Title: Reproducibility of the peritoneal regression grading score for assessment of response to therapy in peritoneal metastasis.

Authors: Solass W, Sempoux C, Carr NJ, Bibeau F, Neureiter D, Jäger T, Di Caterino T, Brunel C, Klieser E, Fristrup CW, Mortensen MB, Detlefsen S

Journal: Histopathology

Year: 2019 Jun

Issue: 74

Volume: 7

Pages: 1014-1024

DOI: 10.1111/his.13829

DR NORMAN JOHN CARR (Orcid ID : 0000-0002-5087-6874)
DR DANIEL NEUREITER (Orcid ID : 0000-0001-9155-5762)
DR SÖNKE DETLEFSEN (Orcid ID : 0000-0002-9466-2333)

Article type : Original Article

Reproducibility of the Peritoneal Regression Grading Score (PRGS) for assessment of response to therapy in peritoneal metastasis

Running title: Reproducibility of PRGS for peritoneal metastasis

Wiebke Solass[1], Christine Sempoux[2], Norman Carr[3], Frédéric Bibeau[4], Daniel Neureiter[5], Tarkan Jäger[6], Tina Di Caterino[7], Christophe Brunel[2], Eckhard Klieser[5], Claus Fristrup[8,9], Michael Bau Mortensen[8,9,10], Sönke Detlefsen[7,8,10]

[1]Institute of Pathology and Neuropathology, Eberhard-Karls-University Tuebingen and National Center for Pleura and Peritoneum, University of Tuebingen, Germany

[2]Service of Clinical Pathology, Institute of Pathology, Lausanne University Hospital, Lausanne, Switzerland

[3]Peritoneal Malignancy Institute, Basingstoke and North Hampshire Hospital, Basingstoke, UK

[4]Institute of Pathology, University Caen and Réseau National des Tumeurs Rares du Péritoine (RENAPE), Caen, France

[5]Institute of Pathology, Paracelsus Medical University, Salzburger Landeskliniken (SALK), Salzburg, Austria

[6]Department of Surgery, Paracelsus Medical University, Salzburg, Austria

[7]Department of Pathology, Odense University Hospital, Odense, Denmark

[8]Odense PIPAC Center (OPC) and Odense Pancreas Center (OPAC), Odense University Hospital, Odense, Denmark

[9]Department of Surgery, HPB and Upper GI Section, Odense University Hospital, Odense, Denmark

[10]Institute of Clinical Research, University of Southern Denmark, Odense, Denmark

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/his.13829

Corresponding author:
Sönke Detlefsen, MD, PhD
Senior Consultant, Assoc. Prof.
Department of Pathology, Odense University Hospital
J.B. Winsløws Vej 15, 5000 Odense C, Denmark
E-mail: Sonke.Detlefsen@rsyd.dk
Phone: +45 6541 4806
Fax: +45 6591 2943

Conflict of interest:
The authors declare no potential conflict of interest.

Abstract

Introduction

The 4-tiered Peritoneal Regression Grading Score (PRGS) is assessing response to chemotherapy in peritoneal metastasis (PM). The PRGS is for example used to assess response to Pressurized IntraPeritoneal Aerosol Chemotherapy (PIPAC). However, the reproducibility of the PRGS is currently unkown. We aimed to evaluate the interobserver and intraobserver variability of the PRGS.

Materials and methods

33 patients who underwent at least 3 PIPAC treatments as part of the PIPAC-OPC1 or PIPAC-OPC2 clinical trials at Odense University Hospital, Denmark, were included. Prior to each therapy cycle, peritoneal quadrant biopsies were obtained, and three H&E stained step sections were scanned and uploaded to a pseudonymized web library. For determining the interobserver variability, eight pathologists assessed the PRGS for each quadrant biopsy, and Krippendorff's alpha and Intraclass Correlation Coefficients (ICCs) were calculated. For determining intraobserver variability, three pathologists repeated their own assessments, and Cohen's kappa and ICCs were calculated.

Results

A total of 331 peritoneal biopsies were analyzed. Interobserver variability for PRGS of each biopsy and for the mean and maximum PRGS per biopsy set was moderate to good/substantial. The intraobserver variability for PRGS of each biopsy and for the mean and maximum PRGS per biopsy set was good to excellent/almost perfect.

Discussion

Our data support the Peritoneal Regression Grading Score (PRGS) as a reproducible and useful tool to assess response to intraperitoneal chemotherapy in peritoneal metastasis. Future studies should evaluate the prognostic and predictive role of the PRGS.

Keywords: colorectal cancer, gastric cancer, interobserver variability, ovarian cancer, pancreatic cancer, peritoneal metastasis, pressurized intraperitoneal aerosol chemotherapy (PIPAC), tumor regression grading

Introduction

Despite of the development of new molecular techniques, histological assessment remains the gold standard in the diagnosis of most human malignancies. The effect of treatment (*e.g.*, chemotherapy) on the malignant tumor tissue – in the primary tumor as well as in its metastases - is assessed by histological characteristics like fibrosis, acellular mucin pools, hyalinosis, and/or infarct-like necrosis, resulting in a relative reduction of viable tumor cells [1]. Hence, these regressive features can be used to identify subpopulations of patients who are most likely to benefit from a given therapy. Most published scoring systems for the assessment of the histological

response to neoadjuvant treatment are based on surgical resection specimens of the primary tumor or metastases [2-7]. In 2016, the 4-tiered Peritoneal Regression Grading Score (PRGS) for the histological assessment of response to therapy in peritoneal metastasis (PM) was proposed by a group of European pathologists [8]. The PRGS score is potentially clinically important in the assessment of histological response to intraperitoneal chemotherapy, particularly when such a therapy is given several times and the decision whether the patient should receive additional treatments depends on the histological response. A novel example of such a treatment is the Pressurized IntraPeritoneal Aerosol Chemotherapy (PIPAC), where chemotherapeutics are aerosolized within the peritoneal cavity during a standard laparoscopy at a capnoperitoneum [9-11]. Currently, PIPAC is an experimental treatment, and randomized, controlled trials are lacking at present [12, 13]. However, PIPAC seems to be a safe procedure, able to induce objective histological regression, to improve quality of life, and to result in improved survival [14-18]. The interobserver and intraobserver variability in assessing the PRGS in PM as well as its prognostic or predictive value are not known. However, the accuracy of current imaging systems for detection and therapy response assessment of PM is limited, and the PRGS is gaining rapidly clinical acceptance [17-21].

In this study, we evaluated the reproducibility of the PRGS in PM. Specific questions were the interobserver variability, the intraobserver variability, possible changes in the accuracy during the course of therapy, and the reproducibility of the maximal regression score vs. the mean regression score. Our study included peritoneal biopsies with PM deriving from a wide range of different primary tumors, scored by a group of pathologists with varying experience.

Methods

Study design

We performed an observational, retrospective, longitudinal, single-blinded study. The study has been approved by the Data Protection Agency of the Region of Southern Denmark (17/30427). One pathologist from each participating center signed a Data Processor Agreement, issued by the Data Protection Agency of the Region of Southern Denmark. All patients were part of the PIPAC-OPC1 (NCT02320448, n=27) or PIPAC-OPC2 (EudraCT provided, GCP monitored (EudraCT 2016-003394-18), n=6) clinical trials, approved by the Ethics Committee of the Region of Southern Denmark (S-20140211, S-20160100).

All peritoneal biopsies were obtained from 33 patients with PM treated at Odense PIPAC Centre (OPC), Odense University Hospital, Denmark, during the course of repeated Pressurized IntraPeritoneal Aerosol Chemotherapy (PIPAC) cycles. Based on current evidence, patients with PM of colorectal or appendiceal origin were treated with oxaliplatin 92 mg/m^2 in 150 ml dextrose, while patients with PM of other origin were treated with a combination of cisplatin 7.5 mg/m^2 in 150 ml saline and doxorubicin 1.5 mg/ m^2 in 50 ml saline [18]. A total of 331 peritoneal biopsies were evaluated, with three step sections per biopsy, resulting in a total of 993 step sections. The included patients had PM deriving from different primary tumors of different origin (Table 1).

Peritoneal biopsy specimens

All patients included in this study underwent at least three PIPAC procedures, and from all included patients, peritoneal quadrant biopsies taken prior to each PIPAC procedure were included. According to current recommendations, biopsies were

taken from macroscopically tumor suspect areas in all four abdominal quadrants, if technically possible. In some instances, however, only 1, 2 or 3 peritoneal biopsies could be taken for technical reasons. After obtaining the first set of biopsies prior to PIPAC treatment 1, the biopsy sites were marked with metal clips to ensure that subsequent biopsies were collected from the same sites. In order to ensure optimal fixation for reliable histopathological analysis, biopsies were fixed in 10% buffered formalin for 24–48 hours. Then, samples were embedded in paraffin using a controlled temperature. Two series of three 4-5 µm thick step sections from each biopsy were stained with hematoxylin and eosin (H&E) at the Department of Pathology, Odense University Hospital (OUH), Denmark.

Web library

All quantitative evaluation was performed on digitalized H&E stained slides. From the two available H&E stained step sections, the slide with the greater tissue area was scanned using a 20x objective on the NanoZoomer 2.0HT whole slide scanner (Hamamatsu Photonics, Hamamatsu, Japan). The digitalized slides were uploaded to a pseudonymized web library. Each pathologist participating in this study received a personalized code to access the web library, and each access to the web library was logged.

Pathologists

All slides were analyzed online by eight independent pathologists from different institutes, different countries and with diverse levels of practical experience in the assessment of histological regression grading in PM. Four of the participating pathologists were co-authors of the proposal article regarding the PRGS [8]. The other

four pathologists were trained to get familiar with the PRGS system and were given a copy of the reference publication [8]. Besides, the untrained observers were taught to use the histological criteria of regression. Five pathologists were senior consultants, 3 of whom had a special research interest in peritoneal pathology for >10 years, and three pathologists were residents in pathology with 2, 3, and 5 years of working experience in pathology. All eight pathologists were involved in the assessment of interobserver variability and assigned a PRGS score to each slide under investigation. Three pathologists (two senior consultants and one resident) repeated their own assessment, with 5, 10 and 12 weeks between the assessments, for determining the intraobserver variability.

Peritoneal Regression Grading Score (PRGS)

The PRGS defines four categories, based on the presence of residual tumor cells and the extent of regressive features. Major histological features of regression are fibrosis, inflammation, hyalinosis, acellular mucin pools, ischemic necrosis, accumulation of macrophages / multinucleated giant cells, and granulomas [8]. PRGS 1 corresponds to a complete regression with absence of tumor cells (Figure 1A-B); PRGS 2 to a major histological response with regressive features predominant over residual tumor cells (Figure 1C); PRGS 3 to a minor histological response with predominance of residual tumor cells over regressive features (Figure 1D); and PRGS 4 to a lack of histological response to therapy where the tumor cells are not accompanied by any regressive features (Figure 1E-H)[8]. According to the proposal, a PRGS was assessed for each quadrant biopsy. Moreover, the mean PRGS, based on the individual scores from the four quadrant biopsies, was given.

Statistics

In order to determine the interobserver and intraobserver reproducibility, the statistical question is how reliable the measurements are. Reliability is defined as the extent to which measurements can be replicated [22]. For evaluating the interobserver agreement between multiple raters, Krippendorff's alpha using ordinal data was calculated. Krippendorff's alpha can be used with any sample size, number of observers, and kind of data in addition to handling missing data appropriately. For evaluating the intraobserver agreement, Cohen's kappa was used. Cohen's kappa and Krippendorff's alpha take coefficients ranging from 0 (or <0 in extreme cases) to 1. A coefficient of 0 is indicative of no agreement and a coefficient of 1 represents perfect agreement. Coefficients below 0 indicate poor/systematic disagreement, a coefficient between 0 and 0.2 slight agreement, between 0.21 and 0.40 fair, between 0.41 and 0.60 moderate, between 0.61 and 0.80 substantial, and between 0.81 and 1.0 almost perfect agreement.

In addition to the calculations above, Intraclass Correlation Coefficients (ICCs) were calculated for interobserver variability and intraobserver variability. ICC is a reliability index that reflects both degree of correlation and agreement between measurements. It has been widely used in conservative care medicine to evaluate interobserver, test-retest, and intraobserver reliability of numerical or continuous measurements. For the interobserver variability, the ICCs were reported with 95% confidence intervals based on a single rater, absolute-agreement, using the two-way random-effects model. For the intraobserver variability, the ICCs were reported with 95% confidence intervals based on a single rater, absolute agreement, using the two-way mixed-effects model. ICCs less than 0.5 are indicative of poor reliability,

ICCs between 0.5 and 0.75 indicate moderate reliability, ICCs between 0.75 and 0.9 indicate good reliability, and ICCs greater than 0.90 indicate excellent reliability.

The statistical analyses were performed using Stata v. 15 (StataCorp LLC, College Station, Texas), with the addition of *kappaetc* (Daniel Klein, INCHER-Kassel, University of Kassel, Germany) to calculate Krippendorff's Alpha as well as Cohens Kappa.

Results

A total of 331 slides from 33 patients were prepared for evaluation. There were 106, 112 and 113 slides from PIPAC 1, 2, and 3. All but 6 slides were rated by all 8 pathologists. Altogether, 2642 ratings were performed. The combined gradings from all pathologists at the different time points (i.e. PIPAC treatments) is shown in Table 2, demonstrating increasing frequency of lower PRGS scores from PIPAC 1 to PIPAC 3 ($p < 0.001$).

The interobserver variability for the PRGS of each quadrant biopsy (Table 3) as well as for the mean (Table 4) and maximum (Table 5) PRGS per quadrant biopsy set are given. The ICC ranged from 0.63 and 0.76, indicating a moderate to good reliability. The Krippendorff's alpha coefficients ranged from 0.60 to 0.74 regarding each quadrant biopsy score and the mean score per biopsy set (Table 3 & 4), indicating a substantial agreement. The agreement regarding the maximum PRGS per biopsy set was slightly worse, with Krippendorff's alpha coefficients ranging from 0.57 to 0.63, meaning moderate to substantial agreement. The difference between the mean PRGS per quadrant biopsy set from each single pathologist and the average mean PRGS per quadrant biopsy set from all eight pathologists' scorings is visualized in Figure 2A. Likewise, the difference between the PRGS for each

quadrant biopsy from each single pathologist and the average PRGS for each quadrant biopsy, calculated from all eight pathologists' scorings, is visualized in Figure 2B. These differences were normally distributed, and the large majority of scorings did not differ more than 0.5 PRGS from the mean values. Even though the PRGS was decreasing from PIPAC 1 to PIPAC 3, the reliability did not deteriorate over time and was not modified by increased regression. Figure 1 gives histological examples of cases where there was high agreement between the participating pathologists.

The intraobserver variability for the PRGS of each quadrant biopsy (Table 6) as well as for the mean (Table 7) and maximum (Table 8) PRGS per quadrant biopsy set are given. The ICC varied between 0.87 and 1.00, reflecting good to excellent intra-observer reproducibility. Kappa coefficients varied from 0.89 to 0.98, indicating almost perfect agreement.

Table 9 shows the interobserver variability between groups at PIPAC no. 1 and 2. We compared senior consultants (n=5) with residents (n=3) and "authors of the proposal article" with "others". In Table 10, agreement regarding the scoring of the first 33% of the biopsies at each PIPAC was compared with the agreement regarding the last 67% among all pathologists (n=8).

Discussion

In this observational, retrospective, longitudinal, single-blinded study, we found that the reproducibility of the PRGS for assessing histological response of PIPAC of PM is substantial. A total of 331 quadrant biopsies obtained from 33 patients with PM

taken at three different time points (prior to PIPAC treatment 1, 2 and 3) were evaluated. The interobserver agreement was moderate to good / substantial, and slightly better regarding the assessment of the mean PRGS per biopsy set compared to the maximum PRGS per biopsy set. When comparing the agreement between groups, residents had a slightly better agreement than senior consultants, and "others" had a slightly better agreement than the authors of the article proposing the PRGS [8]. The intraobserver agreement was good to excellent / almost perfect. We found no training effect when comparing the agreement at the first 33% percent of the scored biopsies with the remaining 67%.

The results of this study are encouraging, particularly when bearing in mind that the participating pathologists had less clinical information than in the clinical setting. First, they did not have access to immunohistochemistry. Second, they did not have access to microscopic slides from the primary tumors. Third, the participating pathologists were blinded regarding prior to which PIPAC treatment the biopsies were taken. Although the mean PRGS decreased from PIPAC 1 to PIPAC 3, there was no change in the accuracy during the course of therapy. Thus, our study supports that the PRGS is a reproducible and useful tool to assess response to intraperitoneal chemotherapy in PM.

Most regression grading systems published so far do not require complementary immunohistochemical analysis. However, immunohistochemistry is an important adjunct in routine practice of clinical pathology. In the setting of PRGS, immunohistochemistry might allow identification of isolated tumor cells in inflamed scar tissue or clusters of tumor cells in heavily inflamed tissue that could not be visualized by H&E staining. Thus, it is likely that the reproducibility of the PRGS would have been higher if the pathologists participating in the present study had had

access to immunohistochemistry. This may explain why one pathologist scored PRGS 2 instead of PRGS 1 in Figure 1A, and probably also why one pathologist scored PRGS 1 instead of PRGS 2 in Figure 1B. But also for the differentiation of PRGS 3 from PRGS 4 immunohistochemistry seems to be a useful tool, as illustrated in Figures 1E-H. We do not know whether the higher agreement between „others“ compared to agreement between the co-authors of the PRGS proposal article and, regarding the biopsies taken at PIPAC 1, the slightly higher agreement between residents compared to agreement between senior consultants means that these scores are more correct [8]. It may, however, be speculated that the pathologists primarily not related to the PRGS development used the proposed PRGS criteria more stringently and categorically, while the scoring of pathologists who were involved in the PRGS proposal may have depended a bit more on their subjective opinion.

For a long time, it has been acknowledged that the degree of histological regression may give clues to the effectiveness of chemotherapy for a given tumor. Several histological tumor regression systems (TRGs) have been developed for the quantification of response to chemotherapy of various primary and metastatic cancers. The Mandard system, developed for esophageal cancer and published in 1994, was later on used in a wide range of other primary malignancies [5]. Examples of TRG systems for rectal cancer are the Dworak (1997) and Rödel (2005) systems, and for colorectal liver metastases, the Rubbia-Brandt system (2007) [4, 6, 7]. In 2014, Trakarnsanga et al. compared the concordance indices of four different TRGs (Mandard (3- and 5-category), Dworak/Rödel (3- and 5-category), Memorial Sloan Kettering Cancer Center (3-category) and American Joint Committee on Cancer and College of American Pathologists (AJCC/CAP) (4-category) in a cohort of 563

patients with locally advanced rectal cancer and concluded that the 4-category AJCC/CAP TRG was the most accurate and should be adopted as the standard [4, 5, 7, 23-26]. For gastric cancer, six different TRG systems have been proposed so far, including the results of a recent Delphi survey [2, 27-31]. Recently, a six-tiered and a condensed three-tiered chemotherapy response score (CRS) for tuboovarian high-grade serous carcinoma after neoadjuvant chemotherapy and interval debulking surgery has been proposed and proved high reproducibility with a Kappa coefficient of 0.76 when using the condensed 3-tiered system [3, 32].

Tumor response of PM from colon cancer was explored in terms of tumor growth and histology in tumor-bearing rats treated with hyperthermic intraperitoneal chemotherapy [33]. The mean number of apoptotic cells and bodies in the entire cancer cell population was determined by counting their numbers in 5 high-power fields of non-necrotic areas. The index represented the number of visible apoptotic cancer cells in these fields. In the clinical setting, the histological response in patients with PM from colorectal cancer (n=144) treated with preoperative systemic chemotherapy was examined by determination of the percentage of viable tumor cells with respect to the area of each nodule [34]. The assessment was independent of the presence of chemotherapy-related tissue injury, fibrosis, or necrosis. In gastric cancer patients, a four-category classification system was used to examine the histologic effects of neoadjuvant bidirectional intraperitoneal-systemic chemotherapy on primary tumors and PM nodules [35].

Regardless of the approach used to quantify tumor response after neoadjuvant therapy, there is an urgent need for an objective, practical, reproducible and clinically relevant regression grading system for PM with acceptable interobserver and intraobserver variability. To our knowledge, the PRGS is the first biopsy-based

scoring system focusing on the assessment of histological response in the palliative setting in PM [8]. The fact that we included PM from a wide range of different primary malignancies, that all biopsies were taken by the same team of surgeons and processed at the same pathological department should be considered a strength. To date, the clinical value of the PRGS has not been fully elucidated, but several clinical trials using the PRGS as primary or secondary outcome are currently ongoing [21, 36]. Besides, a few studies reported a reduction of the mean PRGS after PIPAC treatment in 67-80% of the patients [17, 18, 37]. Besides, initial data indicate a trend for prognostic significance of the PRGS [37]. It is currently not known whether the mean PRGS or the maximum PRGS bears the highest clinical value.

In conclusion, our study shows that the Peritoneal Regression Grading Score (PRGS) has moderate to good / substantial interobserver variability and good to excellent / almost perfect intraobserver variability for the assessment of response to treatment of peritoneal metastasis. Our study also shows that PRGS can be used by younger pathologists without loss of accuracy. Future studies should address the prognostic and predictive role of the PRGS in peritoneal metastasis.

Acknowledgements:

We thank Pia H. Jensen, Kurt Gammelgaard Nielsen and Niels Faurskov Andersen, IT-Service Department at the University of Southern Denmark (SDU), for their help with the online platform giving the study pathologists access to the digitalized slides.

Author's contributions:

WS and SD initiated and conceptualized the study; SD coordinated the study; all authors contributed to data acquisition; SD, CF and WS analysed and interpreted the data, SD, CF and WS wrote the manuscript; all authors critically revised the manuscript; all authors approved the final version of the manuscript.

Tables

Table 1. Demographic data and baseline characteristics regarding the patients included in this study.

Age, years (range)	62 (41-85)
Sex, male / female	15 / 18
Previous treatment	
Palliative SC	
No palliative SC	3
One line palliative SC	23
Two lines palliative SC	7
> Two lines palliative SC	0
Combination PIPAC / SC	6
Primary tumor origin	
Colorectal adenocarcinoma	12*
Pancreatic adenocarcinoma	4
Gastric adenocarcinoma	4[#]
Serous ovarian adenocarcinoma	4
Appendix	4[£]
Malignant mesothelioma, epitheloid type	1
Small bowel adenocarcinoma (duodenum, jejunum)	2
Metastasis of unknown primary (MUP), adenocarcinoma	1
Extrahepatic bile ducts, adenocarcinoma	1
Peritoneal Cancer Index (PCI)	
PCI when ≥ 11 regions evaluated, mean (SD), n=26	15.5 (11.7)
PCI when < 11 regions evaluated, mean (SD), n=7	6 (4.0)
PCI, total (SD), n=33	13.5 (11.2)
Ascites volume	
0 ml	22
1 – 500 ml	5
501 – 1000 ml	3
> 1000 ml	3

* Hereof, three mucinous adenocarcinomas. # Hereof, one diffuse adenocarcinoma. £ Three low-grade appendiceal mucinous neoplasms (LAMNs) and one mucinous adenocarcinoma.

Table 2. Combined grading of the Peritoneal Regression Grading Score (PRGS) from all pathologists prior to the different PIPAC treatments. A total of 2642 scorings were performed. All grades were used at all time points, even though PRGS 4 was relative rarely used. There was an increasing frequency of lower PRGS scores from PIPAC 1 to PIPAC 3 ($p<0.001$).

PRGS	**PIPAC no. 1**	**PIPAC no. 2**	**PIPAC no. 3**	**Total**
1	306	432	488	1226
2	299	324	260	883
3	185	120	129	434
4	59	16	24	99
Total	849	892	901	2642

Table 3. Interobserver variability of the Peritoneal Regression Grading Score (PRGS) for scoring each quadrant biopsy.

8 unique raters					
Time point	**N**	**ICC**	**95% CI**	**Krippendorff Alpha**	**95% CI**
PIPAC no. 1	106	0.70	0.63-0.76	0.66	0.59-0.73
PIPAC no. 2	108	0.64	0.56-0.71	0.60	0.53-0.66
PIPAC no. 3	110	0.64	0.57-0.71	0.60	0.54-0.66

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way random-effects model.

Table 4. Interobserver variability for rating the mean Peritoneal Regression Grading Score (PRGS) per quadrant biopsy set.

8 unique raters					
Time point	**N**	**ICC**	**95% CI**	**Krippendorff Alpha**	**95% CI**
PIPAC no. 1	33	0.76	0.65-0.85	0.74	0.65-0.85
PIPAC no. 2	32	0.69	0.56-0.81	0.68	0.61-0.75
PIPAC no. 3	33	0.71	0.60-0.82	0.71	0.60-0.82

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way random-effects model.

Table 5. Interobserver variability for rating the maximum Peritoneal Regression Grading Score (PRGS) per quadrant biopsy set.

8 unique raters					
Time point	**N**	**ICC**	**95% CI**	**Krippendorff Alpha**	**95% CI**
PIPAC no. 1	33	0.65	0.52-0.77	0.59	0.43-0.76
PIPAC no. 2	32	0.68	0.54-0.80	0.63	0.54-0.71
PIPAC no. 3	33	0.63	0.50-0.76	0.57	0.47-0.67

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way random-effects model.

Table 6. Intraobserver variability for scoring the Peritoneal Regression Grading Score (PRGS) for each quadrant biopsy.

PIPAC no. 1					
Observer	**N**	**ICC**	**95% CI**	**Kappa**	**95% CI**
A	106	0.93	0.90-0.95	0.92	0.86-0.98
B	106	0.98	0.98-0.99	0.98	0.96-1.00
C	106	0.89	0.85-0.93	0.89	0.80-0.97
PIPAC no. 2					
Observer	**N**	**ICC**	**95% CI**	**Kappa**	**95% CI**
A	112	0.96	0.94-0.97	0.95	0.90-0.99
B	112	0.98	0.97-0.98	0.97	0.94-1.00
C	108	0.90	0.86-0.93	0.89	0.80-0.99
PIPAC no. 3					
Observer	**N**	**ICC**	**95% CI**	**Kappa**	**95% CI**
A	112	0.96	0.94-0.97	0.95	0.90-0.99
B	113	0.96	0.94-0.97	0.94	0.91-0.00
C	111	0.86	0.80-0.90	0.84	0.75-0.92

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way mixed-effects model.

Table 7. Intraobserver variability for scoring the mean Peritoneal Regression Grading Score (PRGS) per quadrant biopsy set.

PIPAC no. 1					
Observer	**N**	**ICC**	**95% CI**	Kappa	95% CI
A	33	0.94	0.88-0.97	0.92	0.85-0.98
B	33	1.00	1.00-1.00	0.99	0.98-1.00
C	33	0.96	0.92-0.98	0.95	0.92-1.00
PIPAC no. 2					
Observer	N	ICC	95% CI	Kappa	95% CI
A	33	0.96	0.93-0.98	0.95	0.88-1.00
B	33	1.00	1.00-1.00	0.99	0.98-1.00
C	32	0.87	0.75-0.93	0.88	0.71-1.00
PIPAC no. 3					
Observer	N	ICC	95% CI	Kappa	95% CI
A	33	0.96	0.92-0.98	0.96	0.92-1.00
B	33	0.97	0.95-0.99	0.95	0.90-1.00
C	33	0.93	0.87-0.97	0.92	0.85-0.98

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way mixed-effects model.

Table 8. Intraobserver variability for scoring the maximum PRGS per quadrant biopsy set.

PIPAC no. 1					
Observer	**N**	**ICC**	**95% CI**	**Kappa**	**95% CI**
A	33	0.96	0.92-0.98	0.95	0.88-1.00
B	33	1.0	*	1.00	1.00-1.00
C	33	0.91	0.82-0.95	0.90	0.75-1.00
PIPAC no. 2					
Observer	**N**	**ICC**	**95% CI**	**Kappa**	**95% CI**
A	33	1.0	*	1.00	1.00-1.00
B	33	0.96	0.91-0.98	0.94	0.88-1.00
C	32	0.98	0.95-0.99	0.97	0.89-1.00
PIPAC no. 3					
Observer	**N**	**ICC**	**95% CI**	**Kappa**	**95% CI**
A	33	0.93	0.87-0.97	0.91	0.81-1.00
B	33	0.96	0.93-0.98	0.95	0.89-1.00
C	33	0.85	0.72-0.92	0.82	0.67-0.96

ICC: Intraclass correlation coefficient. The given values are based on a single rater, absolute-agreement, 2-way mixed-effects model. *: complete agreement

Table 9. Interobserver variability between groups for scoring the mean PRGS per quadrant biopsy set at PIPAC no. 1 and PIPAC no. 2. The following groups were compared: Senior consultants (n=5) vs. residents (n=3) and “authors of the proposal article” (n=4) vs. “others” (n=4).

PIPAC no. 1						
Groups of observers	**Number of observers**	**N**	**ICC**	**95% CI**	**Krippendorff Alpha**	**95% CI**
Senior consultants	5	33	0.73	0.59-0.84	0.72	0.61-0.83
Residents	3	33	0.81	0.69-0.90	0.79	0.68-0.90
Proposal authors	4	33	0.71	0.54-0.83	0.69	0.58-0.81
Others	4	33	0.84	0.74-0.91	0.82	0.71-0.94
PIPAC no. 2						
Groups of observers	**Number of observers**	**N**	**ICC**	**95% CI**	**Krippendorff Alpha**	**95% CI**
Senior consultants	5	32	0.7	0.55-0.82	0.69	0.60-0.78
Residents	3	33	0.66	0.46-0.80	0.61	0.50-0.72
Proposal authors	4	33	0.66	0.50-0.80	0.66	0.57-0.74
Others	4	32	0.74	0.58-0.86	0.72	0.60-0.83

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way random-effects model.

Table 10. Training effect for scoring the mean PRGS per quadrant biopsy at PIPAC no. 1, 2 and 3. Level of agreement at each PIPAC divided between the first 33% and last 67% of the biopsies.

Order of scorings	N	ICC	95% CI	Krippendorff Alpha	95% CI
PIPAC no. 1					
First 33%	35	0.67	0.55-0.79	0.62	0.48-0.76
Last 67%	71	0.71	0.63-0.78	0.67	0.59-0.74
PIPAC no. 2					
First 33%	34	0.73	0.62-0.84	0.68	0.53-0.83
Last 67%	71	0.59	0.50-0.69	0.56	0.49-0.62
PIPAC no. 3					
First 33%	32	0.6	0.46-0.74	0.57	0.44-0.70
Last 67%	74	0.66	0.58-0.75	0.61	0.54-0.68

ICC: Intraclass correlation coefficient. The given coefficients are based on a single rater, absolute-agreement, 2-way random-effects model.

Legends

Figure 1. Histological examples of peritoneal biopsy specimens where there was relatively high (A-D) or relatively low (E-H) agreement between the participating pathologists. **A**. Peritoneal metastasis (PM) from gastric adenocarcinoma, PRGS 1. Seven pathologists scored PRGS 1 and one pathologist scored PRGS 2 (H&E). **B**. Peritoneal metastasis (PM) from colorectal mucinous adenocarcinoma, PRGS 1. Seven pathologists scored PRGS 1 and one pathologist scored PRGS 2 (H&E). **C**. PM from colorectal adenocarcinoma, PRGS 2. Seven pathologists scored PRGS 2 and one pathologist scored PRGS 1 (H&E). **D**. PM from colorectal adenocarcinoma, PRGS score 3. Seven pathologists scored PRGS 3 and one pathologist scored PRGS 2. **E.** PM from ovarian serous adenocarcinoma, high grade, PRGS 4. Four

pathologists scored PRGS 4, three pathologists scored PRGS 3 and one pathologist scored PRGS 2 (H&E). **F**. Serial section of biopsy shown in Figure 1E, with immunohistochemical (IHC) staining of Ep-CAM, highlighting the numerous cancer cells present. IHC would probably have improved the interobserver agreement in this case. **G**. PM from gastric adenocarcinoma, PRGS 4. Five pathologists scored PRGS 4 and three pathologists scored PRGS 3 (H&E). **H**. Serial section of biopsy shown in Figure 1G, with IHC of Ep-CAM, highlighting the numerous cancer cells. There are no clear-cut features of regression.

Figure 2. Interobserver variability among eight pathologists assessing the PRGS in peritoneal metastasis (PM). **A**. The difference between the mean PRGS per quadrant biopsy set from each single pathologist and the average mean PRGS per quadrant biopsy set from all eight pathologists' scorings (792 plotted values). **B**. The difference between the PRGS for each quadrant biopsy from each single pathologist and the average PRGS for each quadrant biopsy, calculated from all eight pathologists' scorings (2642 plotted values).

References

1. Langer R, Becker K. Tumor regression grading of gastrointestinal cancers after neoadjuvant therapy. *Virchows Arch.* 2018;**472**;175-186.
2. Becker K, Mueller JD, Schulmacher C *et al.* Histomorphology and grading of regression in gastric carcinoma treated with neoadjuvant chemotherapy. *Cancer* 2003;**98**;1521-1530.
3. Böhm S, Faruqi A, Said I *et al.* Chemotherapy response score: Development and validation of a system to quantify histopathologic response to neoadjuvant chemotherapy in tubo-ovarian high-grade serous carcinoma. *J. Clin. Oncol.* 2015;**33**;2457-2463.
4. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int. J. Colorectal Dis.* 1997;**12**;19-23.
5. Mandard AM, Dalibard F, Mandard JC *et al.* Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994;**73**;2680-2686.
6. Rubbia-Brandt L, Giostra E, Brezault C *et al.* Importance of histological tumor response assessment in predicting the outcome in patients with colorectal liver metastases treated with neo-adjuvant chemotherapy followed by liver surgery. *Ann. Oncol.* 2007;**18**;299-304.
7. Rödel C, Martus P, Papadoupolos T *et al.* Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J. Clin. Oncol.* 2005;**23**;8688-8696.
8. Solass W, Sempoux C, Detlefsen S, Carr NJ, Bibeau F. Peritoneal sampling and histological assessment of therapeutic response in peritoneal metastasis:

Proposal of the peritoneal regression grading score (prgs). *Pleura Peritoneum* 2016;**1**;99-107.

9. Solass W, Kerb R, Murdter T *et al.* Intraperitoneal chemotherapy of peritoneal carcinomatosis using pressurized aerosol as an alternative to liquid solution: First evidence for efficacy. *Ann. Surg. Oncol.* 2014;**21**;553-559.
10. Solass W, Herbette A, Schwarz T *et al.* Therapeutic approach of human peritoneal carcinomatosis with dbait in combination with capnoperitoneum: Proof of concept. *Surg. Endosc.* 2012;**26**;847-852.
11. Solass W, Hetzel A, Nadiradze G, Sagynaliev E, Reymond MA. Description of a novel approach for intraperitoneal drug delivery and the related device. *Surg. Endosc.* 2012;**26**;1849-1855.
12. Dueckelmann AM, Fink D, Harter P *et al.* The use of pipac (pressurized intraperitoneal aerosol chemotherapy) in gynecological oncology: A statement by the german "arbeitsgemeinschaft gynaekologische onkologie studiengruppe ovarialkarzinom (ago-ovar)", the swiss and austrian ago, and the north-eastern german society of gynaecologic oncology. *Arch. Gynecol. Obstet.* 2018;**297**;837-846.
13. Grass F, Vuagniaux A, Teixeira-Farinha H, Lehmann K, Demartines N, Hubner M. Systematic review of pressurized intraperitoneal aerosol chemotherapy for the treatment of advanced peritoneal carcinomatosis. *Br. J. Surg.* 2017;**104**;669-678.
14. Blanco A, Giger-Pabst U, Solass W, Zieren J, Reymond MA. Renal and hepatic toxicities after pressurized intraperitoneal aerosol chemotherapy (PIPAC). *Ann. Surg. Oncol.* 2013;**20**;2311-2316.

15. Demtröder C, Solass W, Zieren J, Strumberg D, Giger-Pabst U, Reymond MA. Pressurized intraperitoneal aerosol chemotherapy with oxaliplatin in colorectal peritoneal metastasis. *Colorectal Dis.* 2016;**18**;364-371.
16. Odendahl K, Solass W, Demtroder C *et al.* Quality of life of patients with end-stage peritoneal metastasis treated with pressurized intraperitoneal aerosol chemotherapy (pipac). *Eur. J. Surg. Oncol.* 2015;**41**;1379-1385.
17. Graversen M, Detlefsen S, Bjerregaard JK, Pfeiffer P, Mortensen MB. Peritoneal metastasis from pancreatic cancer treated with pressurized intraperitoneal aerosol chemotherapy (PIPAC). *Clin. Exp. Metastasis* 2017;**34**;309-314.
18. Graversen M, Detlefsen S, Bjerregaard JK, Fristrup CW, Pfeiffer P, Mortensen MB. Prospective, single-center implementation and response evaluation of pressurized intraperitoneal aerosol chemotherapy (PIPAC) for peritoneal metastasis. *Ther. Adv. Med. Oncol.* 2018;**10**;1-11.
19. Khosrawipour T, Khosrawipour V, Giger-Pabst U. Pressurized intra peritoneal aerosol chemotherapy in patients suffering from peritoneal carcinomatosis of pancreatic adenocarcinoma. *PLoS One* 2017;**12**;e0186709.
20. Somashekhar SP, Rajagopal AK, Zaveri SS, Chandrashekhar RK, Rauthan A, Rakshit SH. First indian study on pressurized intraperitoneal aerosol chemotherapy (pipac) procedure for advanced peritoneal carcinomatosis secondary to epithelial ovarian cancer. *Indian Journal of Gynecologic Oncology* 2018;**16**.
21. Graversen M, Detlefsen S, Asmussen J *et al.* Treatment of peritoneal carcinomatosis with pressurized intraperitoneal aerosol chemotherapy PIPAC-OPC-2. *Pleura Peritoneum* 2018;**3**; 20180108.

22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 2016;**15**;155-163.
23. Jäger T, Neureiter D, Urbas R *et al.* Applicability of american joint committee on cancer and college of american pathologists regression grading system in rectal cancer. *Dis. Colon Rectum* 2017;**60**;815-826.
24. Mace AG, Pai RK, Stocchi L, Kalady MF. American joint committee on cancer and college of american pathologists regression grade: A new prognostic factor in rectal cancer. *Dis. Colon Rectum* 2015;**58**;32-44.
25. Quah HM, Chou JF, Gonen M *et al.* Pathologic stage is most prognostic of disease-free survival in locally advanced rectal cancer patients after preoperative chemoradiation. *Cancer* 2008;**113**;57-64.
26. Trakarnsanga A, Gonen M, Shia J *et al.* Comparison of tumor regression grade systems for locally advanced rectal cancer after multimodality treatment. *J. Natl. Cancer Inst.* 2014;**106**.
27. Association JGC. Japanese classification of gastric carcinoma: 3rd english edition. *Gastric Cancer* 2011;**14**;101-112.
28. Charalampakis N, Nogueras Gonzalez GM, Elimova E *et al.* The proportion of signet ring cell component in patients with localized gastric adenocarcinoma correlates with the degree of response to pre-operative chemoradiation. *Oncology* 2016;**90**;239-247.
29. Kiyabu M, Leichman L, Chandrasoma P. Effects of preoperative chemotherapy on gastric adenocarcinomas. A morphologic study of 25 cases. *Cancer* 1992;**70**;2239-2245.

30. Mansour JC, Tang L, Shah M *et al.* Does graded histologic response after neoadjuvant chemotherapy predict survival for completely resected gastric cancer? *Ann. Surg. Oncol.* 2007;**14**;3412-3418.
31. Tsekrekos A, Detlefsen S, Riddell R *et al.* Histopathologic tumor regression grading in patients with gastric carcinoma submitted to neoadjuvant treatment: Results of a delphi survey. *Hum. Pathol.* 2019;**84**;26-34.
32. Said I, Böhm S, Beasley J *et al.* The chemotherapy response score (crs): Interobserver reproducibility in a simple and prognostically relevant system for reporting the histologic response to neoadjuvant chemotherapy in tuboovarian high-grade serous carcinoma. *Int. J. Gynecol. Pathol.* 2017;**36**;172-179.
33. Pelz JO, Doerfer J, Dimmler A, Hohenberger W, Meyer T. Histological response of peritoneal carcinomatosis after hyperthermic intraperitoneal chemoperfusion (hipec) in experimental investigations. *BMC Cancer* 2006;**6**;162.
34. Passot G, You B, Boschetti G *et al.* Pathological response to neoadjuvant chemotherapy: A new prognosis tool for the curative management of peritoneal colorectal carcinomatosis. *Ann. Surg. Oncol.* 2014;**21**;2608-2614.
35. Yonemura Y, Elnemr A, Endou Y *et al.* Effects of neoadjuvant intraperitoneal/systemic chemotherapy (bidirectional chemotherapy) for the treatment of patients with peritoneal metastasis from gastric cancer. *Int. J. Surg. Oncol.* 2012;148420.
36. Bakrin N, Tempfer C, Scambia G *et al.* Pipac-ov3: A multicenter, open-label, randomized, two-arm phase iii trial of the effect on progression-free survival of cisplatin and doxorubicin as pressurized intra-peritoneal aerosol chemotherapy (pipac) vs. Chemotherapy alone in patients with platinum-resistant recurrent

epithelial ovarian, fallopian tube or primary peritoneal cancer. *Pleura Peritoneum* 2018;**3**;20180014.

37. Solass W, Kurtz F, Struller F *et al.* Prognostic value of peritoneal grading regression score (PRGS) in peritoneal metastasis: An exploratory analysis on registry data. [Published abstract. *Pleura Peritoneum* 2018; Special Suppl **S1**, sA27].