

Amsterdam International Consensus Meeting: Tumor Response Scoring in the Pathology Assessment of Resected Pancreatic Cancer after Neoadjuvant Therapy

Boris V. Janssen, BSc^{1,2*}, Faik Tutucu, MD^{1,3*}, Stijn van Roessel, MD, MSc², Volkan Adsay, MD³, Olca Basturk, MD⁴, Fiona Campbell, MD, FRCPath⁵, Claudio Doglioni, MD⁶, Irene Esposito, MD⁷, Roger Feakins, MB, BCh, BAO, BA, MD, FRCPI, FRCPath⁸, Noriyoshi Fukushima, MD, PhD⁹, Anthony J Gill, MD, FRCPA¹⁰, Ralph H. Hruban, MD¹¹, Jeffrey Kaplan, MD¹², Bas Groot Koerkamp, MD, PhD¹³, Seung-Mo Hong, MD, PhD¹⁴, Alyssa Krasinskas, MD¹⁵, Claudio Luchini, MD, PhD¹⁶, Johan Offerhaus, MD, PhD¹⁷, Arantza Fariña Sarasqueta, MD, PhD¹, Chanjuan Shi, MD, PhD¹⁸, Aatur Singhi, MD, PhD¹⁹, Thomas F. Stoop, BSc^{2,20}, Eline C. Soer, MD¹, Elizabeth Thompson, MD, PhD¹¹, Geertjan van Tienhoven, MD, PhD²¹, Marie-Louise F. Velthuysen, MD, PhD²², Johanna W. Wilmink, MD, PhD²³, Marc G. Besselink, MD, MSc, PhD^{2§}, Lodewijk A.A. Brosens, MD, PhD^{17,24§}, Huamin Wang MD, PhD^{25§}, Caroline S. Verbeke, MD, PhD^{26§}, Joanne Verheij, MD, PhD^{1§} for the International Study Group of Pancreatic Pathologists (ISGPP)

* These authors share first authorship

§ These authors share senior authorship

Running title:

Amsterdam International Consensus Meeting

Affiliations:

¹Department of Pathology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

²Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

³Department of Pathology, Koc University and KUTTAM Research Center, Istanbul, Turkey.

⁴Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁵Department of Pathology, Royal Liverpool University Hospital, Liverpool, UK.

⁶Department of Pathology, IRCCS San Raffaele Scientific Institute, Milan, Italy.

⁷Institute of Pathology, Heinrich-Heine-University and University Hospital of Duesseldorf, Germany.

⁸Department of Pathology, Royal Free London NHS Trust, London, UK.

⁹Department of Pathology, Jichi Medical University Hospital, Tochigi, Japan.

¹⁰Cancer Diagnosis and Pathology Group Kolling institute of Medical Research Royal North Shore Hospital St Leonards NSW Australia and University of Sydney, Sydney NSW Australia.

¹¹Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

¹²Department of Pathology, University of Colorado Hospital, Denver, CO, USA.

¹³Department of Surgery, Erasmus Medical Center, Rotterdam, The Netherlands.

- ¹⁴Department of Pathology, Asan Medical Center, Seoul, Republic of Korea.
- ¹⁵Department of Pathology, Emory University, Atlanta, GA, USA.
- ¹⁶Department of Diagnostics and Public Health, University and Hospital Trust of Verona, Verona, Italy.
- ¹⁷Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands.
- ¹⁸Department of Pathology, Duke University Medical Center, Durham, NC, USA.
- ¹⁹Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.
- ²⁰Division of Surgical Oncology, Department of Surgery, University of Colorado, Anschutz Medical Campus, Aurora, Colorado, USA.
- ²¹Department of Radiation Oncology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.
- ²²Department of Pathology, Erasmus Medical Center, Rotterdam, The Netherlands.
- ²³Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.
- ²⁴Department of Pathology, Radboud UMC, Nijmegen, the Netherlands.
- ²⁵Department of Anatomical Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA.
- ²⁶Department of Pathology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

Corresponding author:

J. Verheij, MD, PhD

Amsterdam UMC, University of Amsterdam

Department of Pathology, Cancer Center Amsterdam

P.O. Box 22660, 1105 AZ Amsterdam

Phone: +31 20 566 5650

E-mail: j.verheij@amsterdamumc.nl

ABSTRACT

Histopathologically scoring the response of pancreatic ductal adenocarcinoma (PDAC) to neoadjuvant treatment can guide the selection of adjuvant therapy and improve prognostic stratification. However, several tumor response scoring (TRS) systems exist, and consensus is lacking as to which system represents best practice. An international consensus meeting on TRS took place in November 2019 in Amsterdam, The Netherlands. Here, we provide an overview of the outcomes and consensus statements that originated from this meeting. Consensus ($\geq 80\%$ agreement) was reached on a total of 7 statements: 1) TRS is important because it provides information about the effect of neoadjuvant treatment that is not provided by other histopathology-based descriptors. 2) TRS for resected PDAC following neoadjuvant therapy should assess residual (viable) tumor burden instead of tumor regression. 3) The CAP scoring system is considered the most adequate scoring system to date because it is based on the presence and amount of residual cancer cells instead of tumor regression. 4) The defining criteria of the categories in the CAP scoring system should be improved by replacing subjective terms including “minimal” or “extensive” with objective criteria to evaluate the extent of viable tumor. 5) The improved, consensus-based system should be validated retrospectively and prospectively. 6) Prospective studies should determine the extent of tissue sampling that is required to ensure adequate assessment of the residual cancer burden, taking into account the heterogeneity of tumor response. 7) In future scientific publications, the extent of tissue sampling should be described in detail in the “Materials and Methods” section.

INTRODUCTION

Neoadjuvant therapy is increasingly used for the treatment of patients with resectable, borderline resectable, and locally advanced pancreatic ductal adenocarcinoma (PDAC), as it may improve the margin-negative resection (R0) rate, disease-free survival (DFS), and overall survival (OS).¹⁻⁵ Histologic examination of pancreatic cancer resection specimens following neoadjuvant therapy offers the opportunity to assess the effect of treatment at the tissue and cellular levels. This histologic assessment can serve two purposes. First, in the context of randomized controlled trials, comparison of the histologic tumor response allows evaluation of the effectiveness of different neoadjuvant treatment regimens. Second, for the individual patient, the histologic tumor response reflects the sensitivity or resistance of the patient's cancer cells to the neoadjuvant treatment, which may in turn help with the selection of adjuvant treatment. For both purposes, tumor response scoring (TRS) systems must be reproducible and must correlate with patient outcome.

Several TRS systems for PDAC following neoadjuvant therapy have been proposed.⁶⁻¹⁷ However, there is no international agreement on which system represents best practice. The difficulties in reaching consensus regarding the ideal TRS system are multifold. First, the extent of tissue sampling, which is key to compensating for intratumor heterogeneity of the tumor response (as detailed in Figure 1), is often not specified but probably varies between published studies. Second, interobserver agreement is poor, which is probably related to the difficulty in applying the criteria that define the different categories in the TRS systems.^{17, 18} For example, criteria for the recognition of histologic features such as "rare small groups of cancer cells" are subjective and may contribute to interobserver variability. Third, existing TRS systems differ in the number of distinct categories and thereby vary in discriminative prognostic power and in ease of application.^{19, 20} In view of these unresolved issues, it is not surprising that comparative studies on the predictive performance of the proposed scoring systems are rare and provide limited evidence as to which TRS system performs best with regard to the trade-off between prognostic accuracy, interobserver agreement, and applicability in daily practice.²¹

In an attempt to identify, discuss, and overcome the challenges encountered in the assessment of pathologic tumor response following neoadjuvant therapy in resected pancreatic cancer, an

international consensus meeting was organized. This article provides an overview of the considerations, outcomes, and consensus statements that originated from this meeting.

MATERIALS AND METHODS

The consensus meeting took place on November 22, 2019 at Schiphol Airport, Amsterdam, the Netherlands and was attended by 18 of the 23 invited expert pancreatic pathologists from 9 countries including 4 continents: North America (United States of America [USA]), Europe (Italy, the Netherlands, Norway, Turkey, and the United Kingdom [UK]), Oceania (Australia), and Asia (Japan and Korea). Also invited and attending were a medical oncologist, two surgeons, a radiation oncologist, two experts in artificial intelligence, and five junior researchers. In total, 29 clinicians and researchers attended the meeting, all on invitation. Supplementary Table S1 lists all attendees, areas of expertise, and affiliations. A pre-meeting survey was conducted by S.v.R., A.F.S., C.S.V., L.A.B., J.V., and the consensus meeting was organized and (co-)chaired by S.v.R., M.G.B., J.V., and L.A.B. During the meeting the participants decided to form the International Study Group of Pancreatic Pathologists (ISGPP).

Pre-meeting survey

The **digital** pre-meeting survey was conducted among all invited pathologists (n=23) to obtain a baseline impression of the areas of agreement and disagreement, using **Google Forms**. The survey consisted of 21 **statements** with additional comment sections, and one open question (Table 1). All **statements** were related to the assessment of tumor response following neoadjuvant chemo(radio)therapy in pancreatic cancer: i) the clinical significance of TRS, ii) histopathologic technique/sampling, and iii) various TRS scoring systems. **Participants were asked to state whether they agreed or disagreed with the given statements and were given the opportunity to expand on their agree/disagree answer, or to state uncertainty, in form of a supplemental free-text answer. When participants provided only the latter, the free-text answers were assigned to the agree/disagree/other categories. In case this was unclear, participants were contacted. A sub-analysis was performed to provide insight into differences in opinion between pathologists from different continents.**

Proceeding of the meeting

The pre-meeting survey outcomes were presented at the beginning of the meeting and served as a starting point for further discussion. In addition, a summary of the current literature on TRS in PDAC (S.v.R.) and issues of controversy in the histopathologic assessment of tumor regression in PDAC

were presented (C.S.V.). Eventually, following constructive discussion, consensus was reached on some topics. The topics on which there was disagreement were identified as requiring future study.

Post-meeting survey

Two months after the meeting, a post-meeting survey, prepared and conducted by S.v.R., A.F.S., M.G.B., L.A.B., H.W., C.S.V., and J.V., was sent to confirm consensus on several statements. The statements in this survey were constructed after the consensus meeting and aimed to represent the outcomes formulated by all participants at the end of the consensus meeting. **Table 2 lists the post-meeting survey statements**. Statements were considered consensus statements when $\geq 80\%$ of respondents agreed.

RESULTS

Pre-meeting survey

All 23 pathologists invited to participate in the pre-meeting survey completed the pre-meeting survey. Eight of the 21 pre-meeting statements reached $\geq 80\%$ agreement. Seven statements had between ≥ 60 and $< 80\%$ agreement. For six statements agreement was $< 60\%$. Table 1 shows the statements and pre-meeting survey outcomes. Supplementary Table S2 provides information on the comparative analysis between continents.

Open question: "What is the main goal and clinical relevance of TRS?"

Overall, respondents regard TRS as a tool that could be clinically relevant in three ways. First, TRS may help predict clinical outcome. Respondents expect the degree of tumor response to correlate with DFS and OS after surgery. Hence, TRS could be used to stratify patients in post-operative clinical trials, and to identify those patients who are most likely to benefit from adjuvant treatment due to a high likelihood of disease recurrence. Second, TRS is viewed as a tool that could potentially guide the choice of adjuvant (chemo)therapy. Little or no tumor response in the resection specimen would indicate that the tumor is resistant to the administered neoadjuvant agent(s). As such, TRS may guide clinicians to use different adjuvant therapies that target biological mechanisms other than those targeted by the neoadjuvant agent/regimen. Third, TRS is viewed as a potential objective parameter in studies comparing the effectiveness of different neoadjuvant regimens. More extensive tumor response in one treatment group could indicate superior treatment effect.

Post-meeting survey and statements

Twenty of the 23 (87%) pathologists invited to participate in the post-meeting survey completed the post-meeting survey. Each of the seven post-meeting statements scored $\geq 80\%$ agreement. Table 2 shows the statements and the exact agreement rates. Here, we provide contextual descriptions of the consensus statements and their underlying arguments.

1) TRS is important because it provides information about the effect of neoadjuvant treatment that is not provided by other histopathology-based descriptors.

All pathologists of the ISGPP agreed that TRS is a clinically relevant parameter. Indeed, as the most direct measure of the efficacy of therapy, the tumor response score represents information that is only indirectly provided by other parameters. Furthermore, in addition to other “traditional” histopathologic parameters (such as tumor grade, margin status etc.), TRS, regardless of the scoring system, is widely seen as predictor of clinical outcome.^{10, 22-27} However, current TRS systems do not allow prognostic stratification of the vast majority of patients, in whom neoadjuvant treatment results in less than (near) complete tumor regression.

2) TRS for resected PDAC following neoadjuvant therapy should assess residual (viable) tumor burden instead of tumor regression.

Most TRS systems are based on an evaluation of either the proportion of the cancer cells that remain viable following treatment or the proportion of tumor cells that have been destroyed by therapy. A problematic issue with both approaches is that the denominator, i.e. the tumor burden before therapy, is unknown. Moreover, it is unclear how the residual viable cancer cells should be assessed after therapy; for instance by counting residual cancer cell numbers or by measuring the size of the foci of residual cancer, and how this can be done in daily practice. Comparison with the original tumor size measured on imaging is inadequate, because tumor size measurements based on gross pathology and radiology often yield divergent results, even when no treatment has been given. Some TRS systems require determination of the amount of residual viable cancer cells in relation to the treatment-induced fibrosis.²⁸ However, fibrosis, for reasons other than neoadjuvant treatment (i.e. concurrent pancreatitis, obstructive changes of the surrounding parenchyma and/or extensive stromal reaction inherent to pancreatic cancer, i.e. desmoplasia) is likely to be histologically indistinguishable from fibrosis secondary to tumor regression (as detailed in Figure 1). To circumvent the inherent problems related to estimating tumor regression, the ISGPP recommends the use of scoring systems that assess residual (viable) tumor burden only.

3) The CAP (College of American Pathologists) scoring system is considered the most adequate scoring system to date because it is based on the presence and amount of residual cancer cells instead of tumor regression.

Multiple TRS systems have been proposed in the past decades.⁴⁻¹⁵ Of these, the CAP, MDACC (MD Anderson), and Evans systems have been used and investigated most frequently (Table 3).^{4,6,8} The 2014 CAP system is a 4-tiered descriptive system based on the amount of residual cancer remaining after therapy, and is adopted from a modified Ryan scheme originally proposed for neoadjuvantly treated rectal cancer.²⁹ The Evans system is 5-tiered and based on the percentage of destroyed tumor. The MDACC system is 3-tiered and based on the percentage of the remaining tumor. The MDACC system originated from a modification to the CAP system that merges CAP Grade 2 and 3 into one category. This was done because PDAC patients with a CAP grade 2 response and those with a grade 3 response had a comparable prognosis.^{10, 22} Although the MDACC system is easy to use and is predictive for patient survival and prognosis, the ISGPP prefers the CAP TRS system because a) it is based on the amount of residual cancer cells, instead of the measurement of percentage-based tumor regression and b) the MDACC system classifies the majority (more than 80%) of PDAC patients as Grade 2 response (poor performer, >5% residual cancer). The ISGPP considered that provision of stratification/categories additional to those of the MDACC Grade 2 group would be more informative. Therefore, the majority of the ISGPP considers the CAP system the most informative to date (90% agreement).

4) The defining criteria of the categories in the CAP scoring system should be improved by replacing subjective terms including “minimal” or “extensive” with objective criteria to evaluate the extent of viable tumor.

Although the CAP system is endorsed by the ISGPP as the most informative scoring system to date, it lacks clear definitions of each grade in terms of microscopic findings. Ideally, the defining criteria of the categories in the CAP scoring system should be improved by replacing subjective terms such as “minimal” or “extensive” with objective criteria to evaluate the extent of viable tumor. For example, the definition of score 1, “single cells or rare small groups of cancer cells”, leaves room for interpretation, especially when the tumor response is heterogeneous throughout the tumor and results in multiple small residual cancer foci. Similar subjective descriptive criteria are also used to define CAP scores 2 and 3. Future studies are necessary to define each category of the current CAP TRS system better and more objectively. For example, strategies may be explored wherein terms as “single cells or rare small groups of cancer cells” could be defined by a maximum diameter, a maximum area, in terms of

(square) millimeters, the absolute number of individual tumor cells in a given area, or by the number of cells per high power field.

5) The improved, consensus-based system should be validated retrospectively and prospectively.

Ideally, validation of a new scoring system should be based on a comparison with other currently available scoring systems in order to test its superiority. Both the inter- and intra-observer agreement and prognostic significance will require evaluation. For validation studies, international collaboration within a multidisciplinary context is highly desirable in order to demonstrate wide applicability.

6) Prospective studies should determine the extent of tissue sampling that is required to ensure adequate assessment of the residual cancer burden, taking into account the heterogeneity of tumor response.

The response of neoplastic cells to neoadjuvant treatment varies within a tumor, and the resulting patchy distribution of residual cancer cells, is a major challenge when scoring tumor response. As such, the measured response to neoadjuvant therapy may differ, depending on the areas sampled and the extent of tumor sampling. While extensive tissue sampling is important to ensure adequate representation of the heterogeneity of treatment effect, the scoring of this heterogeneous, often patchy, process remains challenging (as detailed in Figure 1). Pathologists may either report the poorest response seen in a part of the tumor or estimate the average of the responses observed throughout the entire specimen. The ISGPP hypothesizes that the second approach is most appropriate, as it is likely more representative of the entire tumor, although there is currently no evidence that supports which approach correlates better with patient outcome. As such, new studies are needed to identify the extent of tissue sampling that ensures optimal assessment of the residual cancer burden. Uniform studies can only be achieved by establishing clear rules on the minimum requirements for sampling and reporting. Only then can sufficient evidence be obtained to compare various systems and to draw definitive conclusions. To report complete pathological response reliably (CAP 0, MDACC 0 or Evans 4), we expect that extensive, if not complete, sampling approaches are required. However, complete examination (e.g. embedding the entire pancreas, including all adjacent structures and tissues) is challenging in practice, and its benefits in relation to prognostic significance

are unknown. Further studies are needed to identify the optimal balance between the costs of extensive sampling and the risk of missing clinically important foci of residual cancer.

7) In future scientific publications, the extent of tissue sampling should be described in detail in the "Materials and Methods" section.

While the extent of sampling is a key determinant of the accuracy of TRS, there is currently wide variation in practice. In many studies describing and/or comparing TRS systems, detailed explanations of the extent of sampling are lacking. This makes it impossible to compare the meaning of even complete tumor response in different studies. Hence, the extent of sampling should be included in the "Material and Methods" section of all scientific publications such that the comparability of data from different studies can be evaluated. Before a case can be classified as "complete remission", total sampling of the pancreas is advised (with additional sections from the blocks with fibrous changes), and the prior biopsy diagnosis ought to be reviewed in consultation. Unless there is overtly abundant residual cancer that can be readily documented, extensive sampling of the pancreas is warranted to assess the amount and distribution of the residual cancer properly.

Areas of disagreement and future research

Some topics that were covered during the meeting did not reach agreement and remain open for further discussions and studies. The optimal number of tiers within a scoring system remains a topic of debate. An ideal TRS system 1) has strong prognostic value, 2) allows clinically relevant patient stratification, and 3) is reproducible, reliable and practical. When attempting to establish these 3 characteristics, the number of tiers in a TRS system plays a central role. The majority of participants agreed that distinguishing complete response from non-complete response (either substantial or no response) is reliable and easy to do, provided that total sampling and careful evaluation are performed. However, since complete response is observed in only a small minority of patients, a two-tiered system does not allow clinically relevant stratification in the majority of patients and is therefore of limited use. However, as more tiers are added, prognostic systems generally face a tradeoff between adding prognostic value in terms of more discriminatory power (more tiers) and maintaining simplicity so as to optimize inter-observer agreement and applicability in daily practice (fewer tiers).

Further research is needed to investigate whether additional tiers provide more relevant patient stratification without affecting interobserver agreement.

Information gathered by the pre-meeting survey and comparison of opinions between delegates from different continents highlight that there is significant divergence in practice e.g., in terms of dissection method and the use of photo documentation. For example, 5/13 (38.5%) of European pathologists agreed that axial slicing is the dissection method of choice to accurately score response after neoadjuvant therapy in PDAC, versus 0/7 (0%) of American pathologists. These issues were not addressed during the consensus meeting because they are not exclusively relevant to the assessment of tumor regression but rather present a potential source of nonuniform reporting on any surgical pancreatic cancer specimens, including those from treatment-naïve patients.

During the meeting, the use of ancillary or novel techniques to improve the prognostic value of TRS was discussed. Immunohistochemistry and other markers deemed to relate to response to treatment deserve further investigation to test their relevance for clinical practice. Another potential area of exploration concerns machine learning (ML) and artificial intelligence (AI) strategies. During the meeting, the potential of ML and AI strategies in TRS was widely acknowledged. These strategies have become increasingly versatile in recent years, and some have been successfully implemented for pathology purposes.^{30, 31} As ML and AI utilize an algorithmic approach, interobserver variability might be reduced.

DISCUSSION

The Amsterdam International Consensus Meeting obtained consensus by 23 expert pathologists from 4 continents on 7 statements regarding TRS to assess the effect of neoadjuvant treatment in resection specimens with PDAC. Objective criteria, adequate interobserver agreement, and standardized sampling and reporting are desirable for both clinical practice and clinical research. Objective definitions and easy-to-apply evaluation criteria are necessary for accurate and reproducible evaluation of the tumor response to treatment.

This consensus did not use a formal evidence-based approach nor did it provide a new TRS system. Rather, the consensus was intended as the start of a process to improve TRS in PDAC. In that spirit, a formal Delphi process was not used, since the authors preferred to have a face-to-face meeting to discuss current problems and dilemmas in this particular field.

Adequate scoring of tumor response after neoadjuvant therapy in resected pancreatic cancer aims to result in 1) more accurate assessment of treatment response and outcome prognostication, 2) a useful measure to guide adjuvant regimens, and 3) a valuable tool in comparative trials of neoadjuvant therapies. Given the complexity of challenges with current TRS practices, the development of an improved, easy-to-apply and objective TRS system and a consensus on the extent of tissue sampling are necessary. Once the improved system is established, retrospective and prospective validation is of paramount importance. The International Study Group of Pancreatic Pathologists (ISGPP), which was formed during the consensus meeting, aims to achieve these outcomes by facilitating international multidisciplinary collaborative research in alliance with the Neo-adjuvant Working Group of Pancreatobiliary Pathology Society (PBPath.org).

DISCLOSURE/CONFLICT OF INTEREST

This research meeting was funded by a grant from the Dutch Cancer Society (grant number UVA2013-5842).

Supplementary information is available at Modern Pathology's website.

REFERENCES

1. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med*. 2014;371(11):1039-49.
2. Versteijne E, Vogel JA, Besselink MG, Busch ORC, Wilmink JW, Daams JG, et al. Meta-analysis comparing upfront surgery with neoadjuvant treatment in patients with resectable or borderline resectable pancreatic cancer. *Br J Surg*. 2018;105(8):946-58.
3. Versteijne E, Suker M, Groothuis K, Akkermans-Vogelaar JM, Besselink MG, Bonsing BA, et al. Preoperative chemoradiotherapy versus immediate surgery for resectable and borderline resectable pancreatic cancer: results of the Dutch randomized phase III PREOPANC trial. *J Clin Oncol*. 2020;Jco1902274.
4. Jang JY, Han Y, Lee H, Kim SW, Kwon W, Lee KH, et al. Oncological benefits of neoadjuvant chemoradiation with gemcitabine versus upfront surgery in patients with borderline resectable pancreatic cancer: a prospective, randomized, open-label, multicenter phase 2/3 trial. *Ann Surg*. 2018;268(2):215-22.
5. Suker M, Beumer BR, Sadot E, Marthey L, Faris JE, Mellon EA, et al. FOLFIRINOX for locally advanced pancreatic cancer: a systematic review and patient-level meta-analysis. *Lancet Oncol*. 2016;17(6):801-10.
6. Evans DB, Rich TA, Byrd DR, Cleary KR, Connelly JH, Levin B, et al. Preoperative chemoradiation and pancreaticoduodenectomy for adenocarcinoma of the pancreas. *Arch Surg*. 1992;127(11):1335-9.
7. Moutardier V, Magnin V, Turrini O, Viret F, Hennekinne-Mucci S, Goncalves A, et al. Assessment of pathologic response after preoperative chemoradiotherapy and surgery in pancreatic adenocarcinoma. *Int J Radiat Oncol Biol Phys*. 2004;60(2):437-43.
8. Washington K, Berlin J, Branton P, Washington K, Berlin J, Branton P, et al. Protocol for the examination of specimens from patients with carcinoma of the pancreas. *College of American Pathology*. 2016.
9. Chun YS, Cooper HS, Cohen SJ, Konski A, Burtness B, Denlinger CS, et al. Significance of pathologic response to preoperative therapy in pancreatic cancer. *Ann Surg Oncol*. 2011;18(13):3601-7.

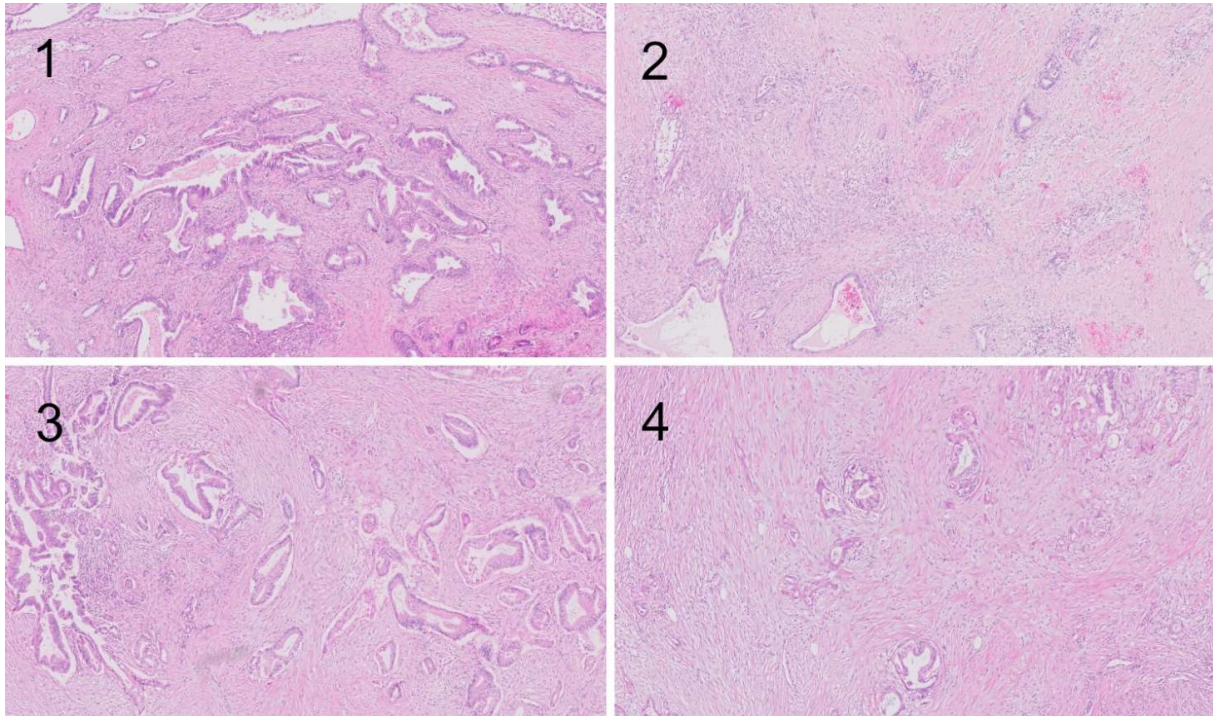
10. Chatterjee D, Katz MH, Rashid A, Varadhachary GR, Wolff RA, Wang H, et al. Histologic grading of the extent of residual carcinoma following neoadjuvant chemoradiation in pancreatic ductal adenocarcinoma: a predictor for patient outcome. *Cancer*. 2012;118(12):3182-90.
11. Hartman DJ, Krasinskas AM. Assessing treatment effect in pancreatic cancer. *Arch Pathol Lab Med*. 2012;136(1):100-9.
12. Townend P, de Reuver PR, Chua TC, Mittal A, Clark SJ, Pavlakis N, et al. Histopathological tumour viability after neoadjuvant chemotherapy influences survival in resected pancreatic cancer: analysis of early outcome data. *ANZ J Surg*. 2018;88(3):E167-e72.
13. Panni RZ, Gonzalez I, Hartley CP, Williams GA, Liu J, Hawkins WG, et al. Residual tumor index: a prognostically significant pathologic parameter in neoadjuvant-treated pancreatic ductal adenocarcinoma. *Am J Surg Pathol*. 2018;42(11):1480-7.
14. He J, Blair AB, Groot VP, Javed AA, Burkhart RA, Gemenetzis G, et al. Is a pathological complete response following neoadjuvant chemoradiation associated with prolonged survival in patients with pancreatic cancer? *Ann Surg*. 2018;268(1):1-8.
15. Rowan DJ, Logunova V, Oshima K. Measured residual tumor cellularity correlates with survival in neoadjuvant treated pancreatic ductal adenocarcinomas. *Ann Diagn Pathol*. 2019;38:93-8.
16. Zhang ML, Kem M, Rodrigues C, Sandini M, Ciprani D, Hank T, et al. Microscopic size measurements predict outcomes in post-neoadjuvant resections of pancreatic ductal adenocarcinoma (PDAC). *Histopathology*. 2020.
17. Cacciato Insilla A, Vivaldi C, Giordano M, Vasile E, Cappelli C, Kauffmann E, et al. Tumor regression grading assessment in locally advanced pancreatic cancer after neoadjuvant FOLFIRINOX: interobserver agreement and prognostic implications. *Frontiers in Oncology*. 2020;10(64).
18. S NK, Serra S, Dhani N, Hafezi-Bakhtiari S, Szentgyorgyi E, Vajpeyi R, et al. Regression grading in neoadjuvant treated pancreatic cancer: an interobserver study. *J Clin Pathol*. 2017;70(3):237-43.
19. Meredith KL, Weber JM, Turaga KK, Siegel EM, McLoughlin J, Hoffe S, et al. Pathologic response after neoadjuvant therapy is the major determinant of survival in patients with esophageal cancer. *Ann Surg Oncol*. 2010;17(4):1159-67.

20. Wu TT, Chirieac LR, Abraham SC, Krasinskas AM, Wang H, Rashid A, et al. Excellent interobserver agreement on grading the extent of residual carcinoma after preoperative chemoradiation in esophageal and esophagogastric junction carcinoma: a reliable predictor for patient outcome. *Am J Surg Pathol.* 2007;31(1):58-64.
21. Verbeke C, Häberle L, Lenggenger D, Esposito I, Verbeke C, Häberle L, et al. Pathology assessment of pancreatic cancer following neoadjuvant treatment: Time to move on. *Pancreatology.* 2018;18(5):467-76.
22. Lee SM, Katz MH, Liu L, Sundar M, Wang H, Varadhachary GR, et al. Validation of a proposed tumor regression grading scheme for pancreatic ductal adenocarcinoma after neoadjuvant therapy as a prognostic indicator for survival. *Am J Surg Pathol.* 2016;40(12):1653-60.
23. Macedo FI, Ryon E, Maithel SK, Lee RM, Kooby DA, Fields RC, et al. Survival Outcomes Associated With Clinical and Pathological Response Following Neoadjuvant FOLFIRINOX or Gemcitabine/Nab-Paclitaxel Chemotherapy in Resected Pancreatic Cancer. *Ann Surg.* 2019;270(3):400-13.
24. Truty MJ, Kendrick ML, Nagorney DM, Smoot RL, Cleary SP, Graham RP, et al. Factors Predicting Response, Perioperative Outcomes, and Survival Following Total Neoadjuvant Therapy for Borderline/Locally Advanced Pancreatic Cancer. *Ann Surg.* 2019.
25. Peng JS, Wey J, Chalikonda S, Allende DS, Walsh RM, Morris-Stiff G. Pathologic tumor response to neoadjuvant therapy in borderline resectable pancreatic cancer. *Hepatobiliary Pancreat Dis Int.* 2019;18(4):373-8.
26. Murata Y, Mizuno S, Kishiwada M, Hamada T, Usui M, Sakurai H, et al. Impact of histological response after neoadjuvant chemoradiotherapy on recurrence-free survival in UICC-T3 pancreatic adenocarcinoma but not in UICC-T4. *Pancreas.* 2012;41(1):130-6.
27. Cloyd JM, Wang H, Egger ME, Tzeng CD, Prakash LR, Maitra A, et al. Association of Clinical Factors With a Major Pathologic Response Following Preoperative Therapy for Pancreatic Ductal Adenocarcinoma. *JAMA Surg.* 2017;152(11):1048-56.
28. Verbeke C, Lohr M, Karlsson JS, Del Chiaro M, Verbeke C, Lohr M, et al. Pathology reporting of pancreatic cancer following neoadjuvant therapy: challenges and uncertainties. *Cancer Treat Rev.* 2015;41(1):17-26.

29. Ryan R, Gibbons D, Hyland JM, Treanor D, White A, Mulcahy HE, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology*. 2005;47(2):141-6.
30. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*. 2016;33:170-5.
31. Chang HY, Jung CK, Woo JI, Lee S, Cho J, Kim SW, et al. Artificial intelligence in pathology. *Journal of pathology and translational medicine*. 2019;53(1):1-12.

TABLES AND FIGURES

Figure 1. Illustrations of intratumoral heterogeneity and the difficulty of recognizing tumor regression



Legend:

Image descriptions: The amount and distribution of cancer cells and fibrosis is often very heterogeneous across the different tumor areas, both in non-treated and treated patient. Representative micrographs 1 and 2 show two distinct areas from the same pancreatic specimen that was treated with neoadjuvant chemotherapy. Representative micrographs 3 and 4 show two distinct areas from the same pancreatic specimen that was *not* treated with neoadjuvant chemotherapy.

Table 1. Pre-meeting survey outcomes regarding the assessment of tumor response following neoadjuvant therapy in pancreatic cancer

		Agree	Disagree	Other
<i>≥ 80% similar responses</i>				
1.	TRS is important because it provides information about the effect of neoadjuvant treatment that is not provided by the pathology-based descriptors.	100%	0.0%	0.0%
2.	Microscopic heterogeneity of both tumor morphology and response to treatment is a major issue in the evaluation of tumor response scoring.	95.7%	4.3%	0.0%
3.	To date there is sufficient evidence to support the criteria, thresholds and categories that are at the basis of current tumor regression scoring systems.	0.0%	91.7%	8.3%
4.	The best scoring system does not exist yet.	87.0%	13.0%	0.0%
5.	The extent of fibrosis in comparison to the extent of viable tumor, such as it is used in the Mandard system is a reliable scoring criterion.	13.0%	87.0%	0.0
6.	TRS is important to predict response to post-operative therapies e.g. in case of recurrence.	87.0%	8.7%	4.3%
7.	Interobserver variability is a major issue in the evaluation of tumor regression.	82.6%	13.0%	4.3%
8.	To date there is adequate evidence regarding the extent of tissue sampling that is required for accurate evaluation of tumor regression.	13.0%	82.6%	4.3%
<i>≥ 60% and < 80% similar responses</i>				
9.	The best approach to quantify regression is estimating an average of the diverse changes, taking into account the extent of areas showing diverse grades of regression.	78.3%	13.0%	8.7%
10.	For accurate evaluation of the effect of neoadjuvant treatment, information on tumor size on imaging before treatment is necessary.	73.9%	17.4%	8.7%
11.	The pancreas should be submitted in a stepwise manner. If no tumor is found in initial tissue sections, it has to be entirely submitted.	69.6%	17.4	13.0%
12.	Ideally TRS correlates with improved R0 rates.	65.2%	30.4%	4.3%
13.	There is currently sufficient evidence regarding the prognostic value of TRS.	30.4%	65.2%	4.3%
14.	Axial slicing is the dissection method of choice to accurately score response after neoadjuvant therapy in PDAC.	26.1%	60.9%	8.7%
15.	Immunohistochemistry is of added value when assessing TRS independently of the system used.	26.1%	60.1%	13.0%
<i>< 60% similar responses</i>				
16.	The scoring system should be based on numeric values i.e. the percentage of viable tumor cells, rather than being descriptive.	39.1%	56.5%	4.3%
17.	Detailed photo documentation of the resection specimen is essential for the accurate assessment of response.	34.8%	56.5%	8.7%
18.	A four-tiered system (comparable to CAP system) is accurate and reproducible.	30.4%	56.5%	13.0%
19.	For accurate evaluation of the effect of neoadjuvant treatment knowledge of the tumor morphology prior to treatment is necessary.	13.0%	52.2%	34.8%
20.	The preferred scoring system to date is; Chatterjee = Agree, CAP = Disagree.	30.4%	47.8%	17.4%
21.	To ensure reproducibility a three-tiered system as proposed by Chatterjee et al should be used.	43.5%	43.5%	13.0%
<i>Open question</i>				

22. What is according to you the main goal and clinical relevance of TRS?

- - -

Table 2. Consensus statements of the 2019 Amsterdam International Consensus Meeting on tumor response grading in pancreatic cancer specimens following neoadjuvant therapy

Consensus statements	Agreement
1. TRS is important because it provides information about the effect of neoadjuvant treatment that is not provided by other histopathology-based descriptors.	90%
2. TRS for resected PDAC following neoadjuvant therapy should assess residual (viable) tumor burden instead of tumor regression.	100%
3. The CAP (College of American Pathologists) scoring system is considered the most adequate scoring system to date because it is based on the presence and amount of residual cancer cells instead of tumor regression.	90%
4. The defining criteria of the categories in the CAP scoring system should be improved by replacing subjective terms including “minimal” or “extensive” with objective criteria to evaluate the extent of viable tumor.	95%
5. The improved, consensus-based system should be validated retrospectively and prospectively.	95%
6. Prospective studies should determine the extent of tissue sampling that is required to ensure adequate assessment of the residual cancer burden, taking into account the heterogeneity of tumor response.	100%
7. In future scientific publications, the extent of tissue sampling should be described in detail in the “Materials and Methods” section.	100%

Table 3. Most commonly used scoring systems of tumor response in resected pancreatic cancer after neoadjuvant therapy

Scoring system	Grade	Criteria
College of American Pathologists (CAP). Washington et al. (2014) ⁸	0	No viable cancer cells
	1	Single cells or rare small groups of cancer cells
	2	Residual cancer with evident tumor regression, but more than single cells or rare small groups of cancer cells
	3	Extensive residual cancer with no evident tumor regression
MD Anderson (MDACC) Chatterjee et al. (2012) ^{10,22}	0	No residual carcinoma
	1	Minimal residual carcinoma (single cells or small groups of cancer cells; <5% residual carcinoma in treated tumor bed
	2	5% or more carcinoma in treated tumor bed
Evans et al. (1992) ⁶	1	Characteristic cytological changes of malignancy are present, but little (<10%) or no tumor cell destruction is evident.
	2a	Destruction of 10–50% of tumor cells
	2b	Destruction of 51–90% of tumor cells
	3	<10% viable appearing tumor cells present
	3M	Sizable pools of mucin are present.
	4	No viable tumor cells present
	4M	Acellular pools of mucin are present

SUPPLEMENTARY MATERIAL

Table S1. List of attendees

Name	Corresponding institution
Volkan Adsay, MD	Department of Pathology, Koc University and KUTTAM Research Center, Istanbul, Turkey
Olca Basturk, MD	Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, US
Marc G. Besselink, MD, MSc, PhD	Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
Lodewijk Brosens, MD, MSc, PhD	Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands
Fiona Campbell, MD, FRCPath	Department of Pathology, Royal Liverpool University Hospital, Liverpool, UK
Claudio Doglioni, MD	Department of Pathology, IRCCS San Raffaele Scientific Institute, Milan, Italy
Irene Esposito, MD	Institute of Pathology, Heinrich-Heine-University and University Hospital of Duesseldorf, Germany
Roger Feakins MB, BCh, BAO, BA, MD, FRCPI, FRCPath	Department of Pathology, Royal Free London NHS Trust, London, UK
Noriyoshi Fukushima, MD, PhD	Department of Pathology, Jichi Medical University Hospital, Tochigi, Japan
Bart Geerts, MD, MSc, PhD	Healthplus.ai and Department of Anesthesiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
Anthony J Gill, MD, FRCPA	Cancer Diagnosis and Pathology Group Kolling institute of Medical Research Royal North Shore NSW Australia and University of Sydney, Sydney NSW Australia
Joost Huiskens, MD, PhD	SAS Netherlands
Jeffrey Kaplan, MD	Department of Pathology, University of Colorado Hospital, Denver, CO, US
Bas Groot-Koerkamp, MD, PhD	Department of Surgery, Erasmus Medical Center, Rotterdam, The Netherlands
Seung-Mo Hong, MD, PhD	Department of Pathology, Asan Medical Center, Seoul, Republic of Korea
Ralph Hruban, MD	Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA
Boris V. Janssen, BSc	Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands, Department of Pathology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, The Netherlands
Alyssa Krasinkas, MD	Department of Pathology, Emory University, Atlanta, GA, US
Claudio Luchini, MD, PhD	Department of Diagnostics and Public Health, University and Hospital Trust of Verona, Verona, Italy
Johan Offerhaus, MD, PhD, MPH	Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands
Stijn van Roessel, MD, MSc	Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

Arantza Fariña Sarasqueta, MD, MSc, PhD	Department of Pathology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Netherlands
Chanjuan Shi, MD, PhD	Department of Pathology, Duke University Medical Center, Durham, NC, USA
Aatur Singhi, MD, PhD	Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, US
Thomas F. Stoop, BSc	Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Netherlands
Eline Soer, MD	Department of Pathology, Cancer Center Amsterdam, Academic Medical Center, University of Amsterdam, The Netherlands
Elizabeth Thompson, MD, PhD	Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA
Geertjan van Tienhoven, MD PhD	Department of Radiation Oncology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, The Netherlands
Faik Tutucu, MD	Department of Pathology, Cancer Center Amsterdam, Academic Medical Center, University of Amsterdam, The Netherlands
Marie-Louise F. Velthuysen, MD, PhD	Department of Pathology, Erasmus Medical Center, Rotterdam, The Netherlands
Caroline S. Verbeke, MD, PhD	Department of Pathology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway
Joanne Verheij, MD, PhD	Department of Pathology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Netherlands
Huamin Wang, MD, PhD	Department of Anatomical Pathology, University of Texas, MD Anderson Cancer Center, Houston, Texas, USA
Johanna W. Wilmink, MD, PhD	Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, The Netherlands

Supplementary Table S2: Sub-analysis of pre-meeting survey according to continent

			Oceania (n = 1)	Europe (n = 13)	Asia (n = 2)	North-America (n = 7)
1.	TRS is important because it provides information about the effect of neoadjuvant treatment that is not provided by the pathology-based descriptors.	Agree: Disagree: Other	100% 0% 0%	100% 0% 0%	100% 0% 0%	100% 0% 0%
2.	Microscopic heterogeneity of both tumor morphology and response to treatment is a major issue in the evaluation of tumor response scoring.	Agree: Disagree: Other	100% 0% 0%	100% 0% 0%	100% 0% 0%	86% 14% 0%
3.	To date there is sufficient evidence to support the criteria, thresholds and categories that are at the basis of current tumor regression scoring systems.	Agree: Disagree: Other	0% 100% 0%	0% 84.6% 15.4%	0% 100% 0%	0% 100% 0%
4.	The best scoring system does not exist yet.	Agree: Disagree: Other	100% 0% 0%	84.6% 15.4% 0%	100% 0% 0%	86% 14% 0%
5.	The extent of fibrosis in comparison to the extent of viable tumor, such as it is used in the Mandard system is a reliable scoring criterion.	Agree: Disagree: Other	0% 100% 0%	15.4% 84.6% 0%	0% 100% 0%	14% 86% 0%
6.	TRS is important to predict response to post-operative therapies e.g. in case of recurrence.	Agree: Disagree: Other	100% 0% 0%	77% 15.3% 7.7%	100% 0% 0%	100% 0% 0%
7.	Interobserver variability is a major issue in the evaluation of tumor regression.	Agree: Disagree: Other	100% 0% 0%	77% 15.3% 7.7%	100% 0% 0%	86% 14% 0%
8.	To date there is adequate evidence regarding the extent of tissue sampling that is required for accurate evaluation of tumor regression.	Agree: Disagree: Other	0% 100% 0%	15.3% 77% 7.7%	50% 50% 0%	0% 100% 0%
9.	The best approach to quantify regression is estimating an average of the diverse changes, taking into account the extent of areas showing diverse grades of regression.	Agree: Disagree: Other	100% 0% 0%	77% 23% 0%	50% 0% 50%	86% 14% 0%
10.	For accurate evaluation of the effect of neoadjuvant treatment, information on tumor size on imaging before treatment is necessary.	Agree: Disagree: Other	100% 0% 0%	69.2% 15.4% 15.4%	100% 0% 0%	71.4% 28.6% 0%
11.	The pancreas should be submitted in a stepwise manner. If no tumor is found in initial tissue sections, it has to be entirely submitted.	Agree: Disagree: Other	100% 0% 0%	61.5% 23% 15.5%	50% 50% 0%	86% 0% 14%
12.	Ideally TRS correlates with improved R0 rates.	Agree: Disagree: Other	0% 100% 0%	69.2% 23.1% 7.7%	100% 0% 100%	57.2% 42.8% 0%
13.	There is currently sufficient evidence regarding the prognostic value of TRS.	Agree: Disagree: Other	0% 100% 0%	23.1% 69.2% 7.7%	100% 0% 0%	28.6% 71.4% 0%
14.	Axial slicing is the dissection method of choice to accurately score response after neoadjuvant therapy in PDAC.	Agree: Disagree: Other	0% 100% 0%	38.5% 46.1% 15.4%	50% 50% 0%	0% 100% 0%
15.	Immunohistochemistry is of added value when assessing TRS independently of the system used.	Agree: Disagree: Other	0% 100% 0%	15.4% 69.2% 15.4%	50% 50% 0%	42.8% 52.7% 0%
16.	The scoring system should be based on numeric values i.e. the percentage of viable tumor cells, rather than being descriptive.	Agree: Disagree: Other	100% 0% 0%	23% 77% 0%	100% 0% 0%	42.8% 42.8% 14.4%
17.	Detailed photo documentation of the resection specimen is essential for the accurate assessment of response.	Agree: Disagree: Other	0% 100% 0%	38.5% 46.1% 15.4%	100% 0% 100%	14% 86% 0%
18.	A four-tiered system (comparable to CAP system) is accurate and reproducible.	Agree: Disagree: Other	0% 100% 0%	38.5% 38.5% 23%	0% 100% 0%	28.6% 71.4% 0%
19.	For accurate evaluation of the effect of neoadjuvant treatment knowledge of the tumor morphology prior to treatment is necessary.	Agree: Disagree: Other	0% 0% 100%	0% 53.8% 46.2%	100% 0% 0%	14.3% 71.4% 14.3%
20.	The preferred scoring system to date is; Chatterjee = Agree, CAP = Disagree.	Agree: Disagree: Other	100% 0% 0%	15.4% 53.8% 30.8%	100% 0% 0%	28.6% 57.1% 14.3%
21.	To ensure reproducibility a three-tiered system as proposed by Chatterjee et al should be used.	Agree: Disagree: Other	0% 100% 0%	23% 61.5% 15.5%	100% 0% 0%	71.4% 14.3% 14.3%