estimate consensus contours
- dissemination of a complete reference image set to improve point-of-care usability

However, we recognize the unique nature of contouring and recognize that additional adjustments for disease site and institutional feasibility may be necessary and appropriate. For example, radiologist expertise may be more useful in head and neck cancer—which involves complex anatomy and fused MRI[38–39]—than in breast cancer.

Several limitations of the present study exist. First, it is possible that some consensus contouring guidelines were not indexed within the searched databases; some may have been available only on professional-organization websites or within unpublished clinical trial protocols. This practice may be more pragmatic, but it omits journal peer review, which is an important component of this process, particularly regarding consensus methodology. Second, we excluded guidelines that were not published in English, even if they would have otherwise met inclusion. As a result, we may have identified a larger proportion of guidelines from English-speaking countries, potentially skewing the results of the geographic analysis for endorsed guidelines (Fig. E1). Third, although we observed some inconsistencies among recommendations in guidelines developed for the same disease site while we indexed data points, we did not conduct a formal or complete comparison of guideline recommendations. Our ability to comment on differences in recommendations within disease site groups is thus limited. Finally, the availability and reporting of some data points (eg, expertise of panel members) varied among the identified guidelines, thus limiting the analysis. Although precautions were taken to limit discrepancies by selecting straightforward endpoints, reviewing data abstraction techniques as a team, and discussing

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

any uncertainties with additional reviewers, it is possible that certain results of our review may not be completely reproducible.

Conclusions

This systematic review highlights trends in consensus contouring guideline publications and summarizes the methodology and components included in guidelines on contour delineation. With the increase in radiation complexity, guideline publication has increased over time and most commonly focuses on disease sites regarded as difficult to contour. Although guidelines developed for clinical trial protocols were more likely to include image-based methods, quantitative assessments, consensus contour estimations, and distribution of a complete reference-imaging data set, these components were often lacking in the overall population of published guidelines. Although recommendations based on processes used in the setting of clinical trials are provided here, there is an opportunity to develop more formal consensus contouring guideline standards to enhance their validity, dissemination, and usability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

The authors thank Sharif Elguindi, MS, for his advice on and review of the geometric indices of contour variation.

This article was supported by a 2019-2020 Dr. Luther Brady Resident Seed Grant from the American College of Radiation Oncology (M.V.S.). It was also supported by funds from the Memorial Sloan-Kettering Cancer Center Grant. Disclosures: E.F.G. is a cofounder of the educational website eContour.org; K.L. has been awarded a research fellowship, funded by grants for research and education related to eContour.org. D.K. reports grants from NCI, during the conduct of the study, and has other financial relationships with Vedanta Biosciences and Takeda, outside of the submitted work. W.B. reports grants from the US National Cancer Institute during the conduct of the study.

References

1. Berry SL, Boczkowski A, Ma R, et al., Interobserver variability in radiation therapy plan output: Results of a single-institution study, Pract Radiat Oncol 6, 2016, 442–449. [PubMed: 27374191]
2. Segedin B and Petric R, Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them?, Radiol Oncol 50, 2016, 254–262. [PubMed: 27679540]
3. Loo SW, Martin WM, Smith R, et al., Interobserver variation in parotid gland delineation: A study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature, Br J Radiol 85, 2012, 1070–1077. [PubMed: 22815411]
4. Abrams RA, Winter KA, Regine WF, et al., Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704—a phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas, Int J Radiat Oncol Biol Phys 82, 2012, 809–816. [PubMed: 21277694]
5. Peters LJ, O'Sullivan B, Giralt J, et al., Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02, J Clin Oncol 28, 2010, 2996–3001. [PubMed: 20479390]
6. Ohri N, Shen X, Dicker AP, et al., Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of cooperative group clinical trials, J Natl Cancer Inst 105, 2013, 387–393. [PubMed: 23468460]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

7. Cox S, Cleves A, Clementel E, et al.. Impact of deviations in target volume delineation—time for a new RTQA approach?, Radiother Oncol 137, 2019, 1–8. [PubMed: 31039468]
8. Chang ATY, Tan LT, Duke S, et al., Challenges for quality assurance of target volume delineation in clinical trials, Front Oncol 7, 2017, 221. [PubMed: 28993798]
9. Eminowicz G, Hall-Craggs M, Diez P, et al., Improving target volume delineation in intact cervical carcinoma: Literature review and step-by-step pictorial atlas to aid contouring, Pract Radiat Oncol 6, 2016, e203–e213. [PubMed: 27032573]
10. Hague C, Beasley W, Dixon L, et al., Use of a novel atlas for muscles of mastication to reduce inter observer variability in head and neck radiotherapy contouring, Radiother Oncol 130, 2019, 56–61. [PubMed: 30420234]
11. Mavroidis P, Giantsoudis D, Awan MJ, et al., Consequences of anorectal cancer atlas implementation in the cooperative group setting: Radiobiologic analysis of a prospective randomized in silico target delineation study, Radiother Oncol 112, 2014, 418–424. [PubMed: 24996454]
12. Fuller CD, Nijkamp J, Duppen JC, et al., Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting, Int J Radiat Oncol Biol Phys 79, 2011, 481–489. [PubMed: 20400244]
13. Sherer MV, Bryant AK, Wu AJ, et al., Assessment of contouring resource use and awareness of contouring guidelines among radiation oncologists, J Radiat Oncol 7, 2018, 103–109.
14. Gee HE, Moses L, Stuart K, et al., Contouring consensus guidelines in breast cancer radiotherapy: Comparison and systematic review of patterns of failure, J Med Imaging Radiat Oncol 63, 2019, 102–115. [PubMed: 30267561]
15. Kowalski ES, Feigenberg SJ, Cohen J, et al., Optimal target delineation and treatment techniques in the era of conformal photon and proton breast and regional nodal irradiation, Pract Radiat Oncol 10, 2020, 174–182. [PubMed: 31786234]
16. Dapper H, Oechsner M, Munch S, et al., Dosimetric comparison of organs at risk using different contouring guidelines for definition of the clinical target volume in anal cancer, Strahlenther Onkol 196, 2020, 368–375 [PubMed: 32016496]
17. Qaseem A, Forland F, Macbeth F, et al., Guidelines international network: Toward international standards for clinical practice guidelines, Ann Intern Med 156, 2012, 525–531. [PubMed: 22473437]
18. Hong TS, Ritter MA, Tome WA, et al., Intensity-modulated radiation therapy: Emerging cancer treatment technology, Br J Cancer 92, 2005, 1819–1824. [PubMed: 15856036]
19. Sedgwick P, Multiple hypothesis testing and Bonferroni's correction, BMJ349, 2014, g6284. [PubMed: 25331533]
20. Moher D, Liberati A, Tetzlaff J, et al., Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, BMJ339, 2009, b2535. [PubMed: 19622551]
21. Booth A, Clarke M, Dooley G, et al., The nuts and bolts of PROSPERO: An international prospective register of systematic reviews, Syst Rev 1, 2012, 2.
22. Maceira Rozas Model C, Rey Liste T, Garcia Caeiro AL, et al., Recommendations for treatment with IMRT for prostate and head-neck cancer. Axencia de avaliacion de tecnoloxias sanitarias de galicia, Clin Transl Onco. 8, 2006, 262–265.
23. Kong FM, Ritter T, Quint DJ, et al., Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus, Int J Radiat Oncol Biol Phys Si, 2011, 1442–1457.
24. Yi SK, Mak W, Yang CC, et al., Development of a standardized method for contouring the lumbosacral plexus: A preliminary dosimetric analysis of this organ at risk among 15 patients treated with intensity-modulated radiotherapy for lower gastrointestinal cancers and the incidence of radiation-induced lumbosacral plexopathy, Int J Radiat Oncol Biol Phys 84, 2012, 376–382. [PubMed: 22342301]
25. Gay HA, Barthold HJ, O'Meara E, et al., Pelvic normal tissue contouring guidelines for radiation therapy: A radiation therapy oncology group consensus panel atlas, Int J Radiat Oncol Biol Phys 83, 2012, e353–e362. [PubMed: 22483697]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

26. Wright JL, Yom SS, Awan MJ, et al., Standardizing normal tissue contouring for radiation therapy treatment planning: An ASTRO consensus paper, Pract Radiat Oncol 9, 2019, 65–72. [PubMed: 30576843]
27. Jabbour SK, Hashem SA, Bosch W, et al., Upper abdominal normal organ contouring guidelines and atlas: A radiation therapy oncology group consensus, Pract Radiat Oncol 4, 2014, 82–89. [PubMed: 24890348]
28. Waddle MR, Sio TT, Van Houten HK, et al., Photon and proton radiation therapy utilization in a population of more than 100 million commercially insured patients, Int J Radiat Oncol Biol Phys 99, 2017, 1078–1082. [PubMed: 28939229]
29. Corradini S, Alongi F, Andratschke N, et al., MR-guidance in clinical reality: Current treatment challenges and future perspectives, Radiat Oncol 14, 2019, 92. [PubMed: 31167658]
30. Warfield SK, Zou KH and Wells WM, Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation, IEEE Trans Med Imaging 23, 2004, 903–921. [PubMed: 15250643]
31. Gwynne S, Spezi E, Sebag-Montefiore D, et al., Improving radiotherapy quality assurance in clinical trials: Assessment of target volume delineation of the pre-accrual benchmark case, Br J Radiol 86, 2013, 20120398. [PubMed: 23392188]
32. Melidis C, Bosch WR, Izewska J, et al., Radiation therapy quality assurance in clinical trials—global harmonisation group, Radiother Oncol 111, 2014, 327–329. [PubMed: 24813094]
33. Grimshaw JM and Russell IT, Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations, Lancet 342, 1993, 1317–1322. [PubMed: 7901634]
34. Sherer MV, Lin D, Puri K, et al., Development and usage of eContour, a novel, three-dimensional, image-based web site to facilitate access to contouring guidelines at the point of care, JCO Clin Cancer Inform 3, 2019, 1–9.
35. Weber DC, Tomsej M, Melidis C, et al., QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy, Radiother Oncol 105, 2012, 4–8. [PubMed: 22985777]
36. Shaughnessy AF, Cosgrove L and Lexchin JR, The need to systematically evaluate clinical practice guidelines, J Am Board Fam Med 29, 2016, 644–648. [PubMed: 28076244]
37. Cluzeau F, Wedzicha JA, Kelson M, et al., Stakeholder involvement: How to do it right: Article 9 in integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report, Proc Am Thorac Soc 9, 2012, 269–273. [PubMed: 23256170]
38. Terezakis SA, Heron DE, Lavigne RF, et al., What the diagnostic radiologist needs to know about radiation oncology, Radioiogy 261, 2011, 30–44.
39. Dawson LA and Menard C, Imaging in radiation oncology: A perspective, Oncologist 15, 2010, 338–349. [PubMed: 20413639]
40. Allozi R, Li XA, White J, et al., Tools for consensus analysis of experts' contours for radiotherapy structure definitions, Radiother Oncol 97, 2010, 572–578. [PubMed: 20708285]
41. Viera AJ and Garrett JM, Understanding interobserver agreement: The kappa statistic, Fam Med 37, 2005, 360–363. [PubMed: 15883903]
42. Dice LR, Measures of the amount of ecologic association between species, Ecology26, 1945, 297–302.
43. Zou KH, Warfield SK, Bharatha A, et al., Statistical validation of image segmentation quality based on a spatial overlap index - scientific reports, Acad Radiol 11, 2004, 178–189. [PubMed: 14974593]
44. Van de Velde J, Vercauteren T, De Gersem W, et al., Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring, Strahlenther Onkol 190, 2014, 628–632, 634–635. [PubMed: 24715246]
45. Taha AA and Hanbury A, An efficient algorithm for calculating the exact Hausdorff distance, IEEE Trans Pattern Anal Mach Intell 37, 2015, 2153–2163. [PubMed: 26440258]
46. Huttenlocher DP, Klanderman GA and Rucklidge WJ, Comparing images using the Hausdorff distance, IEEE Trans Pattern Anal Mach Intell 15, 1993, 850–863.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

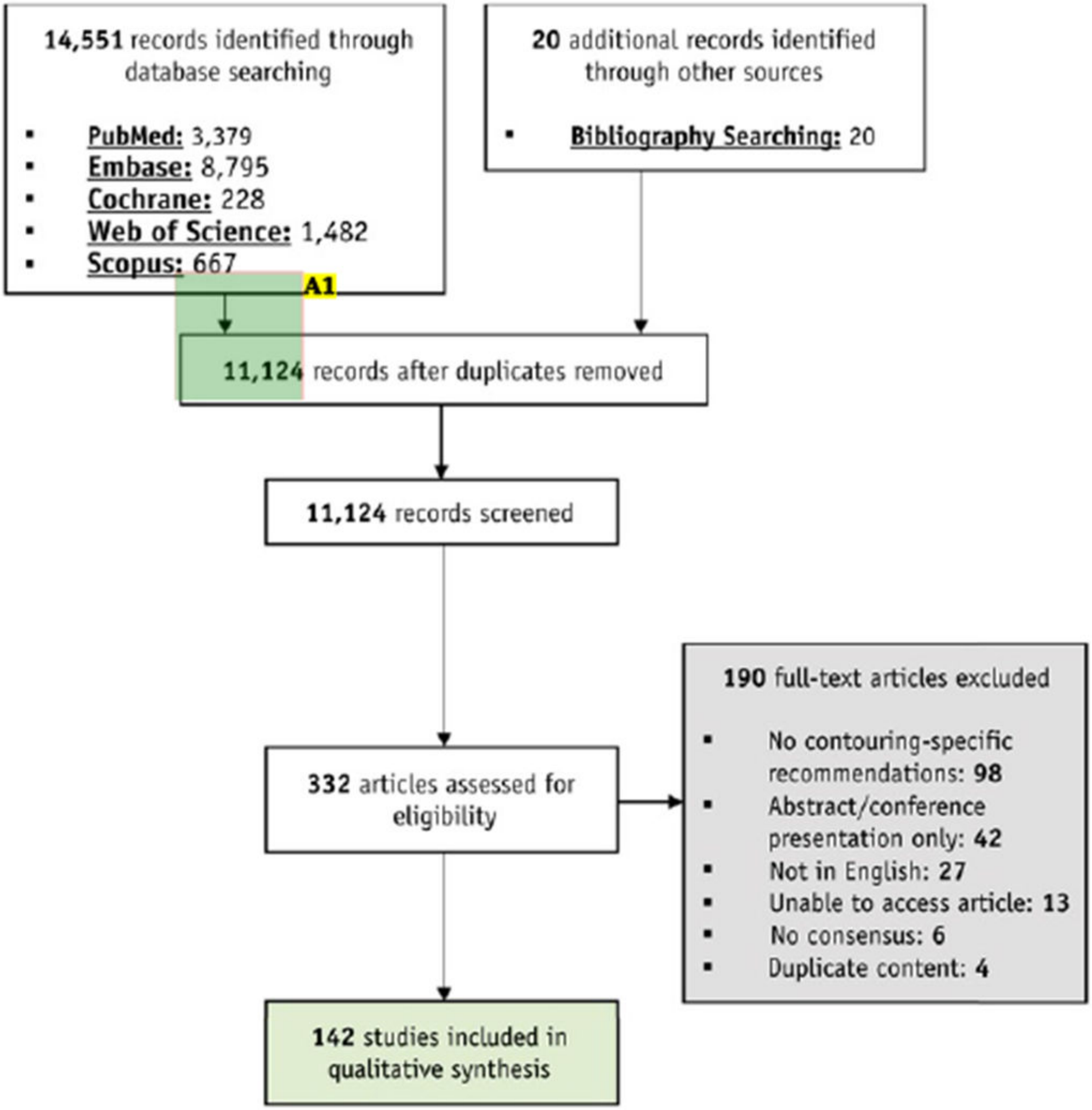


Fig. 1.

Study flow-diagram adapted from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

Annotations:

A1. The arrow sizing is inconsistent, would it be possible to fix this?

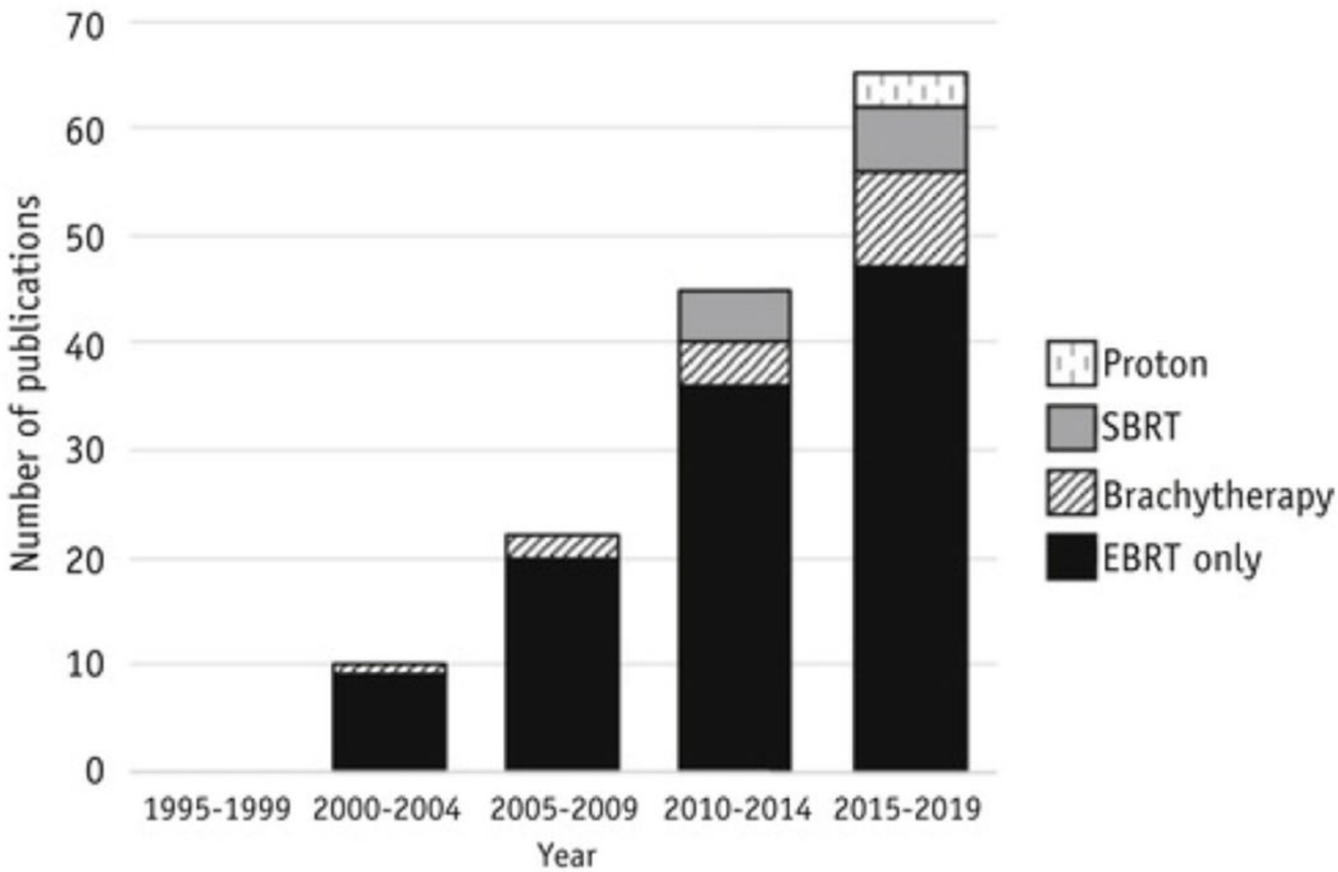


Fig. 2. Publication of guidelines with contour recommendations over time by type of radiation therapy.

External beam radiation therapy (EBRT) includes both 3-dimensional conformal (3DCRT) and intensity modulated radiation therapy (IMRT) techniques.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Type	Count (n = 142)
Head & neck (H&N)	34
Gastrointestinal (GI)	17
Gynecologic (GYN)	17
Genitourinary (GU)	16
Other	14
Normal tissue	5
Non-melanoma skin cancer	3
Metastases	3
Multiple sites	1
Myeloma	1
Breast	12
Lung/Thorax	8
Lymphoma	9
Central nervous system (CNS)	9
Sarcoma	6

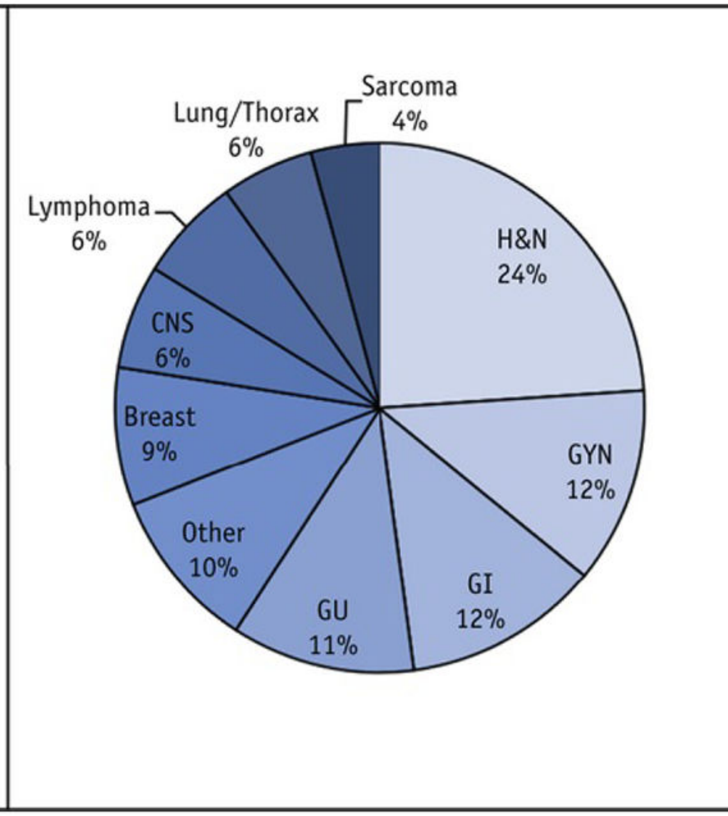


Fig. 3.
Distribution of identified guidelines by disease site of focus.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

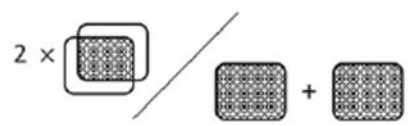
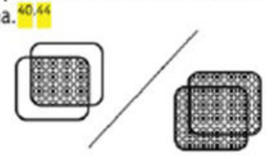
Methods used to reach consensus		Count	% Total (n = 142)
Provided description of methods		131	92%
Performed a literature review		71	50%
Image-based methods		64	45%
Quantitative consensus tools used			
Tools to estimate consensus contours	*Simultaneous Truth and Performance Level Estimation (STAPLE)* • An algorithm used to combine multiple contours to establish a single "true contour" that represents a high-level of agreement amongst the set. Often it is used to create a contour that represents a group consensus.[30,40] 	19	13%
Tools to assess contour variation	Geometric indices used to describe variation among contours commonly include (but are not limited to) the following: *Kappa statistic* • An analysis that measures the degree of agreement that occurs among raters beyond chance agreement.[41] It yields a value that describes interrater agreement between contours for a given image.[40] *Dice similarity coefficient (DSC)* • A metric that can be used to compute the degree of association or spatial similarity between two samples.[40,42] Represents the intersection area of two samples over the sum of their area.[43]  *Jaccard index* • Similar measure to DSC; represents the area of contour overlap or intersection, over the union of the contoured area.[40,44]  *Hausdorff distance* • A measure of maximal distance between two sets of points in space; it is computed by measuring the distance between corresponding points in a set of overlaid samples.[45,46]	25	18%

Fig. 4.

Summary and explanations of methods used to develop guidelines.[40–46]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Summary of characteristics of identified guidelines*

Characteristic	Overall (n = 142)	H&N (n = 34)	GI (n = 17)	GYN (n = 17)	GU (n = 16)	Breast (n = 12)	Lymphoma (n = 9)	Lung/ Thorax (n = 8)	CNS (n = 9)	Sarcoma (n = 6)	Other (n = 14)
Endorsed	87 (61%)	9 (26%)	6 (35%)	14 (82%)	12 (75%)	9 (75%)	9 (100%)	7 (88%)	6 (67%)	4 (67%)	11 (79%)
Developed for trial	15 (11%)	0 (0%)	2 (12%)	5 (29%)	2 (13%)	2 (17%)	2 (22%)	0 (0%)	0 (0%)	0 (0%)	2 (14%)
Multidisciplinary panel	100 (70%)	27 (79%)	12 (71%)	11 (65%)	13 (81%)	8 (67%)	5 (56%)	5 (63%)	7 (78%)	3 (50%)	9 (64%)
Radiologist	56 (39%)	20 (59%)	10 (59%)	7 (41%)	6 (38%)	3 (25%)	0 (0%)	2 (25%)	3 (33%)	0 (0%)	5 (36%)
Surgeon	41 (29%)	13 (38%)	4 (24%)	4 (24%)	7 (44%)	3 (25%)	0 (0%)	1 (13%)	3 (33%)	2 (33%)	4 (29%)
Contoured volume											
Target	88 (62%)	18 (53%)	12 (71%)	12 (71%)	9 (56%)	9 (75%)	7 (78%)	6 (75%)	6 (67%)	4 (67%)	5 (36%)
OARs	20 (14%)	10 (29%)	1 (6%)	0 (0%)	1 (6%)	1 (8%)	0 (0%)	0 (0%)	2 (22%)	0 (0%)	5 (36%)
Both	34 (24%)	6 (18%)	4 (24%)	5 (29%)	6 (38%)	2 (17%)	2 (22%)	2 (25%)	1 (11%)	2 (33%)	4 (29%)
Type of RT											
EBRT	117 (82%)	33 (97%)	16 (94%)	10 (59%)	15 (96%)	9 (75%)	8 (89%)	8 (100%)	7 (78%)	5 (83%)	6 (43%)
Brachytherapy	16 (11%)	1 (3%)	0 (0%)	7 (41%)	1 (6%)	3 (25%)	0 (0%)	0 (0%)	0 (0%)	1 (17%)	3 (21%)
SBRT/SRS	11 (8%)	0 (0%)	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (13%)	3 (33%)	0 (0%)	6 (43%)
Proton	3 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (11%)	2 (22%)	1 (11%)	0 (0%)	0 (0%)
Type of scan											
CT	90 (63%)	18 (53%)	12 (71%)	7 (41%)	12 (75%)	12 (100%)	9 (100%)	8 (100%)	2 (22%)	2 (33%)	8 (57%)
MR-assisted CT	50 (35%)	16 (47%)	4 (24%)	9 (53%)	4 (25%)	0 (0%)	0 (0%)	0 (0%)	7 (78%)	4 (67%)	6 (43%)
MR-based	15 (11%)	5 (15%)	2 (12%)	3 (18%)	1 (6%)	0 (0%)	0 (0%)	0 (0%)	4 (44%)	0 (0%)	0 (0%)
Publicly accessible	59 (42%)	9 (26%)	11 (65%)	9 (53%)	3 (19%)	4 (33%)	2 (22%)	5 (63%)	3 (33%)	3 (50%)	10 (71%)

Abbreviations: CNS = central nervous system; CT = computed tomography; EBRT = external beam radiation therapy; GI = gastrointestinal; GU = genitourinary; GYN = gynecologic; H&N = head and neck; MR = magnetic resonance; OARs = organs at risk; RT = radiation therapy; SBRT/SRS = stereotactic body radiation therapy/stereotactic radiosurgery.

*Not all characteristics were mutually exclusive and percentages may not add up to 100%. If a single guideline fell under more than 1 characteristic category, it was counted within each category.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Comparison of characteristics between clinical trial and nontrial guidelines

Characteristic	Trial (n = 15)	Nontrial (n = 127)	Total	*P* value
	n (%)	n (%)	n	
Endorsed by organization	13 (87%)	74 (58%)	87	.047
Multidisciplinary panel	11 (73%)	89 (70%)	100	1.000
Literature review	3 (20%)	68 (54%)	124	.021
Image-based methods	12 (80%)	52 (41%)	64	.001
STAPLE analysis	6 (40%)	13 (10%)	19	.004
Assessment of contour variation	8 (53%)	17 (13%)	25	<.005
Full image set available	8 (53%)	15 (12%)	23	<.005
Displayed definition table	10 (67%)	68 (54%)	78	.416
Full-text publicly available	9 (60%)	50 (39%)	59	.167

Abbreviation: STAPLE = simultaneous truth and performance level estimation.