Supplementary Material

Supplementary Methods

Literature search criteria

Two investigators (EF and CR) searched four electronic databases (PubMed/MEDLINE, Web of Science, Cochrane Library, Google Scholar) for published articles and abstracts from international meetings from retrospective, prospective and randomised clinical trials investigating organ preservation for rectal cancer, published from inception to 1 April 2020. The search criteria to identify the key outcome measures were as follows: MeSH terms or combined free terms included "rectal cancer" (or "rectal carcinoma") AND "organ preservation" OR "non-operative management (NOM)" OR "watch and wait (W&W)". Additional search terms included "clinical complete response (cCR)", "trial", "retrospective", "prospective", "randomised", "phase 1/2/3" (or phase I/II/III)", "survival", "radiotherapy", "short-course radiotherapy", "chemoradiation (or "chemoradiotherapy" or "radiochemotherapy"), "preoperative", "neoadjuvant", "quality of life (QoL)" and "patient reported outcome (PRO)" or "patient reported outcome measures (PROMs), also with a combination of citation tracking. Unpublished ongoing clinical studies using either the NCT (https://clinicaltrials.gov), the ISRCTN (https://www.isrctn.com) or EudraCT (https://eudract.ema.europa.eu) trail registries were searched as well. Studies with less than ten patients were excluded. We also excluded titles if considered irrelevant to the scope of this review or duplicate publications. In case of duplicate publications, the most recent and informative study was considered. Only papers published in English were reviewed.

Based on their relevance to their present topic and their frequency of reporting, seven key outcome measures were identified in 667 abstracts. The key outcome measures were as follows: 1. Criteria, definition and nomenclature of clinical endpoints 2. Choice of primary endpoint according to the trial phase and design 3. Timepoint of early response assessment to determine complete clinical response 4. Response-based decision and use of biopsy 5. Follow-up procedures and schedule 6. Anorectal function measurement and 7. Quality of life assessment and patient reported outcome measurements. Of those 667 abstracts, and upon full-text assessment, 271 were subsequently excluded due to various reasons (e.g. articles related to preservation of anal sphincter or other organs, such as urinary bladder instead of topics dedicated to organ preservation in rectal cancer; articles referring to non-operative management of postoperative complications due to various reasons and in different clinical settings instead of non-operative management as part of organ preservation in rectal cancer; articles related to

diverse biomarker studies outside the scope of the present study; non-English articles). The flowchart of article selection process is shown in **Figure 1**.

Formation of consensus panel and Delphi method to establish a consensus

All co-authors of the recent review[1] were contacted by EF and CR to invite them to participate. The co-authors were emailed a preliminary survey questionnaire, as well as a request for recommendations for inclusion more clinical experts on the panel and also to include additional studies or outcome measures of organ preservation.

The 1st Delphi round (R1) questionnaire addressing key questions regarding NOM as well as LE strategies following (C)RT was prepared by EF, CR, AA, AG and DSM and emailed to 31 rectal cancer experts including a link to the online questionnaire from the SurveyMonkey program. The panelists were emailed a preliminary survey questionnaire, as well as a request for recommendations for inclusion more clinical experts on the panel and also to include additional studies or outcome measures of organ preservation. Of the 31, 28 (90%) agreed to participate and complete the questionnaire. Also, two experts on clinical trial outcome measures (MB) and QoL and PRO assessment (AG) were involved (but did not vote), and, hence, the consensus panel consisted of 30 members. The final guideline panel comprised a multidisciplinary and interprofessional team, including clinical oncologists, radiation oncologists, medical oncologists, surgical oncologists, pathologists, radiologists with expertise in phase 1-3 trials of rectal cancer as well as bioinformatician. A priori consensus criteria were defined. To reach a consensus on the different outcome measures, a threshold of 70% or more for agreement was required for each item in R1-R3.

In R1, the participants were asked to provide answers on the different outcome measures of organ preservation. Following completion of R1, the same five investigators (EF, CR, AA, AG and DSM) consolidated the responses and prepared the 2nd Delphi round (R2) questionnaire. This included questions that did not reach a consensus in R1 together with additional outcome measures of organ preservation following participant's feedback. Questions that received less than 70% of the votes in R1 were modified accordingly and entered into R2 for voting. Questions that failed to reach consensus and were considered non-essential after R1 were excluded from further analysis.

In R2, the summarised results and individual item histograms of percentage responses from R1 were emailed to all participants, together with the R2 questionnaire. Tables with relevant clinical trials and instructions for voting in R2 were also provided.

In R3, the final consensus recommendations in the manuscript were prepared and assessed by all members of the panel. The summarised results with individual participant's answers of survey rounds were anonymised and exported in Excel format from the SurveyMonkey program (https://www.surveymonkey.com), and distributed to all participants.

Supplementary Results

Consensus procedure and rounds

The detailed questionnaires of the 1st and 2nd Delphi rounds (R1 and R2) on the 7 key outcome measures of organ preservation together with the corresponding answers are provided in **Supplementary Tables 1-2**, respectively. The answers to questions that reached consensus in R1 and R2 together with the corresponding results (% of votes) have been marked with bold and underlined in **Supplementary Tables 1-2**. The flow diagram of the study procedures including R1 and R2 by the expert panel to establish an international consensus is shown in **Figure 2**.

In R1, 18 of 32 (56%) questions received at least 70% of votes and were therefore consented. The 14 of 32 questions where no consensus was established in R1 regarded outcome measures such as the definition of endpoints, the choice of primary endpoint according to the trial phase and design, the timepoint of RA, follow-up methods and QoL assessment (**Supplementary Table 3**). Also, 2 of the 14 questions (questions number 12 and 19 of R1) where no consensus was established in the R1 were considered irrelevant for further voting and were, hence, not included in R2.

To address the items where no consensus was reached in the R1, 12 updated questions were prepared, also considering the feedback of the expert panel, and were entered into R2 (**Supplementary Table 2**). During R2, 11 of the 12 questions (92%) were voted by at least 70% of the panel and, hence, consented. The question that did not reach consensus regarded the definition of TME-free survival (question 2 of R2; received 52% of votes), as described in the main text (**Results**). Please, note that some questions received a consensus (70% threshold) following addition of individual answers that have been provided under the answer option "Other", also with clarification via email correspondence when needed. In R3, the final consensus recommendation manuscript was prepared, evaluated and agreed upon by all members (100%) of the panel.

Supplementary Table 1. List of all 1st DELPHI round (R1) questions and corresponding answer options on the different outcome measures of organ preservation strategy in rectal cancer. The answers to questions that reached consensus in the R1 together with the corresponding results (% of votes) have been marked with bold and underlined.

Outcome measures of organ preservation strategy	**Questions and answer options of the 1st Delphi questionnaire**
Definition of endpoints	1. Do you agree with the following definition of organ preservation as endpoint: Rectum intact, without non-salvageable (i.e. LE with R0 resection) locally progressive disease, no stoma. • **Yes (75%)** • No • Other (please specify) 2. Which criteria are the most appropriate to define cCR as endpoint (see Table 1 provided with the survey)? • Sao Paulo (Habr-Gama et al. Ann Surg 2019) • Amsterdam/Maastricht (Maas et al. JCO 2011) • MSKCC (Smith et al. JAMA Oncol 2019) • ESMO guidelines (Glynne-Jones et al. Ann Oncol 2017) • NCCN guidelines 2020 • Other (please specify) 3. Do you agree with the following definition: Locoregional regrowth refers to an event involving either the bowel wall, mesorectum and/or pelvic organs that occurs after WW. • **Yes (78%)** • No • Other (please specify) 4. Do you agree with the following definition: Locoregional recurrence refers to an event involving either the bowel wall, mesorectum and/or pelvic organs that occurs after LE or TME. • **Yes (86%)** • No • Other (please specify) 5. Do you agree with the following definition: Locoregional/pelvic failure refers to non-curatively (no R0) salvageable locoregional regrowth after WW, or to locoregional recurrence that occurs after LE/TME. • **Yes (78%)** • No • Other (please specify)
Choice of primary endpoint according to the trial phase and design	6. Which primary endpoint do you consider the most relevant for non-operative management (NOM) as part of an organ preservation prospective trial strategy? (Note: time points of assessment are addressed below) • Clinical complete response (cCR) • Sustained cCR • Magnetic-resonance Tumour Regression Grading (mrTRG) • Organ preservation (i.e. rectum intact, no major surgery, no stoma) • TME-free survival • Any local regrowth after initial cCR • Local regrowth not amenable to curative (R0) salvage (by LE or TME) • Locoregional recurrence (after curative salvage operation of local regrowth)

	• DFS (including any local regrowth as event) • DFS (salvageable local regrowth censored/not regarded as event) • Functional anorectal score (to be defined) • Composite endpoint combining some of the above • Other (please specify) 7. Which primary endpoint is the most appropriate after local excision (LE) for organ preservation strategy? • LE with R0 versus R1/R2 • LE with ypT0/1/2 versus 3 or ypT0-1 versus 2-3 • Completion-TME required versus not • Organ preservation (i.e rectum intact, only LE, no stoma) • TME-free survival • Locoregional recurrence after LE or Completion-TME • GRECCAR2 endpoint (composite outcome of death, recurrence, morbidity, and side-effects at 2 years after surgery) • DFS • Functional anorectal score (to be defined), PRO • A combination of above (please suggest selection in writing) • Other (please specify) 8. If the primary endpoint for a phase 2 trial should be different to the primary endpoint of a phase 3 trial of NOM/LE, please enter your preferences below: • Primary endpoint for NOM in phase 2 trial: • Primary endpoint for LE in phase 2 trial: • Primary endpoint for NOM in phase 3 trial: • Primary endpoint for LE in phase 3 trial: • Other (please specify):
Timepoint of response assessment	9. When should we measure the outcome of NOM/LE as part of an organ preservation prospective clinical trial strategy? • **<u>From start of treatment (75%)</u>** • From completion of treatment 10. For the endpoint cCR after CRT/SCRT (without consolidation chemotherapy), which time point should be used to define it after start of treatment? • 4 weeks (1 month) • 6 weeks (1.5 month) • 8 weeks (2 months) • 10 weeks (2.5 months) • 12 weeks (3 months) • Other (please specify) 11. For the endpoint cCR after SCRT/CRT followed by consolidation chemotherapy, which time point should be used to define it after the start of treatment? • 4 weeks (1 month) • 6 weeks (1.5 month) • 8 weeks (2 months) • 10 weeks (2.5 months) • 12 weeks (3 months) • Other (please specify)

	12. For the endpoint of sustained cCR after initial cCR, which time point should be used for statistical case number calculations from the start of treatment? • 12 months • 18 months • 24 months • 30 months • 36 months • Other (please specify) 13. For the endpoint organ preservation/TME-free survival, which time point from the start of treatment should be used to define it? • 6 months • 12 months • 18 months • 24 months • **30 months** • **36 months (71-75%)** • Other (please specify) 14. For the endpoint of locoregional recurrence after curative salvage operation of local regrowth, which time point from salvage operation should be used to define it? • 12 months • 24 months • **36 months (71%)** • Other (please specify) 15. For the endpoint of DFS, which time point from the end (or start) of treatment should be used to define it? • 24 months • **36 months (75%)** • 60 months • Other (please specify)
Response-based decision including the use of biopsy	16. If restaging shows near cCR (or, alternatively, mrTRG1-2), would you consider it acceptable to wait longer before decision on salvage surgery? • **Yes (93%)** • No • Other (please specify) 17. If the answer to the above question is "Yes", which time point after the first assessment would you use for re-assessment of tumour response? • 4 weeks (1 month) • 6 weeks (1.5 month) • 8 weeks (2 months) • **6-12 weeks (3 months) (79%)** • Other (please specify) 18. If endoscopy shows near cCR, should a biopsy be taken? • **No, because (provide reason as comment) (82%):** • Yes, because (provide reason as comment): • Other (please specify): 19. If the biopsy from a patient with near cCR is negative, what would you do?

	• Repeat biopsy • LE • Re-assess after 3 months • Other (please specify) 20. If endoscopy shows cCR (sustained cCR), should a biopsy be taken? • **No, because (provide reason as comment) (89%):** • Yes, because (provide reason as comment): • Other (please specify):
Follow-up methods	21. Which method(s) should be used as part of the follow-up in NOM (or LE) of rectal cancer (more than one option possible)? • **CEA (71%)** • **Digital rectal examination (89%)** • **Endoscopy (93%)** • **MRI pelvis including DWI (93%)** • **CT chest/abdomen (79%)** • Other (please specify) 22. Which of the following time schedules would you prefer to measure CEA? • year 1: monthly; year 2: bimonthly; year 3: every 6 months • **years 1-3: every 3 months; years 4-5: every 6 months (71%)** • years 1-2: every 6 months • years 1-2: every 3 months • Other (please specify) 23. Which of the following time schedules would you prefer to perform DRE? • year 1: monthly; year 2: bimonthly; year 3: every 6 months • year 1: every 3 months; years 2-5: every 6 months • years 1-2: every 4-6 months • years 1-2: every 3 months • year 1: every 3 months; year 2: every 4 months; year 3-5: every 6 months • Other (please specify) 24. Which of the following time schedules would you prefer to perform endoscopy? • year 1: monthly; year 2: bimonthly; year 3: every 6 months • years 1: every 3 months; years 2-5: every 6 months • years 1-2: every 4-6 months • years 1-2: every 3 months • year 1: every 3 months; year 2: every 4 months; year 3-5: every 6 months • Other (please specify) 25. Which of the following time schedules would you prefer to perform MRI of the pelvis (including DWI)? • year 1-3: every 6 months • years 1: every 3 months; years 2-5: every 6 months • years 1-2: every 4-6 months • years 1-3: every 6 months • year 1: every 3 months; year 2: every 4 months; year 3-5: every 6 months • Other (please specify)

	26. Which of the following time schedules would you prefer to perform CT of the chest and abdomen? • year 1-2: every 6 months • years 1-3: every 6 months • years 1: every 6 months; years 2-5: annually • year 1-3: annually • Other (please specify)
Anorectal function tests	27. Which anorectal function score(s) should we use in patients treated with organ preservation strategy? (Note: more than one option is possible. The scores and relevant references are provided together with the survey link) • The Wexner score • **The Low Anterior Resection Syndrome (LARS) score (71%)** • The MSKCC Bowel Function Instrument (MSKCC BFI) score • The Vaizey score • Manometry • Need a new anorectal function score specific for NOM/LE • Other (please specify)
QoL assessment and PROMs	28. Which Quality of Life Questionnaire (QLQ) should we use in addition to QLQ-C30 in patients treated with organ preservation strategies? (Note: more than one option is possible) (Link to https://qol.eortc.org/questionnaires/ and https://euroqol.org/support/how-to-obtain-eq-5d for reference) • QLQ-CR29 • QLQ-CR38 • QLQ-ANL27 • EQ-5Q-5D • Need a new QLQ specific for NOM/LE • Other (please specify) 29. Select the top 10 symptomatic toxicity items that you think are most relevant in organ preservation; where 1 is the most relevant, 2 the second most relevant, 3 the third most relevant and so on. • **Bowel urgency** • **Fecal incontinence** • **Bowel frequency** • **Diarrhea** • **Tenesmus** • **Toilet dependency** • **Night time bowel opening** • Urinary frequency • **Urinary urgency** • Urinary incontinence • Nocturia • **Impotence** • Altered ejaculation • Vaginal dryness • Vaginal stenosis • Libido • Fatigue • **Pain** • Chemotherapy toxicities: e.g. nausea, vomiting, PPE, neuropathy, loss of appetite, • Other (please specify)

	30. What are the key QOL and function issues we should record in patients treated with organ preservation strategy? (Note: more than one option is possible) • **Overall quality of life (85%)** • **Physical function (e.g. ability to self-care, carry out ADLs) (93%)** • **Role function (e.g. ability to go to work, continue hobbies) (85%)** • **Social function (e.g. impact on your social life and/or family life) (96%)** • **Emotional function (e.g. impact on mood) (70%)** • Other (please specify) 31. What are the optimal timings for measurement of symptomatic toxicity to observe differences between treatment schedules? (Note: more than one option is possible) • **Baseline (85%)** • Final week of treatment • Month 1 • Month 2 • **Month 3 (71%)** • Month 6 • Month 9 • **Month 12 (89%)** • Month 15 • Month 18 • **Month 24 (71%)** • Month 30 • **Month 36 (71%)** • Month 42 • Month 48 • Month 54 • **Month 60 (71%)** • Other (please specify) 32. What are the optimal timings for measurement of QOL and function to observe differences between treatment schedules? (Note: more than one option is possible) • **Baseline (89%)** • Final week of treatment • Month 1 • Month 2 • **Month 3 (71%)** • Month 6 • Month 9 • **Month 12 (85%)** • Month 15 • Month 18 • **Month 24 (71%)** • Month 30 • **Month 36 (85%)** • Month 42 • Month 48 • Month 54 • **Month 60 (71%)** • Other (please specify)

Supplementary Table 2. List of 2nd DELPHI round (R2) questions on the different outcome measures of organ preservation strategy in rectal cancer. The answers to questions that reached consensus in R2 together with the corresponding results (% of votes) have been marked with bold and underlined.

Outcome measures of organ preservation strategy	Questions and answer options of the 2nd Delphi questionnaire
Definition of endpoints	1. Do you agree with the following statement: the Amsterdam/Maastricht criteria (Maas et al. JCO 2011) are the most appropriate to define cCR as endpoint. • **Yes (85%)** • No • Other (please specify) $2. Do you agree with the following statement: TME-free survival is defined as time from randomisation to one of the following events: radical TME surgery for non-complete response at re-staging, any locoregional regrowth after initial cCR requiring salvage-TME, any locoregional recurrence after TEM or no-salvageable regrowth (a regrowth that cannot be removed with an R0 resection), distant metastasis or death (all cause), whichever occurs first. • Yes • No • Other (please specify)
Choice of primary endpoint according to the trial phase and design	3. Do you agree with the following statement: cCR rate should be the primary endpoint for early phase 1/2 trials ("pick the winner") aiming to increase cCR rates and enable NOM/LE by more intense RT/CRT/TNT regimen to select tolerable and locally effective treatment regimen for further testing in larger scale trials (e.g. Danish trial: NCT00952926 and CAO/ARO/AIO-16 trial: NCT03561142). • **Yes (89%)** • No • Other (please specify) 4. Do you agree with the following statement: Organ preservation as intermediate endpoint, assessed at 30-36 months after treatment start, should be the primary endpoint for (randomised) phase 2/3 trials using NOM or LE (e.g. WW3, STAR-TREC, ACO/ARO/AIO-18.1 trials). • **Yes (93%)** • No • Other (please specify) 5. Do you agree with the following statement: Organ preservation-adapted DFS (shown in Table at 3 years is an appropriate long-term (primary) endpoint in trials where organ preservation is allowed within but not the primary purpose of a phase 3 trial, especially in locally-advanced tumours. • **Yes (82%)** • No • Other (please specify)
Timepoint of response assessment	6. Do you agree with the following statement: After SCRT/CRT (treatment duration up to 6 weeks) for early-stage tumours, a two-step approach i.e. 12 weeks and 16-20 weeks after start of treatment should be used to determine cCR (e.g. STAR-TREC trial). • **Yes (93%)** • No • Other (please specify)

	*7. Do you agree with the following statement: After CRT and brachytherapy (treatment duration up to 12 weeks), cCR will be determined at 14 weeks after start of treatment (as performed within the OPERA trial). • **Yes (71%)** • No • Other (please specify) 8. Do you agree with the following statement: After TNT with CRT and induction/consolidation chemotherapy (treatment duration 16-20 weeks), cCR will be determined at 24 weeks after start of treatment (e.g. GRECCAR12 and ACO/ARO/AIO-18.1 trials). • **Yes (75%)** • No • Other (please specify) 9. Do you agree with the following statement: After TNT with SCRT/CRT followed by consolidation chemotherapy (total duration 26 and 34 weeks, respectively), cCR will be determined 34-38 weeks after start of treatment (e.g. OPRA and TRIGGER trials, respectively). • **Yes (79%)** • No • Other (please specify)
Follow-up methods	10. Do you agree with the following statement: Regarding follow-up/surveillance, the preferred time schedule to perform DRE, endoscopy and MRI is as follows: year 1-2: every 3-4 months; year 3-5: every 6 months. • **Yes (93%)** • No • Other (please specify) 11. Do you agree with the following statement: The preferred time schedule to perform CT-thorax/abdomen is as follows: year 1: every 6-12 months; year 2-5: every 12 months. • **Yes (71%)** • No • Other (please specify)
QOL assessment and PROMs	12. Do you agree with the following statement: In addition to EORTC QLQ-C30, a new QLQ specific for organ preservation strategies is needed, which will be developed in the near future in collaboration with EORTC Quality of Life Group. • **Yes (100%)** • No • Other (please specify)

[$]Consensus was not reached for this question as there was confusion regarding the precise name of the endpoint of the OPRA trial (TME-free survival instead of TME-free disease-free survival, which was the correct name) when the question was initially conceived. Of note, the definition of the endpoint TME-free disease-free survival was provided separately by the primary investigator of the OPRA trial, JGA (shown in Table 1).

*In R2, six participants requested the use of a second time point for determining cCR to allow consensus to be reached

Supplementary Table 3. List of questions with lack of consensus at the 1st DELPHI round (R1) of organ preservation in rectal cancer survey

Outcome measures of organ preservation strategy	**Questions and answer options of the 1st Delphi questionnaire with lack of consensus in the 1st round***
Definition of endpoints	2. Which criteria are the most appropriate to define cCR as endpoint (see Table 1 provided with the survey)? • Sao Paulo (Habr-Gama et al. Ann Surg 2019) • Amsterdam/Maastricht (Maas et al. JCO 2011) • MSKCC (Smith et al. JAMA Oncol 2019) • ESMO guidelines (Glynne-Jones et al. Ann Oncol 2017) • NCCN guidelines 2020 • Other (please specify)
Choice of primary endpoint according to the trial phase and design	6. Which primary endpoint do you consider the most relevant for non-operative management (NOM) as part of an organ preservation prospective trial strategy? (Note: time points of assessment are addressed below) • Clinical complete response (cCR) • Sustained cCR • Magnetic-resonance Tumour Regression Grading (mrTRG) • Organ preservation (i.e. rectum intact, no major surgery, no stoma) • TME-free survival • Any local regrowth after initial cCR • Local regrowth not amenable to curative (R0) salvage (by LE or TME) • Locoregional recurrence (after curative salvage operation of local regrowth) • DFS (including any local regrowth as event) • DFS (salvageable local regrowth censored/not regarded as event) • Functional anorectal score (to be defined) • Composite endpoint combining some of the above • Other (please specify) 7. Which primary endpoint is the most appropriate after local excision (LE) for organ preservation strategy? • LE with R0 versus R1/R2 • LE with ypT0/1/2 versus 3 or ypT0-1 versus 2-3 • Completion-TME required versus not • Organ preservation (i.e rectum intact, only LE, no stoma) • TME-free survival • Locoregional recurrence after LE or Completion-TME • GRECCAR2 endpoint (composite outcome of death, recurrence, morbidity, and side-effects at 2 years after surgery) • DFS • Functional anorectal score (to be defined), PRO • A combination of above (please suggest selection in writing) • Other (please specify) 8. If the primary endpoint for a phase 2 trial should be different to the primary endpoint of a phase 3 trial of NOM/LE, please enter your preferences below: • Primary endpoint for NOM in phase 2 trial: • Primary endpoint for LE in phase 2 trial: • Primary endpoint for NOM in phase 3 trial: • Primary endpoint for LE in phase 3 trial:

	• Other (please specify):
Timepoint of response assessment	10. For the endpoint cCR after CRT/SCRT (<u>without</u> consolidation chemotherapy), which time point should be used to define it after start of treatment? • 4 weeks (1 month) • 6 weeks (1.5 month) • 8 weeks (2 months) • 10 weeks (2.5 months) • 12 weeks (3 months) • Other (please specify) 11. For the endpoint cCR after SCRT/CRT <u>followed by consolidation</u> chemotherapy, which time point should be used to define it after the start of treatment? • 4 weeks (1 month) • 6 weeks (1.5 month) • 8 weeks (2 months) • 10 weeks (2.5 months) • 12 weeks (3 months) • Other (please specify)
Response-based decision including the use of biopsy	17. If the answer to the above question is "Yes", which time point after the first assessment would you use for re-assessment of tumour response? • 4 weeks (1 month) • 6 weeks (1.5 month) • 8 weeks (2 months) • 12 weeks (3 months) • Other (please specify)
Follow-up methods	23. Which of the following time schedules would you prefer to perform DRE? • year 1: monthly; year 2: bimonthly; year 3: every 6 months • year 1: every 3 months; years 2-5: every 6 months • years 1-2: every 4-6 months • years 1-2: every 3 months • year 1: every 3 months; year 2: every 4 months; year 3-5: every 6 months • Other (please specify) 24. Which of the following time schedules would you prefer to perform endoscopy? • year 1: monthly; year 2: bimonthly; year 3: every 6 months • years 1: every 3 months; years 2-5: every 6 months • years 1-2: every 4-6 months • years 1-2: every 3 months • year 1: every 3 months; year 2: every 4 months; year 3-5: every 6 months • Other (please specify) 25. Which of the following time schedules would you prefer to perform MRI of the pelvis (including DWI)? • year 1-3: every 6 months • years 1: every 3 months; years 2-5: every 6 months • years 1-2: every 4-6 months • years 1-3: every 6 months • year 1: every 3 months; year 2: every 4 months; year 3-5: every 6 months

	• Other (please specify) 26. Which of the following time schedules would you prefer to perform CT of the chest and abdomen? • year 1-2: every 6 months • years 1-3: every 6 months • years 1: every 6 months; years 2-5: annually • year 1-3: annually • Other (please specify)
QoL assessment	28. Which Quality of Life Questionnaire (QLQ) should we use in addition to QLQ-C30 in patients treated with organ preservation strategy? (Note: more than one option is possible) (Link to https://qol.eortc.org/questionnaires/ and https://euroqol.org/support/how-to-obtain-eq-5d for reference) • QLQ-CR29 • QLQ-CR38 • QLQ-ANL27 • EQ-5Q-5D • Need a new QLQ specific for NOM/LE • Other (please specify)

*Consensus was not reached in 12 of the total 32 questions at R1. Also, 2 of the 14 questions (questions number 12 and 19 of R1) where no consensus was established in the R1 were considered irrelevant for further voting and were, hence, not included in R2.

Supplementary Table 4. Summary of the different definitions of cCR for organ preservation strategy

Definition of cCR	Timepoint after treatment	Diagnostic methods and criteria
Sao Paulo[2]	Ca. 8 weeks	DRE and Rectoscopy: no palpable tumour; no residual ulcer, mass or stenosis of the rectum (whitening of the mucosa, teleangiectasias, and subtle loss of pliability of the rectum were considered to be consistent with cCR). Endoscopic biopsy: not routinely performed. MRI: no suspicious mesorectal enlarged, irregular bordered and heterogeneous nodes, and in the presence of findings of fibrotic changes within the rectal (low signal intensity areas with or without submucosal hypertrophy).
Amsterdam /Maastricht[3]	6-8 weeks	DRE and Rectoscopy: no palpable tumour, no residual tumour or only a small residual erythematous ulcer or scar; Endoscopic biopsy: negative biopsies from the scar, ulcer, or former tumour location (not mandatory) MRI: Substantial downsizing with no residual tumour or residual fibrosis only (with low signal on diffusion-weighted imaging MRI), sometimes associated with residual wall thickening due to edema; 2) no suspicious lymph nodes
MSKCC[4]	Not provided	DRE and Rectoscopy: no palpable tumour; a flat white scar with or without telangiectasias and lack of ulceration or nodularity Endoscopic biopsy: no information provided MRI: was not reported as MRI protocol was implemented in 2013
ESMO Guidelines 2017[5]	12 weeks after start of neoadjuvant treatment	DRE and Rectoscopy: no palpable tumour or irregularity; no visible lesion except a flat scar, telangiectasia or whitening of the mucosa; Endoscopic biopsy: negative biopsies from the scar MRI/US: absence of any residual tumour in the primary site and draining lymph nodes
NCCN Guidelines 2020	Not provided	DRE and Rectoscopy: no palpable tumour; no residual tumour Endoscopic biopsy: no information provided MRI: no residual tumour in the primary site and draining lymph nodes

Supplementary Table 5. Anorectal function scores relevant to question Nr. 27 of the 1st DELPHI round (R1)

Anorectal function score
• The Wexner score[6] • The Low Anterior Resection Syndrome (LARS) score[7] • The MSKCC Bowel Function Instrument (MSKCC BFI)[8] • The Vaizey score[9] • Manometry

SUPPLEMENTARY REFERENCES

1. Fokas E, Glynne-Jones R, Appelt A, et al. Outcome measures in multimodal rectal cancer trials. *Lancet Oncol* 2020; **21**(5): e252-e64.
2. Habr-Gama A, Sao Juliao GP, Vailati BB, et al. Organ Preservation in cT2N0 Rectal Cancer After Neoadjuvant Chemoradiation Therapy: The Impact of Radiation Therapy Dose-escalation and Consolidation Chemotherapy. *Ann Surg* 2019; **269**(1): 102-7.
3. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011; **29**(35): 4633-40.
4. Smith JJ, Strombom P, Chow OS, et al. Assessment of a Watch-and-Wait Strategy for Rectal Cancer in Patients With a Complete Response After Neoadjuvant Therapy. *JAMA oncology* 2019: e185896.
5. Glynne-Jones R, Wyrwicz L, Tiret E, et al. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017; **28**(suppl_4): iv22-iv40.
6. Jorge JM, Wexner SD. Etiology and management of fecal incontinence. *Dis Colon Rectum* 1993; **36**(1): 77-97.
7. Emmertsen KJ, Laurberg S. Low anterior resection syndrome score: development and validation of a symptom-based scoring system for bowel dysfunction after low anterior resection for rectal cancer. *Ann Surg* 2012; **255**(5): 922-8.
8. Temple LK, Bacik J, Savatta SG, et al. The development of a validated instrument to evaluate bowel function after sphincter-preserving surgery for rectal cancer. *Dis Colon Rectum* 2005; **48**(7): 1353-65.
9. Vaizey CJ, Carapeti E, Cahill JA, Kamm MA. Prospective comparison of faecal incontinence grading systems. *Gut* 1999; **44**(1): 77-80.