

Beyond One-Time RECIST 1.1 Training: Knowledge Acquisition and Retention

Response Evaluation Criteria in Solid Tumours version 1.1 (RECIST 1.1) provides standardised criteria for evaluating changes in tumour burden in solid-tumour clinical trials. Since their publication in 2009, the criteria have been incorporated into the objectives and endpoints of numerous oncology clinical trials.^{1,2} Subsequent clarifications from the RECIST Committee have addressed recurring questions involving lesion selection, lymph node assessment, non-target disease, new lesions and the determination of progressive disease.³

Standardisation does not, however, eliminate the possibility of incorrect interpretation and mistakes. A recent comparison of routine clinical interpretations with formal RECIST 1.1 assessments found substantial disagreement in progressive disease classification, most commonly related to target lesion measurements and progression of non-target lesions.⁴ Such findings reinforce the need for appropriate training across the clinical trial team.

Clinical research associates (CRAs) do not replace investigators, radiologists or central readers and they do not independently perform a medical interpretation of imaging. Nevertheless, they regularly interact with RECIST-derived source data. Their responsibilities include source data review, verification of data entered into the electronic data capture (EDC) system, identification of inconsistencies between imaging reports and case report forms, follow-up of missing assessments and escalation of potential discrepancies to medical or imaging experts.⁵ These activities require not simply familiarity with RECIST terminology – they require a practical understanding of how the criteria are applied, or in other words – practical skills rather than theoretical knowledge.

Several educational resources have been developed to improve the application of RECIST 1.1 across clinical trial roles.^{6–8} These include training modules for site personnel, role-specific instruction for clinical data managers and institutional toolkits recommending ongoing training with knowledge assessment. Published evaluations remain relatively limited. Andoh and colleagues evaluated educational and feedback interventions intended to improve radiologists' application of RECIST and found that the combination of distributed educational materials and structured feedback reduced the frequency of errors in radiology reports.⁹ However, there remains little published evidence on longitudinal, role-specific RECIST training for CRAs that combines baseline assessment, immediate evaluation, facilitated remediation and subsequent assessment of knowledge retention.

Conventional training is often delivered as a single event and measured only by completion. Completion confirms that the material was accessed, but not that it was understood or retained. To address this limitation, we developed a structured RECIST 1.1 training programme combining an interactive tutorial, baseline and post-training knowledge assessments, targeted remediation and a follow-up assessment approximately three months later. This article describes the programme and summarises objective and participant-reported outcomes collected between March 2021 and June 2026. Preliminary findings from an earlier analysis of this training programme were presented at the 2025 ASCO Annual Meeting.¹⁰ The present article reports an expanded dataset through 24 June 2026 and includes additional longitudinal, year-specific and participant-experience analyses.

Designing Training Around the CRA Role

The programme, RECIST 1.1 Practice Essentials, was designed for CRAs and other clinical research professionals involved in reviewing oncology trial data. It was not intended to qualify participants to

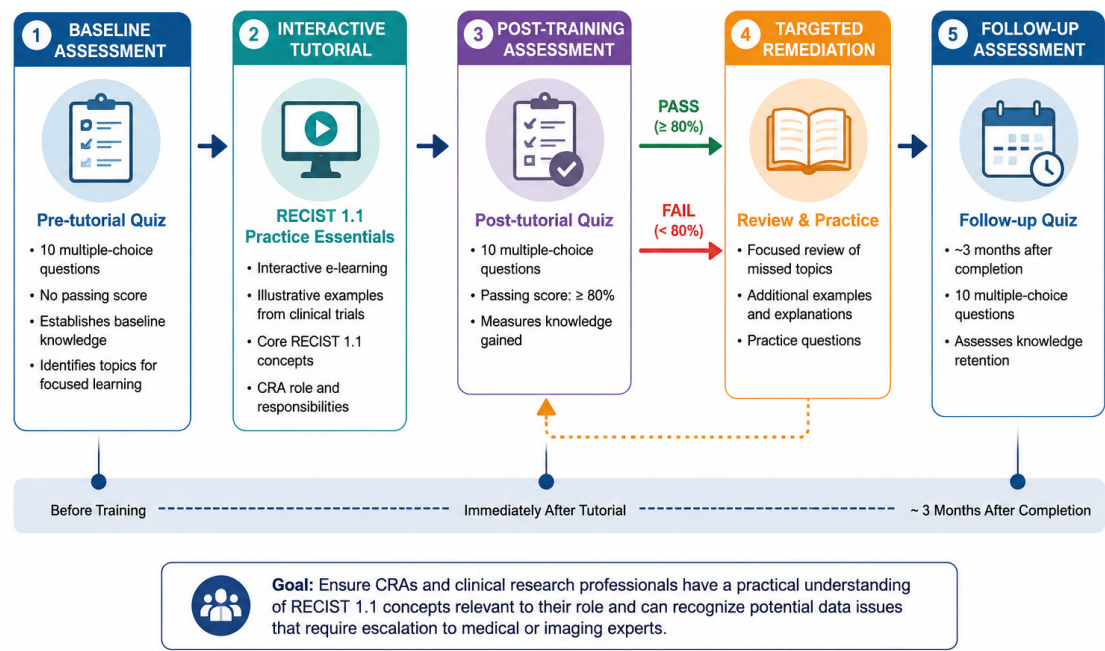


Figure 1. Structure of RECIST 1.1 training programme

perform radiological reads or independently determine tumour response.

The content focuses on practical concepts that CRAs are likely to encounter: selection and measurement of target lesions, evaluation of pathological lymph nodes, assessment of non-target lesions, calculation of changes in the sum of diameters, use of baseline and nadir measurements, identification of new lesions and determination of response to treatment and progressive disease.

The tutorial content and assessment questions were reviewed by several internal and external medical oncologists with experience in applying RECIST 1.1 in clinical trials. Their review focused on clinical accuracy, consistency with RECIST 1.1 principles, relevance to situations encountered by CRAs and clarity of the explanations and response options. Feedback from these reviews was incorporated into the training materials before implementation and during subsequent updates to the programme.

The programme has five components (Figure 1).

First, participants are offered a pre-tutorial quiz without a passing score. This ungraded assessment establishes a baseline against which learning progress can be evaluated by comparing the results with those of the final post-tutorial quiz.

Second, an interactive tutorial explains core RECIST 1.1 concepts through illustrative examples derived from clinical trial practice. The training repeatedly reinforces the boundary between recognising a potential inconsistency and making a medical or radiological determination.

Third, participants complete a multiple-choice post-tutorial quiz. An 80% passing threshold is applied. The assessment measures immediate knowledge acquisition and allows recurring misunderstandings to be identified across trainee cohorts.

Fourth, participants who do not initially achieve the passing score attend a facilitated question-and-answer session with a medical expert. The relevant questions and the rationale for the correct answers are discussed before reassessment. An unsuccessful first attempt is therefore treated as an opportunity for remediation rather than simply recorded as a training deficiency.

Finally, participants are invited to complete a follow-up quiz approximately three months after the tutorial. This assessment evaluates the extent to which the knowledge demonstrated immediately after training is retained after participants return to routine work.

Methods

Training records available as of 24 June 2026 were reviewed. Quiz scores were expressed as percentages. Analyses were restricted to records containing interpretable numeric scores ranging from 0% to 100%. Missing, non-interpretable and clearly invalid entries were excluded.

The pre-tutorial, post-tutorial and follow-up assessments each comprised ten equally weighted multiple-choice questions covering the core concepts addressed in the tutorial. Partially overlapping question sets were used at the three assessment time points. Participants completed the assessments independently. Correct answers and explanations were reviewed during remediation when required; however, participants did not otherwise have access to the answer key before completing the follow-up assessment. Minor

revisions to wording and presentation were introduced over time, but the core RECIST 1.1 concepts assessed remained consistent throughout the evaluation period.

The primary analysis compared paired pre-tutorial and initial post-tutorial scores. A longitudinal complete-record analysis included records with scores available at all three time points: pre-tutorial, post-tutorial, and follow-up. Mean changes were evaluated using paired tests and are presented with 95% confidence intervals. Because quiz scores were discrete and bounded between 0% and 100%, two-sided Wilcoxon signed-rank tests were performed as sensitivity analyses for all paired comparisons. Year-specific results were summarised descriptively because cohort composition, previous oncology experience and follow-up availability varied over time.

Participant-experience data were collected using a five-item survey with response options ranging from strongly disagree to strongly agree. 'Agree' and 'strongly agree,' responses were combined as favourable responses, while 'disagree' and 'strongly disagree,' responses were combined as unfavourable responses. Free-text comments were reviewed to identify recurring themes.

Immediate Knowledge Acquisition

A total of 382 paired pre- and post-tutorial assessment records were available for analysis. The mean pre-tutorial score was 55.3%, compared with 76.3% after the tutorial completion. The mean improvement was 21.0 percentage points (95% confidence interval, 18.6 to 23.4; $p < 0.001$) (Figure 2). The result was confirmed in the Wilcoxon signed-rank sensitivity analysis ($p < 0.001$). Performance improved in 76.7% of paired records, remained unchanged in 11.3% and decreased in 12.0%. The median change was 20 percentage points. All participants who initially scored below the 80% passing threshold completed facilitated remediation and reassessment, after which each achieved the required score. The magnitude of improvement is operationally meaningful. At baseline, participants answered only slightly more than half of the assessed questions correctly on average. Following the tutorial, average performance increased by approximately two correct answers on a ten-question assessment. The baseline quiz is therefore more than an introductory exercise: it provides the comparator needed to demonstrate the learning attributable to the programme.

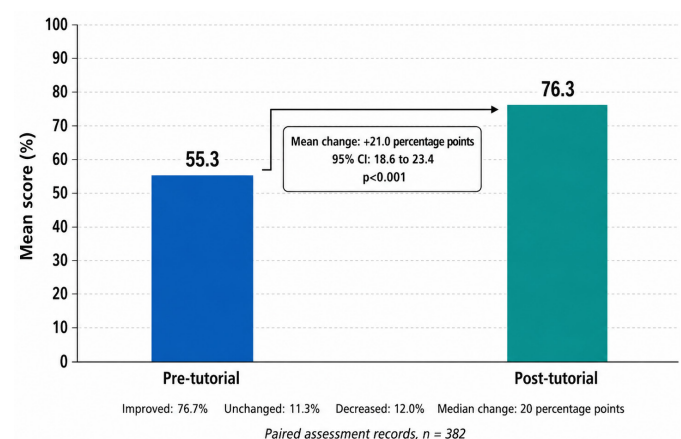


Figure 2. Mean assessment scores before and after the tutorial.

Knowledge Retention

A total of 116 assessment records contained results for all three time points. In this cohort, mean scores were 56.2% before the tutorial, 78.2% post-tutorial and 77.8% at follow-up (Figure 3). The improvement from baseline to the post-tutorial assessment was statistically significant ($p < 0.001$), as was the difference between baseline and follow-up ($p < 0.001$). These findings were confirmed

in Wilcoxon signed-rank sensitivity analyses (both $p<0.001$). In contrast, there was no meaningful decline between the post-tutorial and follow-up assessments: the mean difference was -0.4 percentage points (95% confidence interval, -5.0 to 4.2 ; $p=0.853$). The corresponding Wilcoxon signed-rank test was also non-significant ($p=0.82$). In 61.2% of complete records, the follow-up score was the same as or higher than the immediate post-training score.

These results suggest that the knowledge gain was generally maintained approximately three months after training. This is particularly relevant because participants may not immediately encounter every RECIST scenario in their assigned trials. A delayed assessment provides a more realistic indication of whether knowledge remains available for later application than an immediate post-training quiz alone.

Results Across Training Years

Improvement from baseline to post-tutorial assessment was observed in every annual cohort from 2021 through 2026 (Figure 4). Baseline scores were lowest during the programme’s initial year and generally higher in later cohorts, while mean post-tutorial scores became comparatively stable. These findings may reflect both improvements to the course and differences in composition of annual trainee groups. Year-to-year comparisons should therefore remain descriptive. They do not establish that changes were caused solely by revisions to the training. Nevertheless, the recurring pre-to-post improvement across successive cohorts indicates that the overall result was not limited to a single group or implementation period.

Using Assessment Errors to Improve the Programme

Assessment data can identify not only whether learning occurred, but also which concepts require greater emphasis. An earlier analysis of baseline errors found that 44.2% involved defining or measuring target lesions, 32.2% involved determining progressive disease, 21.2% involved non-target lesions and 2.4% involved methods of assessment. These findings were used to restructure the tutorial and expand explanations and examples in the most difficult areas. The pattern is consistent with the practical complexity of RECIST 1.1. Target-lesion selection and measurement require several linked rules, while progressive-disease determination may require consideration of nadir, absolute and relative change, non-target disease and new lesions.^{4,11,12} This illustrates an important advantage of a multi-stage training model: aggregated assessment findings create a feedback

Statement	Responses, n	Favourable	Uncertain	Unfavourable
Well engaged during the session	313	88.2%	9.3%	2.6%
Appropriately challenged by the presentation and quiz	312	86.5%	8.3%	5.1%
Comfortable with pace and duration	313	85.9%	7.7%	6.4%
Tutorial was easy to understand	313	79.9%	11.8%	8.3%
Able to apply the learning immediately	312	76.6%	16.7%	6.7%

Table 1. Participant survey results

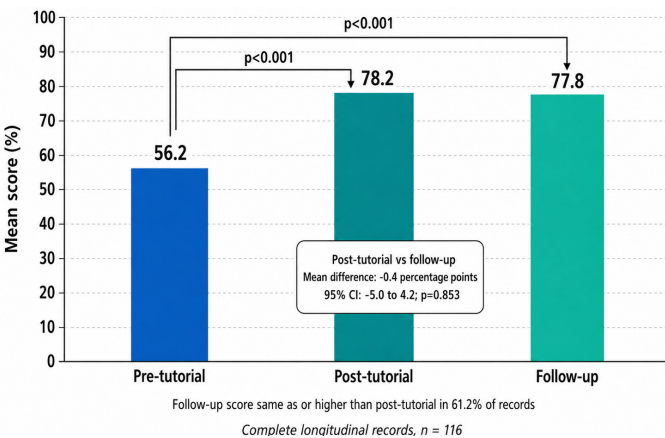


Figure 3. Mean assessment scores at baseline, post-tutorial and follow-up.

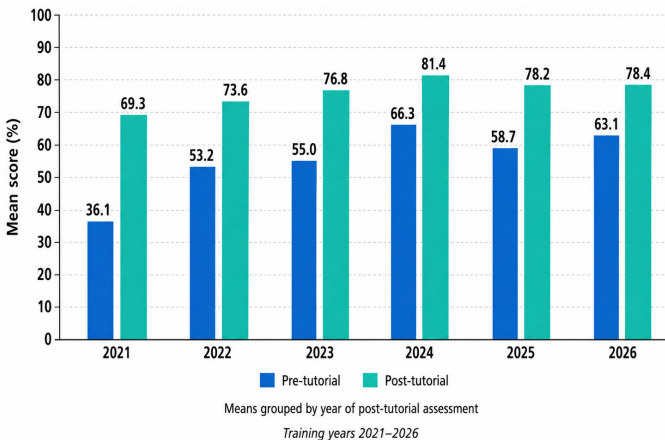


Figure 4. Pre- and post-tutorial assessment scores by training year, 2021–2026.

loop. The course can evolve in response to documented learning needs rather than relying only on faculty impressions.

Participant Experience and Perceived Practical Value

A total of 313 survey responses were collected between 19 March 2021 and 24 June 2026, representing 296 individuals; 17 individuals submitted responses following two training occurrences. Results are summarised in Table 1. Engagement received the highest rating: 88.2% of responses agreed or strongly agreed that participants were well engaged during the session. A further 86.5% reported that the presentation and quiz questions were appropriately challenging, and 85.9% were comfortable with the pace and duration. The tutorial was considered easy to understand in 79.9% of responses, while 76.6% indicated that the participant expected to apply the acquired knowledge immediately. The comparatively lower rating for immediate applicability is understandable in a heterogeneous population. Some participants may not have been assigned to an oncology trial at the time of training, and the opportunity to use individual RECIST concepts varies by role and project.

Seventy-nine respondents provided free-text comments. Positive comments commonly described the training as useful, clear and relevant to clinical trial work. Suggestions included adding more practical examples, providing reference materials or a written script after the training, allowing more time to discuss difficult questions and dividing complex material into shorter learning segments. Some respondents noted that the course was more challenging for participants without a medical background or for those completing it in a non-native language. These comments generated practical improvement opportunities. They supported greater use of applied scenarios, clearer explanations of quiz answers and additional reinforcement of the distinction between identifying a potential inconsistency and independently making a response assessment.



Practical Lessons for CROs

Several principles from this programme may be transferable to other areas of clinical trial training. Training should be role specific. CRAs need enough knowledge to review RECIST-derived data critically and escalate discrepancies, but they do not require the same curriculum as investigators, radiologists or central readers. Defining this boundary improves both relevance and compliance with assigned responsibilities.

Training completion should not be treated as evidence of competence. A completion record confirms access to the material, whereas a knowledge assessment provides evidence of understanding. A baseline assessment also allows the magnitude of learning to be measured rather than inferred. Remediation should be built into the training pathway. When participants do not meet the passing threshold, facilitated discussion can distinguish between a terminology problem, a calculation error, misunderstanding of a clinical scenario or confusion between similar concepts. This is more educationally useful than repeated testing without explanation. Knowledge retention should be evaluated. Immediate post-training performance reflects short-term recall. A delayed assessment helps determine whether learning remains available for later use and can identify the need for refresher training. Finally, further course development should be based on the training data and results. Recurring errors and participant feedback can be translated into revised examples, clearer explanations and targeted supplementary materials. This creates a continuously improving programme rather than a static annual requirement. We adapted the same structure to other oncology response criteria, including iRECIST, Lugano and RANO, while retaining the core elements of baseline assessment, applied learning, remediation and follow-up evaluation.

Limitations

This evaluation was observational and did not include an untrained control group. Quiz scores are indirect measures of professional performance and were not linked to trial-level quality indicators such as query rates, protocol deviations, data corrections or concordance with central imaging review. Follow-up results were available for fewer records than pre- and post-tutorial results, creating the possibility of selection bias.

Future analyses could link training participation with operational quality indicators, compare outcomes by professional role and previous oncology experience and assess the effect of targeted refresher modules.

Conclusion

Consistent use of RECIST 1.1 depends not only on the expertise of investigators and imaging specialists, but also on the knowledge of clinical research professionals who review and manage oncology trial data. A structured training model combining a practical tutorial, baseline and post-training assessments, facilitated remediation and a three-month follow-up produced a substantial improvement in RECIST 1.1 knowledge. Among records with complete longitudinal data, the improvement was maintained at follow-up. Participant feedback also showed high engagement, appropriate challenge and favourable perceptions of the course's clarity and relevance.

The findings support moving beyond one-time, completion-based training toward a model that measures knowledge acquisition, addresses misunderstandings, evaluates retention and uses evidence to improve the curriculum. Such an approach may strengthen the ability of CRAs to recognise inconsistencies, escalate potential issues and support the quality of oncology clinical trial data.

REFERENCES

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*. 2009;45(2):228–247. doi:10.1016/j.ejca.2008.10.026.
2. US Food and Drug Administration. Clinical Trial Imaging Endpoint Process Standards: Guidance for Industry. April 2018.
3. Schwartz LH, Litière S, de Vries E, et al. RECIST 1.1—Update and clarification: from the RECIST Committee. *European Journal of Cancer*. 2016;62:132–137. doi:10.1016/j.ejca.2016.03.081.
4. Siegel MJ, Ippolito JE, Wahl RL, Siegel BA. Discrepant assessments of progressive disease in clinical trials between routine clinical reads and formal RECIST 1.1 interpretations. *Radiology: Imaging Cancer*. 2023;5(5):e230001. doi:10.1148/rycan.230001.
5. Gouel P, Callonnec F, Levêque É, et al. Evaluation of the capability and reproducibility of RECIST 1.1 measurements by technologists in breast cancer follow-up: a pilot study. *Scientific Reports*. 2023;13:9148. doi:10.1038/s41598-023-36315-w
6. NRG Oncology. RECIST 1.1 Criteria Toolkit: A Resource for Success in Clinical Trials with Solid Tumor Evaluation Requirements. Available at: <https://www.nrgoncology.org/Home/News/Post/recist-11-criteria-toolkit-a-resource-for-success-in-clinical-trials-with-solid-tumor-evaluation-requirements/> Accessed 14 July 2026
7. Society for Clinical Research Sites. RECIST Training Modules: Oncology Training Programs for Site Staff Conducting Clinical Research [Internet]. Training modules developed by the Clinical Imaging Team at Merck Research Laboratories. Available from: <https://myscrs.org/learning-campus/irecist-modules/>. Accessed 14 July 2026
8. Stephenson B. RECIST—A Practical Guide for Data Managers [Internet]. Association for Clinical Data Management; 14 May 2019. Available from: <https://acdmglobal.org/courses/recist-a-practical-guide-for-data-managers/>. Accessed 14 July 2026
9. Andoh H, McNulty NJ, Lewis PJ. Improving accuracy in reporting CT scans of oncology patients: assessing the effect of education and feedback interventions on the application of the Response Evaluation Criteria in Solid Tumors criteria. *Academic Radiology*. 2013;20(3):351–357. doi:10.1016/j.acra.2012.12.002
10. Kosov M, Bowman N, Maximovich A, Field KM, Murphy F, Semeniuk N. Importance of RECIST 1.1 training program for improving performance of contract research organization staff in solid tumor clinical trials. *J Clin Oncol*. 2025;43(16_suppl):e21018. doi:10.1200/JCO.2025.43.16_suppl.e21018.
11. Morse B, Jeong D, Ihnat G, Silva AC. Pearls and pitfalls of response evaluation criteria in solid tumors (RECIST) v1.1 non-target lesion assessment. *Abdom Radiol (NY)*. 2019 Feb;44(2):766–774. doi: 10.1007/s00261-018-1752-4. PMID: 30196362; PMCID: PMC7771310.
12. Teslenko I, Belotserkovskiy M, Kumar A. Common pitfalls of RECIST 1.1 application in clinical trials. *Journal for Clinical Studies*. 2015;7 (3):26–32.

Maxim Kosov

Maxim Kosov, MD, PhD, is Senior Medical Advisor at PSI CRO AG (USA). Trained in paediatrics, anesthesiology and intensive care, he has more than 30 years of combined medical and research experience, including over 20 years in the clinical trials industry. His expertise spans a broad range of therapeutic areas, and he has authored or co-authored more than 60 scientific publications.

Email: maxim.kosov@psi-cro.com

