HISTOLOGIC RESPONSE GRADING AFTER CHEMORADIATION IN LOCALLY ADVANCED RECTAL CANCER: A PROPOSAL FOR STANDARDIZED REPORTING

ROB GLYNNE-JONES, FRCR., FRCP., AND NICOLA ANYEMENE, FRCR., MRCP.

Centre for Cancer Treatment, Mount Vernon Hospital, Northwood, Middlesex, UK

In a previous editorial, we criticized pathological complete response (pCR) as an endpoint in rectal cancer after neoadjuvant chemoradiation, because only a small proportion of patients (10–20%) are captured by a pCR. We also pointed out the confusion as to what constitutes a pCR: ypT0 alone, ypT0 plus ypTmic, or ypT0 plus ypN0. Further confusion arises from authors who define a pCR as “the presence of residual cells, which are considered non-viable” (1), and “no evidence of viable malignant cells in the specimen or the presence of scattered isolated malignant cells in a background of fibrosis” (2).

The evaluation of clinical response is recognized to be inaccurate. Standard response criteria in terms of size according to the World Health Organization or Response Evaluation Criteria in Solid Tumors criteria do not take into account minor responses. There is poor differentiation on imaging of residual areas within the radiation field to distinguish viable tumor from replacement with fibrosis, and clinical response has not been shown to act as a robust surrogate endpoint to predict outcome in this setting. If we had an objective, practical and clinically relevant regression classification system, which encompassed both the degree of damage sustained by the tumor and the residual tumor burden in primary and nodes, then we could precisely compare the efficacy of various treatment protocols—either in terms of the cytotoxic drugs or radiation dose. We could also expect to squeeze more relevance from a molecular response prediction.

Alternative strategies to categorize treatment effects from radiation and chemoradiation describe the observed histologic changes in the resected specimen. Measures such as T-microscopic (Tmic), and tumor regression grades (TRGs), define other less-marked levels of response, apart from pCR, although they are not yet recognized in the American Joint Committee on Cancer staging recommendations. Different authors have used TRG systems in rectal cancer, which were developed and based on appearances after radiotherapy alone (3), concurrent chemoradiation in oesophageal cancer (4), and chemotherapy alone in gastric cancer (5).

These histopathologic appearances have been poorly documented, and a standardized approach to characterizing these changes is lacking. Yet their use is increasing. Only one randomized trial (6) has prospectively assessed histologic tumor response. We have found only a single previous review of tumor regression grading, and no Cochrane systematic reviews or Cochrane protocols in development. This brief editorial aims to focus on the lack of consistency regarding the terms Tmic and TRGs, despite their increasing use.

TMIC

A few microscopic foci of cancer outside the muscularis propria have been categorized as Tmic rather than ypT3 (7). Often such cells appear only in a single block. Yet there is no robust evidence that the outcome for pCR, ypTmic, or ypT3 is different.

The original definition of Tmic categorized “isolated foci of microscopic tumor (usually less than 3)” (1). Since then, others have defined Tmic as “rare isolated residual cancer cells”; “≤20% residual cancer cells in the previously tumor involved rectal area”; “microfoci only of tumor remaining (near pCR)”; “a minimal microscopic residuum”; “microscopic residues”; or “minimal microfocal residual disease.” Some authors use the term Tmic, but fail to define it at all.

RESIDUAL CELL DENSITY AND TRG

The various TRG systems are pathologic evaluations based on the relative amount of tumor cells present, and

Reprint requests: Dr. Rob Glynne-Jones, Mount Vernon Centre for Cancer Treatment, Northwood, Middlesex, United Kingdom HA6 2RN. Tel: 0044 1923 844767; Fax: 0044 1923 844960; E-mail: Rob.glynnejones@nhs.net

Honoraria received from Roche, Sanofi-Aventis, Merck Serono, and Pfizer for attendance at advisory boards, speakers’ bureau, and support received to attend international meetings (all less than $10,000) (R.G.-J.).

Conflict of interest: none.

Received June 17, 2008, and in revised form Oct 2, 2008. Accepted for publication Oct 19, 2008.
the extent of the fibrosis/desmoplastic reaction in the stroma, and also describe morphologic changes in cells such as nuclear pleomorphism and cytoplasmic vacuolation. Early TRGs estimated the intramural and adventitial cell kill as minimal, moderate, marked, or complete. Other investigators have apparently measured the percentage of residual tumor in the specimen and lymph nodes or have tried to grade tumor necrosis as a percentage of the tumor mass.

In esophageal cancers, a five-point scale based the grades on residual tumor cells and the extent of fibrosis (4). The most favorable and lowest TRG represents pCR. This system also captures the remaining 85% of patients in different categories up to the absence of any histologic changes.

Simplified variations with three or four grades have also been employed and have different cutoffs (8, 9). These systems each claim to be more easily reproducible, but there is no internationally accepted system of classification. The definitions used in the different studies vary and it is acknowledged that reproducibility is poor (10). A pCR can be defined as TRG0, TRG1a, or TRG5, so studies cannot be compared. Some pathologists when blinded have expressed difficulty in distinguishing between TRG4 and TRG5 in unirradiated patients.

It remains controversial whether lymph nodes can be graded in a similar fashion as the primary tumor. Histologic regression has been documented in lymph nodes by the residual presence of mucin deposits, and tumor regression grades in mesorectal lymph nodes have been correlated with the primary tumor.

Retrospective studies have suggested that on multivariate analysis tumor regression grade was an independent prognostic indicator for long-term local tumor control (3). A regression grading system used in the German CAO/ARO/AIO-94 (the German Study) protocol suggests that the degree of regression at 6 weeks influences subsequent outcome, but there is no robust evidence that the outcomes for TRG1 (i.e., pCR) and TRG2 are different (9).

Despite these disadvantages and lack of validation, TRGs are used as a measure of efficacy to determine whether a local excision after chemoradiation is sufficient or radical surgery should be performed and to discriminate between the efficacy of different neoadjuvant regimens. Finally, recent authors have attempted to correlate TRGs with metabolic response on positron emission tomography/computed tomography or genetic polymorphisms as a predictor of response.

It is vital that tumor regression grading is a practical tool, and so we need standardized sampling methods across the residual tumor. Current standard histologic practice only examines a very small part of the tumor, even when the whole suspect area is embedded. Hence, there should be a standard sampling procedure, and histologic regression should be assessed not simply on the central section of each tumor, but there should be an agreed framework to deal with different appearances in different parts of the tumor (central vs. peripheral; primary vs. lymph nodes).

**OUR RECOMMENDATION**

The macroscopically visible tumor or suspected areas should be localized, measured, and specimens completely embedded in paraffin. Three levels should be taken and examined from each tumor block. Serial 5-μm sections are stained with hematoxylin and eosin for evaluation of radiochemotherapy-induced regression. In case of diagnostic uncertainties, further staining procedures or immunohistochemical analysis for cytokine expression can be performed to identify individual cells. Studies have shown a significant association between regression grades and ypN categories, highlighting that regression should not be restricted to the primary tumor; therefore, lymph nodes should be extensively scrutinized.

We therefore recommend a four-point scale called tumor response grade, which is defined with a standard time frame of 6–8 weeks postsurgery. Grades 1, 2, and 3 could be subdivided into A (ypN0) and B (ypN1-2). Ideally, 0 and 3 would account for 15–20% each, and 1 and 2, respectively, 30% and 35% each.

0 = Complete histomorphologic regression (i.e., pT0, ypNO).
1 = Major histomorphologic regression with few hard to find scattered microscopic foci <2 mm (i.e., <10% residual tumor).
2 = Minor histomorphologic regression with fibrosis outweighing residual cancer cells.
3 = Minimal histomorphologic regression with no/negligible evidence of any tumor response.

There is no published information on the extent of inter- and intraobserver variability in defining TRGs, although it is acknowledged that reproducibility is poor. The use of systems that were based on the different appearances after radiotherapy, concurrent chemoradiation, and chemotherapy are unlikely to deliver consistency—particularly if these systems were developed in esophageal or gastric cancer. In addition, the inherent heterogeneity of response in different parts of the tumor, between primary and nodes, and varying quality assurance employed, limit the utility of TRGs.

The terms Tmic and TRGs remain invalidated surrogate endpoints, and the lack of consistency in their reporting, hinders their interpretation as a measure of response within rectal cancer trials. This is a chicken and egg situation, because TRGs are unlikely to be adopted by agencies such as the American Joint Committee on Cancer until they have been validated as having prognostic import. It would be helpful if, in future, pathologists recognize the use of a single four-point scale for TRGs as a standard. International consensus is required, as clarity, consistency, and standardization are urgently needed. Pathologists could use an Internet-based teaching aid such as http://www.virtualpathology.leeds.ac.uk/teaching.php, where pathologists could visit for a virtual comparison to clarify the assessment of each grade. We then need to test this system on multiple datasets to validate the predictive power in terms of outcome, and whether it is reproducible in the wider setting.
REFERENCES