

# Interobserver agreement of the Paris and simplified classifications of superficial colonic lesions: a Western study



## Authors

Francesco Cocomazzi<sup>1,2</sup>, Marco Gentile<sup>1</sup>, Francesco Perri<sup>1</sup>, Antonio Merla<sup>1</sup>, Fabrizio Bossa<sup>1</sup>, Mariano Piazzolla<sup>2,1</sup>, Antonio Ippolito<sup>1</sup>, Fulvia Terracciano<sup>1</sup>, Arcangela Patrizia Giuliani<sup>1</sup>, Rossella Cubisino<sup>1</sup>, Antonella Marra<sup>1</sup>, Sonia Carparelli<sup>1</sup>, Alessia Mileti<sup>2,1</sup>, Rosa Paolillo<sup>2,1</sup>, Andrea Fontana<sup>3</sup>, Massimiliano Copetti<sup>3</sup>, Alfredo Di Leo<sup>2</sup>, Angelo Andriulli<sup>1</sup>

## Institutions

- 1 Fondazione “Casa Sollievo della Sofferenza”, IRCCS, Gastroenterology and Endoscopy Units, San Giovanni Rotondo, Foggia, Italy
- 2 University of Bari, Section of Gastroenterology, Department of Emergency and Organ Transplantation, Bari, Italy
- 3 Unit of Biostatistics, Fondazione “Casa Sollievo della Sofferenza”, IRCCS, San Giovanni Rotondo, Foggia, Italy

submitted 17.8.2020

accepted after revision 9.12.2020

## Bibliography

Endoscopy International Open 2021; 09: E388–E394

DOI 10.1055/a-1352-3437

ISSN 2364-3722

© 2021. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,  
70469 Stuttgart, Germany

## Corresponding author

Marco Gentile, MD, Gastroenterology and Endoscopy,  
Fondazione “Casa Sollievo Sofferenza”, IRCCS, viale  
Cappuccini 1, 75013 San Giovanni Rotondo, Italy  
Fax: + 39 088 2410784  
[m.gentile@operapadrepio.it](mailto:m.gentile@operapadrepio.it)

## ABSTRACT

**Background and study aims** The Paris classification of superficial colonic lesions has been widely adopted, but a simplified description that subgroups the shape into pedunculated, sessile/flat and depressed lesions has been proposed recently. The aim of this study was to evaluate the accuracy and inter-rater agreement among 13 Western endoscopists for the two classification systems.

**Methods** Seventy video clips of superficial colonic lesions were classified according to the two classifications, and their size estimated. The interobserver agreement for each classification was assessed using both Cohen  $\kappa$  and AC1 statistics. Accuracy was taken as the concordance between the standard morphology definition and that made by participants. Sensitivity analyses investigated agreement between trainees (T) and staff members (SM), simple or mixed lesions, distinct lesion phenotypes, and for laterally spreading tumors (LSTs).

**Results** Overall, the interobserver agreement for the Paris classification was substantial ( $\kappa=0.61$ ; AC1=0.66), with 79.3% accuracy. Between SM and T, the values were superimposable. For size estimation, the agreement was 0.48 by the  $\kappa$ -value, and 0.50 by AC1. For single or mixed lesions,  $\kappa$ -values were 0.60 and 0.43, respectively; corresponding AC1 values were 0.68 and 0.57. Evaluating the several different polyp subtypes separately, agreement differed significantly when analyzed by the  $\kappa$ -statistics (0.08–0.12) or the AC1 statistics (0.59–0.71). Analyses of LSTs provided a  $\kappa$ -value of 0.50 and an AC1 score of 0.62, with 77.6% accuracy. The simplified classification outperformed the Paris classification:  $\kappa=0.68$ , AC1=0.82, accuracy=91.6%.

**Conclusions** Agreement is often measured with Cohen's  $\kappa$ , but we documented higher levels of agreement when analyzed with the AC1 statistic. The level of agreement was substantial for the Paris classification, and almost perfect for the simplified system.

## Introduction

Colorectal cancer (CRC) is considered to originate from adenomatous polyps, which phenotypically may appear as pedunculated or sessile. A non-polypoid shape of adenomas also has been recognized more recently, which can also develop into CRC [1].

Superficial colonic lesions are notable for the wide range of morphologic phenotypes, as they may appear as polypoid, flat/depressed or excavated tumors. In addition, mixed lesions are also evident as one subtype may present features of more than one type. Simple or mixed flat lesions, at least 10 mm in diameter, are labelled laterally spreading tumors (LSTs) and divided into four phenotypes, according to granular or nongranular, homogeneous or nonhomogeneous endoscopic appearance [1,2]. The Paris classification, which ensures awareness of subtle differences in the macroscopic subtypes of superficial neoplasms [2,3], is the most used international classification system to report polyp shape and it recently has been endorsed by professional societies [4–6]. Its adoption is an essential quality indicator for endoscopy practice. A full understanding of the Paris classification has several clinical meanings: first, it may assist in determining a minimal standard terminology, which would help reduce subjectivity in the description of lesions between observers; second, it has relevant implications because CRC prevalence is extremely low in some subclasses, but may reach 50% in other subtypes; finally, it provides information likely to guide both polyp management and post-resection surveillance [1–5,7].

Relying so heavily on the Paris classification would ensure an adequate level of agreement between raters as it would support confidence in the diagnoses being made. Few reports verified the interobserver- and/or intra-observer validity of this system. In a recent study, the interobserver agreement between Western endoscopists was only moderate ( $\kappa=0.42$ ) and pairwise agreement before and after training was also low (60%–67%) [8]. Reassuringly, better performance was credited by a South Korean study, where  $\kappa$ -values of 0.533 to 0.713 and accuracy values of 0.715 to 0.846 were scored by expert endoscopists in the pre-training and post-training tests, respectively [9]. In another study, these parameters were also evaluated in difficult-to-define settings, such as complex/mixed polyps [10]: an accuracy value of 66.0% and moderate inter-rater agreement ( $\kappa=0.48$ ) was scored by American specialists in complex polypectomy. Lee et al [11] classified the LSTs into four categories, as suggested by the Kyoto consensus workshop [1]: accuracy values of 0.859 and  $\kappa$ -values of 0.730, respectively, were reported by expert South Korean endoscopists. The four LST categories also may be derived by the Paris classification, but currently no study has reported agreement for LSTs classified according to this system [1,12]. As a general observation, lower values were scored by either trainees or even specialists with lower competence in complex polypectomy [8–11].

In 2002, when the Paris classification was issued by an ad hoc conference, the intent was “to explore the utility and clinical relevance of the Japanese endoscopic classification of superficial

neoplastic lesions of the GI tract.” The intent was to reverse the opinion of Western colonoscopists, who considered the Japanese classification too complex for practical use. Since then, the Paris classification has been endorsed by international societies [4–6] and widely adopted. However, as previously mentioned, available evidence still documents the persistence of difficulties in the inter-rater observation of some endoscopic morphologic features [8–10]. Owing to the wide variation in rater classification according to the Paris system, a simplified description of polyp morphology recently has been proposed, which has three broad categories for shape: pedunculated, sessile/flat (elevated), and depressed lesions [8]. We acknowledge the limited verification of the Paris classification, as only two classification exercises done by Western endoscopists have been carried out so far [8,10]. In addition, the performance of the suggested simplified system has not yet undergone objective evaluation.

We performed a study in which 13 Western gastroenterologists with variable expertise in colonoscopy classified superficial colorectal lesions according to the Paris classification. The aim was to evaluate interobserver agreement and accuracy for this classification system and to determine the effectiveness of a training module for both trainees (Ts) and staff members (SMs). The secondary aim was to assess the the same parameters using the new simplified classification system, as suggested by Van Doorn et al [8].

## Materials and methods

This study was carried out in the Division of Gastroenterology & Endoscopy of the Fondazione “Casa Sollievo della Sofferenza,” IRCCS, in San Giovanni Rotondo, Italy. The Division serves as a teaching unit for the Postgraduate School of Gastroenterology of the University of Bari, Italy. We conducted an observational study of inter-rater reliability performed in accordance with the guidelines for reporting reliability and agreement studies [13]. Thirteen investigators, seven SMs and six Ts, were involved in the study. The SMs each had performed at least 1,000 colonoscopies and two of them were specialists in complex polypectomy; each T had an initial experience with at least 200 colonoscopies.

### Pre-study training

All investigators were initially provided with relevant literature on the topic and attended a 1-hour conference at which the Paris classification was fully elucidated (the first learning phase). Subsequently, a set of 25 endoscopic pictures of superficial lesions, retrieved from the illustrations accompanying available literature, was electronically sent to the observers in a PowerPoint file, preceded by a summary of the classification. The class subtypes of the neoplasms, reported in the legends for these images, served as the reference standard for the “correct” classification. Respondents were blinded to the legend accompanying the retrieved images and had to assess the lesion characteristics using the Paris classification; in addition, to ensure an unbiased review of the pictures, the order in which they were numbered differed from one to another observer. After

receiving the individual response, each rater was made aware of the “correct” classification. A final meeting with all participants was organized to address questions about mistaken attribution of individual images (the second learning phase).

### Study design (video clip evaluation process)

For the post-training study, we used videos of colonoscopies that were recorded previously in our Endoscopic Unit using forward-viewing instruments (CF-Q 180, CF-H 185, CF-H 190 and CF-HQ 190, Olympus Medical Systems, Tokyo, Japan). After selecting 70 high-quality records and viewing the full-length videos, short clips varying in length from 10 seconds to 4 minutes and showing polyps were created and sent with a Google Drive link to the participants. Patients and the histopathology of lesions remained unknown to the observers. Investigators were allowed to watch the video as many times as they preferred, and asked to classify the 70 lesions as polypoid or non-polypoid, simple (Ip, lsp, ls, Ila, Ilb, Ilc, and III) or mixed (e.g. Ila+ls and Ila+Ilc). Answers were sent to the study coordinator in an Excel file. Because there is no standard definition of polyp morphology, the “correct” one was set through discussion between the best performing operator in the pre-study training (100% performance) and the study coordinator. An estimate of the diameter of the single lesion was also required: diminutive (<6 mm), small (6–9 mm), or large (>9 mm). Once the classification was returned by all endoscopists, answers were kept confidential and a feedback form showing the correct classification was sent to each of them. Finally, to evaluate the performance of the simplified classification as proposed by Van Doorn [8], we considered pedunculated polyps the categories Ip and lsp in the Paris classification, elevated the ls, Ila, Ilb and Ila+ls categories, and depressed the Ilc, Ila+Ilc and ls+Ilc Paris categories.

### Outcomes

The main outcome of the study was evaluation of inter-rater agreement of the Paris and simplified classifications of superficial colonic lesions, after a training program. The level of agreement was also evaluated for different size lesions. Several sensitivity analyses were pre-planned to investigate the agreement: 1) between Ts and SMs; 2) for simple or mixed lesions; 3) for each Paris subtype; and 4) for LSTs using the Paris Classification. In addition, with the intent to verify the usefulness of pre-study training, interobserver agreement was assessed for the 25 images. Finally, accuracy analyses of the correct classification also were performed.

### Statistical analysis

Interobserver agreement was estimated using the kappa coefficient ( $\kappa$ ). To overcome a potential kappa paradox [14–17], we also assessed the agreement using Gwet’s AC1 coefficient and 95% confidence intervals (95%CI) were considered. The overall classification accuracy was measured by percentage of correct morphology classifications provided by the study participants, assuming that those provided by the experts were the gold standard. Moreover, we evaluated the classification accuracy for each individual observer.

**► Table 1** Interobserver agreement ( $\kappa$ - and AC1-values with 95% confidence intervals) for the Paris classification of 25 still images of colonic superficial lesions.

Raters	Kappa	95% CI	AC1	95% CI
All	0.54	0.43–0.65	0.60	0.50–0.70
SM	0.47	0.34–0.60	0.53	0.42–0.65
T	0.63	0.50–0.77	0.68	0.55–0.81

CI, confidence interval; SM, staff members; T, trainees

All statistical analyses were performed using SAS Software Release 9.4 (SAS Institute, Cary, North Carolina, United States).

## Results

### Pre-study training: photographs evaluation

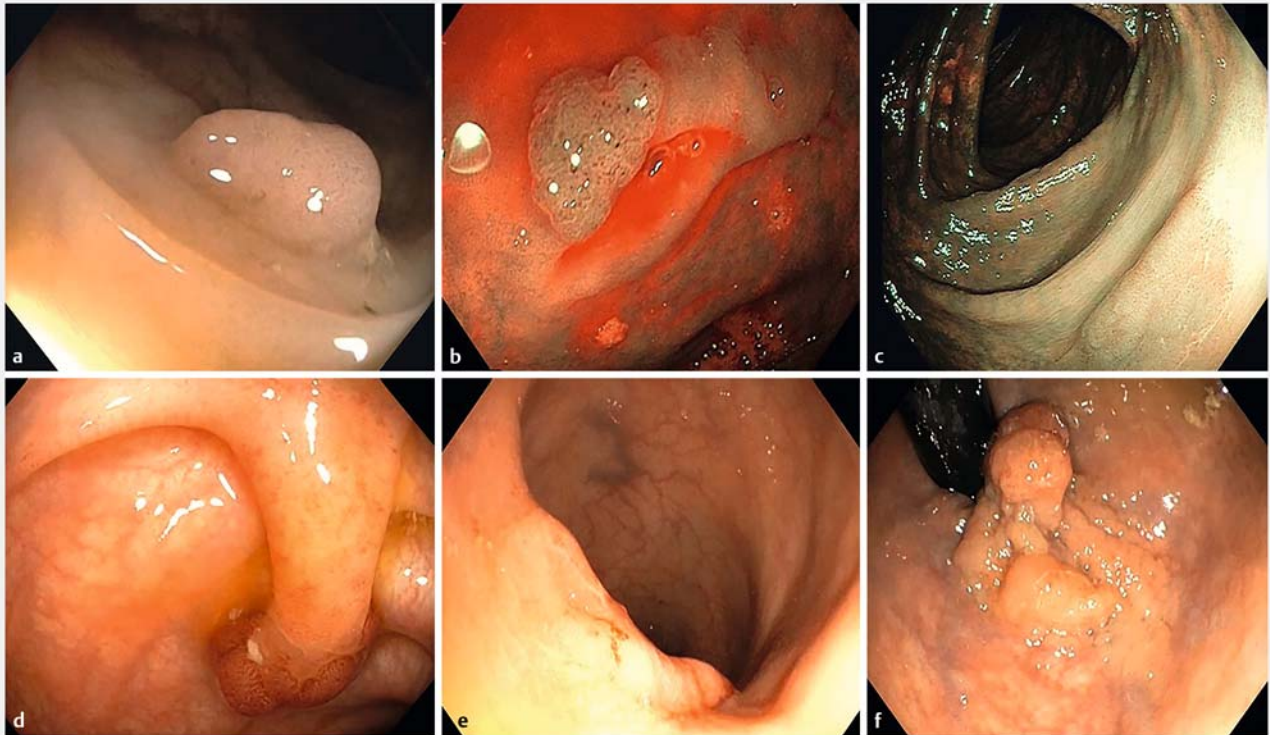
The 25 still images of colonic neoplasms showed 21 simple lesions (4 0-ls, 3 0-lp, 9 0-Ila, 2 0-Ilb and 3 0-Ilc) and four mixed lesions (3 0-Ila+ls and 1 0-Ila+Ilc). The interobserver agreement among the 13 observers for the Paris classification is shown in ► **Table 1**. Data document a moderate level of agreement between raters with a Cohen  $\kappa$ -value of 0.54 (95% CI: 0.43–0.65); a higher  $\kappa$ -value was scored by the six Ts (0.63, 95% CI: 0.50–0.77) as compared to 0.47 (95% CI: 0.34–0.60) for the seven SMs. Corresponding Gwet’s AC1 values amounted to 0.60 (95% CI: 0.50–0.70) for the 13 raters, 0.53 (95% CI: 0.42–0.65) for SMs and 0.68 (95% CI: 0.55–0.81) for Ts. Because the standard “correct” classification was derived from original articles from which these images were retrieved, the accuracy in correct classification amounted to 72% for the 13 observers, 74% for Ts, and 70% for SMs.

### Video clip evaluation

#### The Paris Classification

The 70 video clips referred to 54 single and 16 mixed lesions. Examples of their features are shown in ► **Fig. 1**. The single lesions were defined as 0-ls (no. = 24), 0-lsp (no. = 2), 0-lp (no. = 7), 0-Ila (no. = 18), 0-Ilb (no. = 2), and 0-Ilc (no. = 1). Of the 16 mixed lesions, eight were classified as 0-Ila+ls, seven as 0-Ila+Ilc, and one as 0-ls+Ilc. The inter-rater agreement for the Paris classification is shown in ► **Table 2**. The level was substantial at both the Cohen  $\kappa$ -value (0.61, 95% CI: 0.55–0.67) and the Gwet’s AC1 value (0.66, 95% CI: 0.60–0.71). Because it did not differ between SMs and Ts, all successive results refer to the rates for the 13 endoscopists.

We ran further sensitivity analyses to evaluate the interobserver agreement for single or mixed lesions and distinct polyp phenotypes and LSTs; the results are shown in ► **Table 2**. The first sub-analysis referred to the polyp phenotypes: the  $\kappa$ -value for single lesions (independently from their morphologic subtypes) was 0.60 (95% CI: 0.53–0.67) and 0.43 (95% CI: 0.32–0.54) for mixed lesions; corresponding values with the Gwet’s AC1 statistics were 0.68 (95% CI: 0.62–0.74) and 0.57 (95% CI: 0.45–0.70), respectively. The successive analysis took into ac-



► **Fig. 1** Morphology examples (video stills). **a** 0-Is polyp; **b** 0-IIa lesion (characterized by means of NBI); **c** 0-IIb lesion (characterized by means of NBI); **d** 0-Ip polyp; **e** 0-IIa + IIc laterally spreading lesion; **f** 0-IIa + Is laterally spreading lesion.

► **Table 2** Interobserver agreement ( $\kappa$ - and AC1-values with 95% confidence intervals) for the Paris classification of 70 video clips of colonic superficial lesions.

Design	Raters	Kappa	95% CI	AC1	95% CI
All	All	0.61	0.55–0.67	0.66	0.60–0.71
	SM	0.61	0.54–0.69	0.66	0.59–0.73
	T	0.59	0.51–0.67	0.64	0.58–0.71
Dimension	All	0.48	0.38–0.58	0.50	0.39–0.60
Simple	All	0.60	0.53–0.67	0.68	0.62–0.74
Mixed	All	0.43	0.32–0.54	0.57	0.45–0.70
SUBTYPE Is	All	0.08	0.03–0.12	0.71	0.63–0.80
Ila	All	0.12	0.04–0.21	0.67	0.57–0.78
Ila + Is	All	0.12	0.03–0.20	0.63	0.44–0.83
Ila + IIc	All	0.09	0.02–0.15	0.59	0.44–0.73
LSTs	All	0.50	0.38–0.61	0.62	0.53–0.71

CI, confidence interval; SM, staff members; T, trainees; LSTs, laterally spreading tumors.

count the single categories of the Paris classification and was limited to the four most common shapes (i. e. Is, Ila, Ila + Is and Ila + IIc). As indicated in ► **Table 2**, the Cohen's  $\kappa$ -values for each subtype ranged from 0.08 to 0.12, all pointing toward a slight agreement according to Landis and Koch [18], whereas corresponding values with the Gwet's statistics scored in the range

of 0.59 to 0.71, indicating substantial agreement. When the analysis was restricted to the 23 LSTs (9 0-IIa, 7 0-IIa + Is, 7 0-IIa + IIc), the level of inter-rater agreement was moderate at the Cohen's  $\kappa$  statistics (0.50, 95% CI: 0.38–0.61) and substantial at the Gwet's analysis (0.62, 95% CI: 0.53–0.71). The last sub-analysis was for verification of the agreement for evaluation of

► **Table 3** Interobserver agreement ( $\kappa$  and AC1 values with 95% confidence intervals) for the simplified classification of 70 video clips of colonic superficial lesions.

Design	Raters	Kappa	95% CI	AC1	95% CI
All	All	0.68	0.58–0.78	0.82	0.77–0.88
Elevated	All	0.10	0.05–0.15	0.88	0.83–0.93
Pedunculated	All	0.01	–0.04–0.06	0.93	0.85–1.00
Depressed	All	0.03	–0.06–0.12	0.47	0.21–0.72

CI, confidence interval; SM, staff members; T, trainees.

the size of the lesions: the level was moderate with both the  $\kappa$  (0.48, 95% CI: 0.38–0.58), and to AC1 statistics (0.50, 95% CI: 0.39–0.60).

### Simplified classification

According to previous reports about the limits of the Paris Classification in routine practice [8, 10, 19], considering the specific value of some subtypes in prognosis and therapeutic choice (e.g. pit-pattern Vi in depressed area) [20, 21], and trying to derive an easy-to-use morphological classification, we evaluated the performance of the simplified classification based on only three categories: nine pedunculated (Ip and lsp), 52 elevated (Is, Ila, Ilb and Ila + Is), and nine depressed (Ilc, Ila + Ilc and Is + Ilc) lesions. The results are shown in ► **Table 3**. By using this simplified system, the interobserver agreement amounted to 0.68 (95% CI: 0.58–0.78) at the Cohen's  $\kappa$ -value analysis and to 0.82 (95% CI: 0.77–0.88) with the Gwet's AC1 computation.

### Accuracy

Confidence rates for the correct morphologic classification of lesions shown in the 70 video clips are listed in ► **Table 4**. Overall, the accuracy amounted to 79.3%. Only 26 lesions were correctly classified with a >90% value, and 12 of them with 100%. Lower accuracy values were those for sub-pedunculated lesions (lsp, 54–61%) and for some depressed lesions (Ilc, Is + Ilc, 31–46%). For a few sessile (0-Is), slightly elevated (0-Ila), mixed nodular (0-Ila + Is) and depressed/pseudodepressed (0-Ila + Ilc) lesions, the lowest values for accuracy were 54%, 46%, 38% and 54%, respectively. Mean operator accuracy for correct classification of lesions was also 79.3%, ranging from 64 to 91%. No single operator was 100% accurate, the best performer being correct 91% of the time. The lowest values (64%–66%) were registered for only two observers (a SM and a T), and all remaining colonoscopists had a score >74%. Correct identification of the lesion shape for LST was 77.6%, and that for the new classification system amounted to 91.6%.

### Discussion

In routine endoscopy reports, the descriptions of polyps vary widely between endoscopy units. Although a standardized form has been recommended [4, 5], some endoscopists detail the macroscopic shape of the lesion by using obsolete terminology,

► **Table 4** Diagnostic accuracy of the estimation for polyp morphology using both the Paris and Van Doorn [8] classification systems.

Evaluation	Accuracy	Range
Overall	79.3%	31–100
Operators	79.3%	64–91
Is	83.2%	54–100
Ip	90%	61–100
lsp	57.5%	54–61
Ila	80%	46–100
Ilb	73%	61–85
Ilc	31%	31–31
Ila + Is	78%	38–100
Ila + Ilc	74.6%	54–92
Is + Ilc	46%	46–46
LSTs	77.6%	38–100
New classification [8]	91.6%	54–100

LSTs, laterally spreading tumors.

while other professionals judiciously follow the Paris classification [10]. Knowledge of several morphologies is critical for endoscopists [22]. Over the years, Eastern and Western studies have been conducted to evaluate both the prevalence of the Paris classification subtypes and the risk of invasive cancer associated with the various lesions [1, 2, 23, 24]. A different distribution of non-polypoid lesions (NPLs) between East and West was found, although the variation may be more reflective of lower recognition ability by operators rather than a true difference in prevalence [2]. In regard to the risk of invasive cancer, worldwide data are superimposable, with higher rates for depressed lesions or for those with a depressed component (Ilc) [1, 5, 23].

There currently is debate between Western and Asian endoscopists about the general validity of the Paris classification of colonic lesions: the former operators claim a moderate interobserver agreement, as measured by  $\kappa$ -values of 0.42 and 0.48, and accuracy of 47.5% [8, 10], whereas South Korean

endoscopists report  $\kappa$ -values of 0.713 and accuracy of 0.797 [9]. Relying on previous figures, the value of the classification is considered questionable in clinical practice on one side of the world and far better on the other side. In this context, studies describing the prevalence and corresponding histology of polypoid and NPLs should be interpreted with caution [19], due to the lack of objective evaluation of the interobserver agreement [8, 24, 25]. To our knowledge, this approach to analysis of the prevalence of the several subtypes only is available in the Bianco et al [23], and the Kim et al [26] studies. Owing to the paucity of evidence on which to base a judgment, we carried out the present investigation, in which 13 Western operators working in the same endoscopy unit evaluated 70 video clips of superficial colonic lesions: after two-step, pre-study training, the evaluation produced an interobserver agreement value of 0.61 and an accuracy of 79.3%, values that indicate a substantial concordance among observers and approximate the Asian data [9]. Our study supports the merits of the morphologic classification of superficial colonic lesions and extends the generality of the Paris classification system, even in a Western context.

Several methodological differences may explain the divergent results between our investigation and the Van Doorn study [8]. First, the learning protocol differed. In the latter investigation, a training module was developed containing a classification overview, eight video clips and 32 still images. After evaluating them, the observers received a feedback form with correct answers. Moreover, not all lesion subtypes were presented. On the contrary, we provided face-to-face feedback to all 13 observers in two formal rounds: 1) to explain the classification system; and 2) when the 25 still images were reevaluated and discussed. We acknowledge that our results may reflect the experience of a single endoscopic center and not be indicative of a multicenter practice: all 13 observers in our study were SMs or Ts collaborating in the same unit, whereas the individual international experts involved in the other study were based in Europe or the United States [8]. However, with our approach, a substantial improvement in the rates of correct classification could be achieved after an appropriate training phase, a gain that was not detected in the Van Doorn study [8]. A future study should assess the multicenter agreement among observers working in different units to definitively confirm the accuracy of our rates. Second, the length of video clips and the time allowed for their evaluation were also different: short video clips of 10 to 25 seconds were developed in one investigation in which observers were allowed to watch a video up to three times [8]; in our study, we assembled videos of < 10 seconds to 4 minutes in duration, which could be reviewed separately but ad libitum by each rater.

Although the interobserver agreement in our study could be interpreted as substantial [18], we obtained a  $\kappa$ -value of only 0.61, which is higher than the one reported in the two previous Western studies [8, 10], but inferior to the one emerging from the Asian study [9]. To dig into our data, we ran several sensitivity subanalyses to explore how the variable experience among the observers (SMs vs Ts), polyp phenotypes (single vs mixed lesions), and the different gross morphologic features (Is, Ila,

Ila+Is, Ila+Ilc) might have impacted the results. Useful information was derived from these analyses. The most remarkable one pertains to the slight agreement for the individual lesion subtypes; in this analysis, the Cohen  $\kappa$  values were in the range of 0.08 to 0.12, which would signify a low reliability in describing individual lesions. However, when the same lesions were subjected to the Gwet statistics, most of the AC1 coefficients were indicative of substantial agreement. This problem, known as the “ $\kappa$  paradox,” reflects a situation in which the  $\kappa$ -value is low despite a high level of agreement. Mathematically, this effect is explained by the fact that  $\kappa$  is affected not only by the degree to which coders disagree but also by the skewed distribution of categories due to a prevalence deviating from 0.5 [16]. To fix these problems, Gwet [17] proposed using AC1 as a stable alternative to the unstable, misleading  $\kappa$  coefficient. As a matter of fact, we adopted both the Cohen  $\kappa$  value and the AC1 statistic for our analyses, and found a higher level of agreement with the latter statistic, which should give endoscopists confidence that the evaluations they are doing are reliable.

A further merit of the present investigation is the evaluation for the first time of the simplified polyp classification, as proposed by Van Doorn et al [8]. These authors, acknowledging the difficulties of polyp shape description according to the Paris classification, suggested a new classification system that distinguishes between only three broad categories: pedunculated, sessile/flat, and depressed lesions. In our investigation we have shown a high accuracy (91.6%) and an almost perfect agreement between the 13 coders, according to Landis and Koch [18]. As shown in ► **Table 3**, in the Gwet’s AC1 analysis, this simplified classification turned out to have the highest levels of agreement among the 13 coders (0.82; 95% CI: 0.77–0.88) [18]. However, the value was not perfect for the depressed subtype, with an AC1 score of 0.47 (95% CI: 0.21–0.72). For evaluation of single categories performed with the  $\kappa$  statistics, the paradoxical effect also was evident (0.01–0.10). Because lesions with depressed morphology are associated with risk of invasive cancer [27], more effort should be paid, in future studies, to identifying depressed lesions or depressed parts of a lesion.

As with any new classification system, there will be pros and cons. We think that through this simplified classification, something is gained: 1) greater interobserver agreement and accuracy; 2) an easy-to-use morphological system; 3) a single category including depressed lesion or demarcated depressed area in a lesion, the most relevant feature of the morphology characterization; and 4) the possibility of placing nonpolypoid and polypoid appearance in the same group of elevated lesions, being that their risk of advanced neoplasia is similar and essentially related to their size rather than to their macroscopic appearance [23]; therefore, we would group lesions with the same prognostic significance. However, we acknowledge a minor deficiency of this classification: the exclusion of reporting a nodule (demarcated area; 0-Is, >10 mm) in an elevated lesion (e.g. 0-Ila+Is), a feature that would change the therapeutic approach and the prognostic meaning of this subtle morphology [21, 28, 29]. A future study has to address this particular issue.

## Conclusion

In conclusion, we would stress the concept of continued training to improve communication and ameliorate the visual description of superficial colonic lesions. Endoscopists need to be confident that the classification they are using is valid and reliable. Furthermore, for the first time, we evaluated interobserver agreement, taking into account both simple and mixed lesions, the most important type in clinical practice. Agreement is often measured with Cohen's  $\kappa$ , but we proved a higher level of agreement when data were analyzed with the Gwet's AC1 statistic. In the latter evaluation, the Paris classification of superficial lesions was found to result in substantial agreement between the 13 Western coders; however, the simplified classification outperformed the Paris system by showing almost perfect agreement. Further research should be performed to consider improving the agreement for depressed neoplasms, which are more prone to be associated with invasive cancer.

## Competing interests

The authors declare that they have no conflict of interest.

## References

- [1] Kudo S, Lambert R, Allen JI et al. Nonpolypoid neoplastic lesions of the colorectal mucosa. *Gastrointest Endosc* 2008; 68: S3–S47
- [2] The Paris endoscopic classification of superficial neoplastic lesions. esophagus, stomach, and colon. November 30 to December 1, 2002. *Gastrointest Endosc* 2003; 58: S3–S43
- [3] Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy* 2005; 37: 570–578
- [4] Ferlitsch M, Moss A, Hassan C et al. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2017; 49: 270–297
- [5] Kaltenbach T, Anderson JC, Burke CA et al. Endoscopic removal of colorectal lesions – Recommendations by the US multi-society task force on colorectal cancer. *Gastrointest Endosc* 2020; 91: 486–519
- [6] Japanese Society for Cancer of the Colon and Rectum. Japanese classification of colorectal, appendiceal, and anal carcinoma: the 3rd English version. *J Anus Rectum Colon* 2019; 3: 175–195
- [7] Rao AK, Soetikno R, Raju GS et al. Large sessile serrated polyps can be safely and effectively removed by endoscopic mucosal resection. *Clin Gastroenterol Hepatol* 2016; 14: 568–574
- [8] Van Doorn SC, Hazewinkel Y, East EJ et al. Polyp Morphology: an interobserver evaluation for the Paris Classification among international experts. *Am J Gastroenterol* 2015; 110: 180–187
- [9] Kim JH, Nam KS, Kwon HJ et al. Assessment of colon polyp morphology: is education effective? *World J Gastroenterol* 2017; 23: 6281–6286
- [10] Aziz Adam A, Wani S, Kahi CH et al. Physician assessment and management of complex colon polyps: a multicenter video-based survey study. *Am J Gastroenterol* 2014; 109: 1312–1317
- [11] Lee YJ, Kim ES, Park KS et al. Inter-observer agreement in the endoscopic classification of colorectal laterally spreading tumors: a multicenter study between experts and trainees. *Dig Dis Sci* 2014; 59: 2550–2556
- [12] Bogie RMM, Veldman MHJ, Snijders LARS et al. Endoscopic subtypes of colorectal laterally spreading tumors (LSTs) and the risk of submucosal invasion: a meta-analysis. *Endoscopy* 2018; 50: 263–282
- [13] Kottner J, Audigé L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64: 96–106
- [14] Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543–549
- [15] Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551–558
- [16] Di Eugenio B, Glass M. The kappa statistic: a second look. *Comput Linguist* 2004; 30: 95–101
- [17] Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 2008; 61: 29–48
- [18] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174
- [19] Vleugels JLA, Hazewinkel Y, Dekker E. Morphological classifications of gastrointestinal lesions. *Best Pract Res Clin Gastroenterol* 2017; 31: 359–367
- [20] Matsuda T, Fujii T, Saito Y et al. Efficacy of the invasive/non-invasive pattern by magnifying chromoendoscopy to estimate the depth of invasion of early colorectal neoplasms. *Am J Gastroenterol* 2008; 103: 2700–2706
- [21] Iwatate M, Ikumoto T, Hattori S et al. NBI and NBI combined with magnifying colonoscopy. *Diagn Ther Endosc* 2012; 2012: 173269
- [22] Rex DK, Hassan C, Bourke MJ. The colonoscopist's guide to the vocabulary of colorectal neoplasia: histology, morphology, and management. *Gastrointest Endosc* 2017; 86: 253–263
- [23] Bianco MA, Cipolletta L, Rotondano G et al. Prevalence of nonpolypoid colorectal neoplasia: an Italian multicenter observational study [published correction appears in *Endoscopy*. 2010 Jul; 42(7): 563]. *Endoscopy* 2010; 42: 279–285
- [24] Soetikno RM, Kaltenbach T, Rouse RV et al. Prevalence of nonpolypoid (flat and depressed) colorectal neoplasms in asymptomatic and symptomatic adults. *JAMA* 2008; 299: 1027–1035
- [25] Sanduleanu S, Rondagh EJ, Masclee AA. Development of expertise in the detection and classification of non-polypoid colorectal neoplasia: Experience-based data at an academic GI unit. *Gastrointest Endosc Clin N Am* 2010; 20: 449–460
- [26] Kim BC, Chang HJ, Han KS et al. Clinicopathological differences of laterally spreading tumors of the colorectum according to gross appearance. *Endoscopy* 2011; 43: 100–107
- [27] Rembacken BJ, Fujii T, Cairns A et al. Flat and depressed colonic neoplasms: a prospective study of 1000 colonoscopies in the UK. *Lancet* 2000; 355: 1211–1214
- [28] Puig I, Mårholm C, Bustamante M. Endoscopic imaging techniques for detecting early colorectal cancer. *Curr Opin Gastroenterol* 2019; 35: 432–439
- [29] Bisschops R, East JE, Hassan C et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy* 2019; 51: 1155–1179