ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL ARTICLE

Critical insights on real-life PD-L1 histopathological workflow and assessment in esophageal, esophagogastric junction, and gastric carcinoma in Italy

A. Gambella[1†], F. Grillo[1,2*†], P. Parente[3], A. Vanoli[4,5], A. Caputo[6,7], M. Paudice[1,2], V. Angerilli[8], F. Castri[9], A. Pastorino[2,10], M. Fassan[11,12‡] & L. Mastracci[1,2‡], on behalf of the RELIABL Study Group

[1]Pathology Unit, Department of Surgical Sciences and Integrated Diagnostics (DISC), University of Genoa, Genoa; [2]IRCCS Ospedale Policlinico San Martino, Genoa; [3]Pathology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo; [4]Department of Molecular Medicine, University of Pavia, Pavia; [5]Unit of Anatomic Pathology, IRCCS San Matteo Hospital Foundation, Pavia; [6]Department of Pathology, University Hospital of Salerno, Salerno; [7]Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", University of Salerno, Baronissi; [8]Surgical Pathology Unit, ULSS2 Marca Trevigiana, Treviso; [9]Policlinico Universitario A. Gemelli, Roma; [10]Medical Oncology Unit 1, IRCCS Ospedale Policlinico San Martino, Genova; [11]Veneto Institute of Oncology IOV-IRCCS, Padua; [12]Department of Medicine (DIMED), University of Padua, Padua, Italy

Available online xxx

Background: Immune checkpoint inhibitors (ICIs) targeting programmed cell death protein 1/programmed death-ligand 1 (PD-1/PD-L1) have improved survival in locally advanced and metastatic esophageal, esophagogastric junction, and gastric carcinoma (GEC). Patient selection for ICI-treatment relies on PD-L1 protein expression assessment via immunohistochemistry (IHC). This study aimed to evaluate the real-world assessment of PD-L1 IHC results compared with clinical trial data.

Patients and methods: This multicentric, real-world retrospective study analyzed PD-L1 IHC data from 28 Italian pathology centers of GEC cases diagnosed between October 2023 and September 2024. The study documented PD-L1 expression distribution via combined positive (CPS), tumor proportion (TPS), and tumor area positivity (TAP) scores, and investigated the impact of several factors on IHC results.

Results: We collected 1936 cases: 1802 adenocarcinomas (ADCA), 131 squamous carcinomas (SCC), and 3 carcinomas of non-specific histotype. Most institutions reported CPS and TPS data, whereas a minority used TAP. Overall, CPS, TPS, and TAP scores were in line with the data in literature and clinical trials for both ADCA and SCC, but inter-institutional heterogeneity was observed as represented by CPS $\geq$1 ADCA cases (range among institutions: 43.6%-100%). Inter-institutional heterogeneity was significantly associated with several variables, including (i) PD-L1 IHC case workload, with lower workload centers reporting more CPS $\geq$1 cases on average, and (ii) PD-L1 clone, with the 22C3 clone showing higher CPS scores than the SP263 clone. Tissue block aging was also significantly associated with a lower PD-L1 score, with a critical time window at 24-60 months.

Conclusions: This study confirms the alignment of GEC PD-L1 expression in Italian real-world practice with clinical trials. Inter-institutional variability and the significant influence of preanalytical factors, particularly tissue aging and PD-L1 clone, highlight important challenges in routine PD-L1 testing. Addressing these issues is crucial to enhance the reliability of PD-L1 IHC assessment and ensure optimal patient selection for ICIs in GEC.

Key words: PD-L1 immunohistochemistry, predictive biomarker, national survey, gastroesophageal cancer, interobserver variability, immune checkpoint inhibitors

INTRODUCTION

Cancers of the esophagus, gastroesophageal junction, and stomach (GEC) account for >1.1 million new oncology cases yearly worldwide,[1] and present a dismal prognosis, especially if they are locally advanced or metastatic.[2-4] In metastatic cancers, clinical trials testing immune checkpoint inhibitors (ICIs) have shown improved survival: programmed cell death protein 1 (PD-1) ICIs have improved patient overall survival significantly, both as single agents[5,6]

*Correspondence to: Dr Prof. Federica Grillo, Pathology Unit, Department of Surgical Sciences and Integrated Diagnostics (DISC), University of Genoa, IRCCS Ospedale Policlinico San Martino, Viale Benedetto XV, 6, Genoa 16132, Italy. Tel: +39-010-555-5957

E-mail: federica.grillo@unige.it (F. Grillo).

†These authors equally contributed to this study and shared first authorship.

‡These authors equally contributed to this study and shared last authorship.

2059-7029/© 2025 The Author(s). Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

and combined with standard chemotherapy.[7-9] Based on these data, ICIs are now a cornerstone of both squamous-cell carcinoma (SCC) and adenocarcinoma (ADCA) (both HER2-positive and HER2-negative) in the locally advanced unresectable or metastatic setting.[10] More recently, ICIs have shown relevant efficacy when combined with chemotherapy in the perioperative setting of locally advanced resectable ADCA of the gastroesophageal junction and stomach.[11]

The tissue-tethered assessment of predictive biomarkers[12] is crucial for patient enrollment in clinical trials and therapeutic access in real-world practice using protein expression via immunohistochemical staining (IHC) of CD274 [i.e. programmed death ligand 1 (PD-L1)].[13-16] Biologic rationale, predictive potential, and available scoring systems for PD-L1 IHC assessment have been thoroughly discussed in the literature,[17-20] including the implementation of artificial intelligence-driven approaches.[21-25]

Use of different PD-L1 IHC scoring systems is mostly related to the approved ICI drug and the available antibody clone. Overall, combined positive score (CPS) evaluation is required for patients with either ADCA or SCC for treatment with either pembrolizumab or nivolumab. Evaluation of tumor proportion score (TPS), with its focus on tumor cells, is required for the use of nivolumab in locally advanced unresectable/metastatic SCCs (which show more prominent tumor cell PD-L1 expression than ADCA).[14] More recently, the tumor area positivity (TAP) score has been introduced for patient selection for treatment with tislelizumab in both ADCA and SCC. The TAP score aims to facilitate PD-L1 assessment by prioritizing the area of PD-L1 positivity rather than the specific cell count,[26,27] and the TAP score may become the leading selection system for other ICIs in the near future. Data from clinical trials[28] reported a PD-L1 IHC-positive expression for ADCA that ranged from 78%-85% of analyzed cases via CPS (considered positive if the score is $\geq$1),[7,9,29,30] 55% of analyzed cases via TAP (considered positive if the score is $\geq$5%),[31] and 16% of cases via TPS (considered positive if the score is $>$1%).[7,32] Conversely, SCC presented a CPS of $\geq$10 in 43%-53.6% of cases,[8,33] TPS of $\geq$1% in 51% of cases,[34] and a TAP of $\geq$10% in 30.7%-34.4% of cases.[35,36] Based on these data, different clinically relevant cut-offs are considered to date (e.g. CPS $\geq$1 for HER2-positive ADCA), although there are still significant differences depending on specific ICI labels and regulatory authorities' authorization.[37-39] In addition, PD-L1 IHC workflow in clinical trials is quite different from the clinical setting, as samples are centralized and readily managed with a dedicated workflow for IHC staining and related assessment by a pathologist with specific training. In real-world practice, this is not always the case. PD-L1 IHC is carried out on (i) tissue samples that have often undergone non-standardized preanalytical phases (e.g. risk of under-fixation or over-fixation, variable cold-ischemia time), (ii) potentially long-stored archived tissue blocks, and (iii) possible pre-cut blank slides prepared at the time of diagnosis and then indefinitely archived for subsequent use. Unfortunately, specific data about the preanalytical phases and diagnostic daily routine from the pathology units is mostly lacking, especially regarding their impact on PD-L1 scoring systems and related results.[40]

Considering (i) the clinical relevance of PD-L1 testing,[2,14,16,18] (ii) the impact of preanalytical variables on IHC protein expression,[41-46] and (iii) the workflow variability among pathology units, this study aimed to document the Italian real-world diagnostic practice of PD-L1 IHC assessment of patients affected with GEC. We developed and distributed a survey to multiple Italian institutions to specifically harvest data regarding the overall PD-L1 diagnostic workup carried out in a specific 1-year time window, including the type of samples analyzed, and the types of PD-L1 clones, immunostainers, and scoring systems used. Moreover, we further analyzed data to identify putative causes of inter-institution score variability, such as slide and tissue block aging and volume of PD-L1 IHC assessment per institution.

METHODS

This document originated from the 'The REal-LIfe Assessment of Biomarker in pathology Labs (RELIABL) Study Group', an initiative coordinated by the Italian Group of Gastrointestinal Pathologists.

A survey was circulated among the Italian Pathology Units that routinely assess PD-L1 on GEC; 100 institutions were invited, and 28 provided data. The survey consisted of five introductory questions concerning (i) the overall workload of the institution (overall cases per year), (ii) staff numbers involved in performing the IHC staining procedure, (iii) the number of pathologists evaluating PD-L1 for gastrointestinal (GI) malignances, (iv) whether the pathologists evaluating PD-L1 for GI malignances have or do not have a dedicated training in GI pathology, and (v) how many pathologists sign out the PD-L1 IHC assessment report. The participating institutions were then asked to fill an Excel (Microsoft Corporation, 2025; Microsoft Excel) template with information regarding their institution's PD-L1 assessment practice, including the specimen type, primary tumor site, histotype, sample adequacy, IHC stainer, PD-L1 clone, slide and tissue block aging, and the CPS, TPS, and TAP scores for each case according to published definition/criteria.[13,14,27] In particular, CPS is calculated by dividing the number of PD-L1 positive cells (tumor, lymphocytes, and macrophages) by the total number of viable tumor cells, then multiplying by 100[13,14]; TPS is determined by dividing the number of PD-L1 positive tumor cells by the total number of viable tumor cells, then multiplying by 100 and reporting the value as a percentage[13,14]; TAP is obtained by dividing the area covered by PD-L1 positive tumor and immune cells by the total tumor area and reporting the value as a percentage.[27] Institutions were asked to report CPS, TPS, and TAP in several ways. First, using a semi-quantitative approach and differentiating cases accordingly. Based on this approach, CPS was reported as $<$1 or $\geq$1, and cases with a CPS $\geq$1 were further stratified in $\geq$5, and then $\geq$10. TPS was reported as $<$1% or TPS $\geq$1%. TAP score

was reported as <1%, ≥1%, and ≥5%, the latter further stratified in ≥10%. In addition, institutions were also asked to provide the exact value for each score (i.e. the specific value).

Details of the survey and related variables are reported in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105846. We requested institutions to provide data on cases diagnosed from 1 October 2023 to 30 September 2024.

Statistical analyses were carried out with R and RStudio (version 4.4.3 and 2024.12.1+563, respectively). Descriptive data were provided as mean or median with interquartile range (IQR). Parametric and non-parametric tests were used as appropriate and are specified throughout the document. We used two-tailed tests, and a P value < 0.05 was considered statistically significant for the purpose of results interpretation.

The study adhered to the Italian Medicines Agency determination (20 March 2008) on retrospective observational studies using anonymized data, which does not mandate ethics committee approval or informed consent. The research was conducted in accordance with the Declaration of Helsinki (1964) and its subsequent amendments.

RESULTS

Introductory indices, demographic data, and preanalytical variables

We started our study by performing several introductory and descriptive analyses, which were pivotal for multiple reasons. They (i) provide an updated representation of the GEC demography in Italy over a 1-year period, (II) validate our series by direct comparison with data from the literature, and (iii) prepare data to identify potential causes of PD-L1 scoring variability among institutions.

Twenty-eight institutions participated in the RELIABL initiative, fairly distributed throughout the whole Italian territory (Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2025.105846).

The median workload of the participating centers was 24 300 cases per year (IQR 16 875-37 000 cases), with minimum and maximum values of 11 200 and 71 000, respectively. Most institutions (26/27; 96.3%) had biomedical scientists/technologists performing the IHC stains, and only one institution had a dedicated biologist instead. The median number of pathologists per institution evaluating PD-L1 was 3 (IQR 2-5 pathologists), 63.0% of the institutions (17/27) had pathologists with dedicated GI training evaluating the PD-L1 stains, but 67.9% (19/28) of these had only one pathologist signing out the final report. Among these laboratories, seven specified that an additional pathologist signed out the report if the evaluation was challenging.

Data were collected from 1936 cases evaluated for PD-L1 IHC expression. Of these, 390 (20%) were external pathology consultations, accounting for up to 60.5% of PD-L1 workload in individual centers. To evaluate whether the PD-L1 workload had an impact on subsequent analyses—in particular, PD-L1 score results—we evaluated the number of PD-L1 cases per institution and decided to stratify following a three-tier system: <50 cases, 50-100 cases, and >100 cases. Accordingly, 12 (42.8%) institutions evaluated <50 cases, 10 (35.7%) institutions evaluated 50-100 cases, and 6 (21.5%) institutions evaluated >100 cases. Furthermore, as PD-L1 IHC-expression assessment and scoring systems vary depending on the histological subtype, we evaluated data by providing descriptive statistics of the whole series and considering ADCA and SCC separately. In the entire cohort, 65.5% of cases (1268/1936) were male, with a median age of 70 years (IQR 61.5-78 years). Most specimens (56.3%) were obtained via an endoscopic biopsy, predominantly from the stomach (1245 cases, 64.3%). A smaller subset (201 cases, 10.4%) was assessed on metastatic tissue, including only three nodal samples. Data stratified according to histologic subtype (i.e. ADCA versus SCC) provided similar results, as detailed in Table 1. Most cases per institution (mean percentage 91.2%, median percentage 93.3%) were ADCA, as expected in a Western cohort. Regarding ADCA, all institutions reported using CPS for PD-L1 assessment; TPS and TAP were reported by 25/28 and 8/28 institutions. Similarly, 25/28 centers (89.3%) reported information on at least one case of SCC using CPS, 24/28 (85.7%) using TPS, and 5/28 (17.9%) using TAP.

Subsequently, preanalytical steps were analyzed, such as sample adequacy, presence of on-slide IHC-positive controls, type of immunostainer and PD-L1 clone used, and the aging of both the tissue slides and tissue blocks. From this data collection, detailed in Table 2, the PD-L1 clones and related immunostainers used, and the aging of tissue slides and blocks were of particular interest. Most institutions (23/28; 82%) used the Ventana immunostainer (Ventana, Roche, Indianapolis, IN); one institution used the Dako immunostainer (Dako, Carpinteria, CA), one institution used the Leica immunostainer (Leica Biosystems, Nussloch, Germany), and three institutions had access to multiple platforms. Staining was carried out using either the SP263 clone (Ventana, Roche; 1087/1936; 56.1%) or the 22C3 clone (Dako; 849/1936, 43.9%), with four centers applying the 22C3 clone on the Ventana platform. Slide and block aging were calculated in days and months, respectively. The mean storage times were 11.2 days for slides and 2.6 months for blocks, with extreme cases of slides stored up to 359 days and blocks up to 84 months before staining. To assess the effect of tissue aging on PD-L1 IHC, slides were grouped into: <1 day, 1-3 days, 3-5 days, 5-10 days, 10-15 days, 15-30 days, 30-60 days, and >60 days. Blocks were categorized as: <4 months, 4-8 months, 8-12 months, 12-24 months, 24-60 months, >60 months. Differences between ADCA and SCC are reported in Table 2. Finally, we evaluated the distribution of PD-L1 score use and related results across institutions (Figure 1).

PD-L1 assessment—adenocarcinoma cases

All institutions reported using CPS for PD-L1 assessment; TPS and TAP were reported by 25/28 and 8/28 institutions,

Table 1. Demographic and sampling details about the overall series, further stratified in adenocarcinoma (ADCA) and squamous-cell carcinoma (SCC) subgroups

Characteristic	Overall series (n = 1936)	ADCA (n = 1802)	SCC (n = 131)
Sex, male (%)	1268 (65.5)	1175 (65.2)	90 (68.7)
Age, median years (IQR)	70 (61.5-78)	71 (61.5-78)	69 (61.5-78)
Type of cases, external consults (%)	393 (20.3)	355 (19.7)	35 (26.7)
Specimen type (%)			
Endoscopic biopsy	1090 (56.3)	981 (54.4)	108 (82.4)
EUS-FNB	9 (0.5)	9 (0.5)	0 (0.0)
Lymph node	3 (0.2)	3 (0.2)	0 (0.0)
Needle biopsy	33 (1.7)	30 (1.7)	3 (2.3)
Surgical biopsy	67 (3.5)	67 (3.7)	0 (0.0)
Surgical specimen	730 (37.7)	709 (39.3)	19 (14.5)
Other	3 (0.2)	3 (0.2)	0 (0.0)
Not available	1 (0.1)	0 (0.0)	1 (0.8)
Site of primary disease (%)			
Esophagus, proximal	29 (1.5)	9 (0.5)	20 (15.3)
Esophagus, medial	39 (2.0)	8 (0.4)	31 (23.7)
Esophagus, distal	91 (4.7)	67 (3.7)	24 (18.3)
Esophagus, NOS	118 (6.1)	80 (4.4)	37 (28.2)
Junctional, Siewert 1	79 (4.1)	76 (4.2)	3 (2.3)
Junctional, Siewert 2	196 (10.1)	193 (10.7)	2 (1.5)
Junctional, Siewert 3	113 (5.8)	111 (6.2)	2 (1.5)
Junctional, NOS	17 (0.9)	15 (0.8)	2 (1.5)
Stomach, fundus	36 (1.9)	36 (2.0)	0 (0.0)
Stomach, body	263 (13.6)	260 (14.4)	3 (2.3)
Stomach, antrum	378 (19.5)	373 (20.7)	4 (3.1)
Stomach, pylorus	26 (1.3)	26 (1.4)	0 (0.0)
Stomach, NOS	542 (28.0)	539 (29.2)	3 (2.3)
Not available	9 (0.5)	9 (0.5)	0 (0.0)
Sample of a metastatic lesion	201 (10.4)	189 (10.5)	10 (7.6)
If metastatic, site			
Lymph node	73 (36.3)	69 (36.5)	3 (30.0)
Peritoneum	62 (30.8)	62 (32.8)	0 (0.0)
Liver	27 (13.4)	23 (12.2)	3 (30.0)
Large bowel	12 (6.0)	12 (6.3)	0 (0.0)
Lung	3 (1.5)	2 (1.1)	1 (10.0)
Pleura	1 (0.5)	1 (0.5)	0 (0.0)
Other	23 (11.4)	20 (10.6)	3 (30.0)

EUS-FNB, endoscopic ultrasound-guided fine needle biopsy; IQR, interquartile range; NOS, not otherwise specified.

Table 2. Distribution of preanalytical variables in the overall series and stratified according to histotype

Characteristic	Overall series (n = 1936)	ADCA (n = 1802)	SCC (n = 131)
Adequacy, tissue quantity			
<100 cells	53 (2.7)	48 (2.7)	5 (3.8)
100-500 cells	598 (30.9)	550 (30.5)	48 (36.6)
>500 cells	1018 (52.6)	958 (53.2)	57 (43.5)
Not available	267 (13.8)	246 (13.6)	21 (16.1)
PD-L1 control			
Positive tonsil control	1228 (63.4)	1140 (63.3)	87 (66.4)
Other positive control	493 (25.5)	464 (25.7)	28 (21.4)
No positive control	70 (3.6)	69 (3.8)	1 (0.8)
Not available	145 (7.5)	129 (7.2)	15 (11.5)
IHC stainer			
Ventana	1307 (67.5)	1206 (66.9)	98 (74.8)
Dako	624 (32.2)	592 (32.9)	32 (24.4)
Leica	5 (0.3)	4 (0.2)	1 (0.8)
PD-L1 clone			
22C3	849 (43.9)	799 (44.3)	49 (37.4)
SP263	1087 (56.1)	1003 (55.7)	82 (62.6)
Slide aging, days			
Median (IQR)	1 (1-2)	1 (1-2)	1 (1-4.25)
Mean	11.2	10.6	19.0
Minimum	0	0	0
Maximum	359	359	325
Slide aging, cases			
<1 day	1283 (66.3)	1205 (66.9)	77 (58.8)
1-3 days	261 (13.5)	244 (13.5)	15 (11.5)
3-5 days	29 (1.5)	24 (1.3)	5 (3.8)
5-10 days	83 (4.3)	73 (4.1)	10 (7.6)
10-15 days	40 (2.1)	36 (2.0)	4 (3.1)
15-30 days	53 (2.7)	48 (2.7)	5 (3.8)
30-60 days	22 (1.1)	19 (1.1)	3 (2.3)
>60 days	77 (4.0)	68 (3.8)	9 (6.9)
Not available	88 (4.5)	85 (4.7)	3 (2.3)
Tissue block aging, months			
Median (IQR)	0 (0-1)	0 (0-1)	0.5 (0-3)
Mean	2.6	2.5	3.4
Minimum	0	0	0
Maximum	84	71	84
Tissue block aging, cases			
<4 months	1527 (78.9)	1433 (79.5)	91 (69.5)
4-8 months	121 (6.3)	105 (5.8)	16 (12.2)
8-12 months	40 (2.1)	36 (2.0)	4 (3.1)
12-24 months	54 (2.8)	50 (2.8)	4 (3.1)
24-60 months	32 (1.7)	31 (1.7)	1 (0.8)
>60 months	6 (0.3)	5 (0.3)	1 (0.8)
Not available	156 (8.1)	142 (7.9)	14 (10.7)

ADCA, adenocarcinoma; IHC, immunohistochemical; IQR, interquartile range; PD-L1, programmed death-ligand 1; SCC, squamous-cell carcinoma.

respectively (Table 3). Focusing on CPS, it was available for 1784 of the 1802 (99.0%) ADCA cases collected with the survey, as 11 were not evaluated (complete absence of neoplastic cells on immunostained cells), and in 7, only a TPS score was provided. Of those with tumor cells, 36 (2.1%) presented <100 tumor cells, 545 (30.5%) presented 100-500 tumor cells, and 958 (53.7%) had >500 tumor cells. The remaining 245 cases (13.7%) were evaluated with different metrics for adequacy (e.g. biopsy dimensions, number of fragments with ADCA). We were also interested in evaluating the use of PD-L1 positive controls for the IHC stain. Among CPS-scored cases (n = 1784), 63.4% used on-slide tonsil tissue as the positive control, 25.7% used other control tissue, and in 10.9% (194 cases), positive control data were unavailable.

CPS positivity was defined as ≥1; higher thresholds (≥5, ≥10) correlate with increased likelihood of benefit from ICIs.[7,9,14,29,47] ADCA cases showed 76.1% (1358) with CPS ≥1, 55.5% (991) with CPS ≥5, and 35.8% (639) with CPS ≥10 (see Table 3 and Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2025.105846). Inter-institutional analysis revealed substantial variability, with CPS ≥1 rates in ADCA ranging from 43.6%-100% across centers. Similar data were observed for CPS ≥5 (range 18.0%-95.0%) and CPS ≥10 (range 0%-85.0%; Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105846).

To identify potential causes of this inter-institution variability, we explored the impact of several variables, including the number of PD-L1 IHC cases per institution,

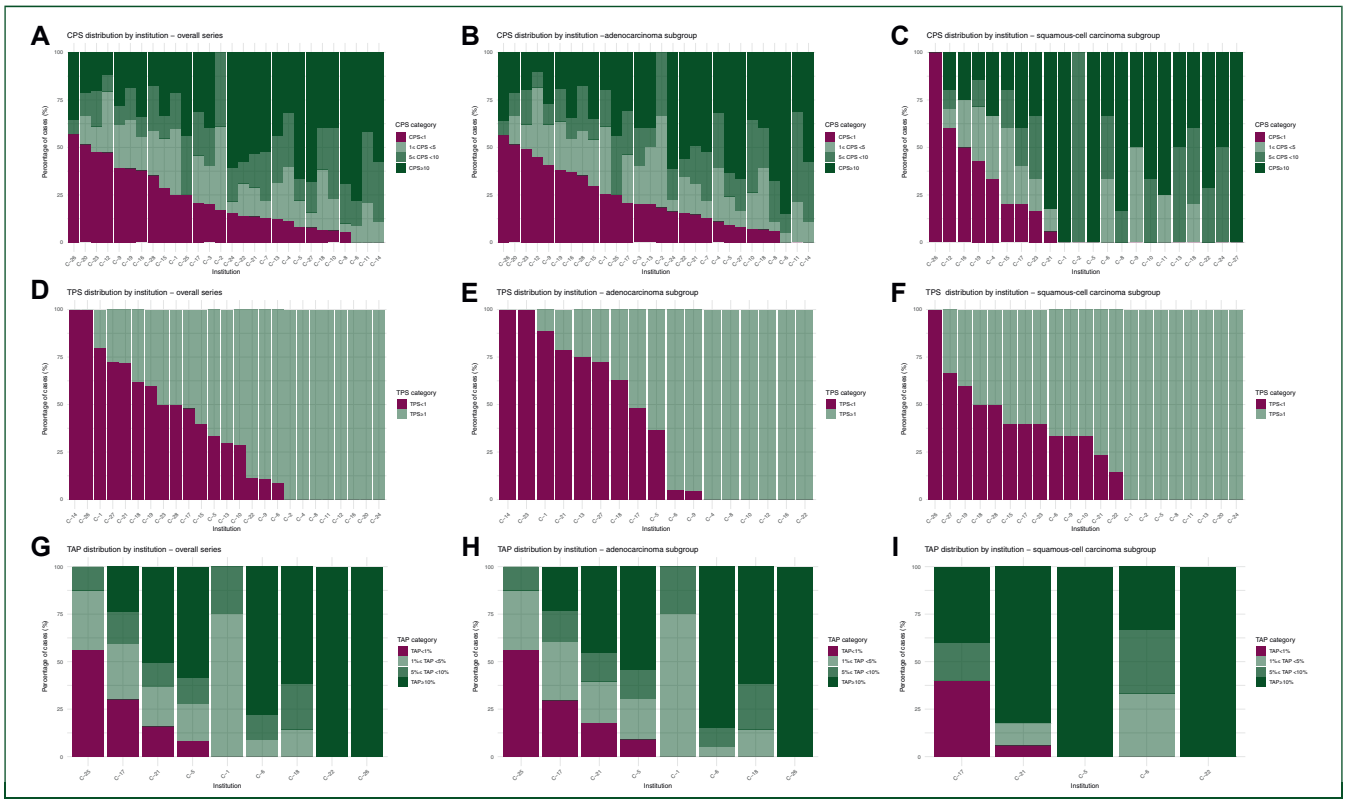

Figure 1. Distribution of PD-L1 scoring systems across the overall series and histological subgroups adenocarcinoma and squamous-cell carcinoma. Stacked bar charts show the distribution of combined positive score (CPS) (A, B, C), tumor proportion score (TPS) (D, E, F), and tumor area positivity score (TAP) (G, H, I) in the overall series (A, D, G), the adenocarcinoma subgroup (B, E, H), and the squamous-cell carcinoma subgroup (C, F, I). Each bar represents the relative proportion of samples falling within predefined PD-L1 expression categories for each scoring system, as represented in the related legend

PD-L1 clone used (SP263 versus 22C3), specimen origin (external consults versus internal cases), and aging of tissue slides and tissue blocks. Interestingly, we observed a statistically significant association ($P = 0.019$) between PD-L1 IHC cases workload and CPS results; institutions with a lower workload (<50 cases) reported more cases with CPS ≥1 (81.8%) than those with a higher workload (50-100 cases 74.8%; >100 cases 75.3%). Focusing on PD-L1 clones, CPS evaluated with the 22C3 clone was higher compared with CPS evaluated with SP263. This association was confirmed using continuous values ($P < 0.001$) and semiquantitative stratification ($P < 0.001$). Of note, both clones were almost equally represented in the ADCA subgroup (ADCA cases stained with 22C3 and SP263 clones and evaluated with CPS were 797 and 987, respectively). Tissue slide aging was associated with CPS ($P < 0.001$), and lower scores were observed as the age of tissue slide increased, but the correlation was not significant when considering continuous data ($r = 0.023$; $P = 0.38$). Similarly, cases with tissue slide aging >15 days did not show significantly different CPS values (Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105846). On the other hand, tissue block aging was significantly associated with CPS score both as a semiquantitative ($P < 0.001$) and continuous ($r = -0.07$, $P = 0.014$; Figure 2A) variable. In particular, the older the tissue block, the lower the CPS values, and we identified 24-60 months as the most critical time window (Figure 2B). The specimen origin (external consults/internal cases) was associated with higher values of CPS results as a continuous variable ($P = 0.001$) but not as semiquantitative subgroups ($P = 0.07$). Similar data were observed when considering CPS ≥5 and CPS ≥10 as cut-offs discriminating positive and negative cases (Supplementary Figure S4, available at https://doi.org/10.1016/j.esmoop.2025.105846). Due to the relevance of tissue aging and workload that emerged from these analyses, we were interested in testing whether any correlations existed. Indeed, we identified a significant inverse correlation between institutions' workload and both tissue block ($r = -0.09$; $P < 0.001$) and tissue slide ($r = -0.14$; $P < 0.001$) aging. In particular, the lower the workload, the higher the aging.

Considering the unidirectional nature of these correlations, we wanted to test which variables were driving the differences in CPS scores. To this end, we developed a multivariate model that identified the PD-L1 clone [odds ratio (OR) 0.34, 95% confidence interval (CI) 0.26-0.44, $P < 0.001$] and tissue block aging (OR 0.98, 95% CI 0.96-0.98, $P < 0.001$) as independent variables affecting CPS results (Supplementary Figure S5, available at https://doi.org/10.1016/j.esmoop.2025.105846).

TAP scoring was reported by 8 of 28 institutions (28.6%). Mean and median scores were 13.0 and 5 (IQR 2-15), respectively. TAP <1% was recorded in 19.7% (59 cases) and TAP ≥1% in 80.3% (241 cases), with 169 and 122 cases exceeding ≥5% and ≥10%, respectively. Like CPS, inter-

Table 3. Details of all scoring systems in the overall series and stratified according to histotype

PD-L1 IHC protein expression scoring systems	Overall	ADCA subgroup	ADCA	SCC subgroup	SCC
			Clinical trial range of percentage of cases [reference]		Clinical trial range of percentage of cases [reference]
CPS	$n = 1897$	$n = 1784$	—	$n = 113$	
<1, % (*n*)	23.3 (443)	23.9 (426)	—	15.0 (17)	—
≥1, % (*n*)	76.7 (1454)	76.1 (1358)	78-85[7,9,29,30]	85.0 (96)	—
≥5, % (*n*)	73.7 (1071)	72.9 (991)	60-61[7,47]	83.3 (80)	—
≥10, % (*n*)	47.9 (696)	47.0 (639)	35[9,29]	59.4 (57)	43-53.6[8,33]
Mean value	14.6	14.3	—	20.8	—
Median value (IQR)	7 (2.6-18)	7 (2-17)	—	12 (4-25)	—
Minimum value	0	0	—	0	—
Maximum value	100	100	—	100	—
Number of institutions reporting at least one case (%)	28/28 (100)	28/28 (100)	—	25/28 (89.3)	—
TPS	$n = 524$	$n = 427$	—	$n = 97$	—
<1%, % (*n*)	51.5 (270)	56.9 (243)	—	27.8 (27)	—
≥1%, % (*n*)	48.5 (254)	43.1 (184)	16[7,32]	72.2 (70)	51[34]
Mean value	9.6	8.0	—	17.0	—
Median value (IQR)	1 (0-8.2)	0.5 (0-5)	—	5 (1-20)	—
Minimum value	0	0	—	0	—
Maximum value	100	100	—	100	—
Number of institutions reporting at least one case (%)	25/28 (89.3)	17/28 (60.7)	—	24/28 (85.7)	—
TAP	$n = 329$	$n = 300$	—	$n = 29$	—
<1%, % (*n*)	18.8 (62)	19.7 (59)	—	10.3 (3)	—
≥1%, % (*n*)	81.1 (267)	80.3 (241)	—	89.6 (26)	—
≥5%, % (*n*)	71.9 (192)	70.1 (169)	55[31]	88.5 (23)	—
≥10%, % (*n*)	53.5 (143)	50.6 (122)	—	80.8 (21)	30.7-34.4[35,36]
Mean value	14.12	13.0	—	25.3	—
Median value (IQR)	6.0 (2-15)	5 (2-15)	—	15 (6-30)	—
Minimum value	0	0	—	0	—
Maximum value	100	100	—	100	—
Number of institutions reporting at least one case (%)	9/28 (32.1)	8/28 (28.6)	—	5/28 (17.9)	—

ADCA, adenocarcinoma; CPS, combined positive score; IHC, immunohistochemical; IQR, interquartile range; PD-L1, programmed death-ligand 1; SCC, squamous-cell carcinoma; TAP, tumor area positivity score; TPS, tumor proportion score.

institution variability was observed for the TAP score (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105846). We then analyzed putative factors that may influence TAP scoring. Again, the PD-L1 clone was related to TAP yield, but differently from the CPS; higher TAP values were associated with the SP263 clone rather than the 22C3 ($P < 0.001$). Slide aging was also associated with TAP PD-L1 yield, and, specifically, the expression dropped in the 1- to 3-day window ($P = 0.04$). None of the other variables were significantly related to TAP values (Supplementary Figure S6, available at https://doi.org/10.1016/j.esmoop.2025.105846). A multivariate model identified the PD-L1 clone (OR 2.19, 95% CI 1.04-4.90, $P < 0.05$) as an independent variable affecting the TAP score (Supplementary Figure S7, available at https://doi.org/10.1016/j.esmoop.2025.105846).

PD-L1 assessment—SCC cases

Overall, we collected data on 131 SCC cases, and 25/28 centers (89.3%) reported information on at least one case of SCC using CPS, 24/28 (85.7%) using TPS, and 5/28 (17.9%) using TAP. ADCA and SCC presented similar data distribution according to the PD-L1 scoring systems despite the evident sample size disparity (1802 ADCA cases versus 131 SCC cases, respectively), as detailed in Table 3. Most cases were CPS ≥1 (85.0% of cases, 59.4% of which were CPS ≥10), TPS ≥1% (72.2% of cases), and TAP ≥1% (89.6% of cases, 88.5% of which were TAP ≥5%). The most striking but expected variation compared with ADCA was related to the TPS, as most SCC cases (70/97, 72.2%) presented TPS ≥1% as opposed to ADCA (184/427, 43.1%).

Differences between institutions were then analyzed. We observed a relevant inter-institutional heterogeneity, especially for TPS (TPS ≥1%, range 33.3%-100%) and CPS (CPS ≥10, range 14.3%-100%) systems. Like ADCA, we explored the impact of several variables while contemporarily acknowledging the limitation of this subgroup analysis (i.e. the small sample size of SCC). Starting with the CPS, PD-L1 clones and specimen origin were associated with CPS results, and SCC cases that were stained with SP263 clones presented lower CPS values than those stained with 22C3, both as continuous ($P < 0.001$) and semiquantitative ($P < 0.001$) variables. Similarly, internal cases presented lower CPS values than external cases but only when considering CPS as a continuous variable ($P =$

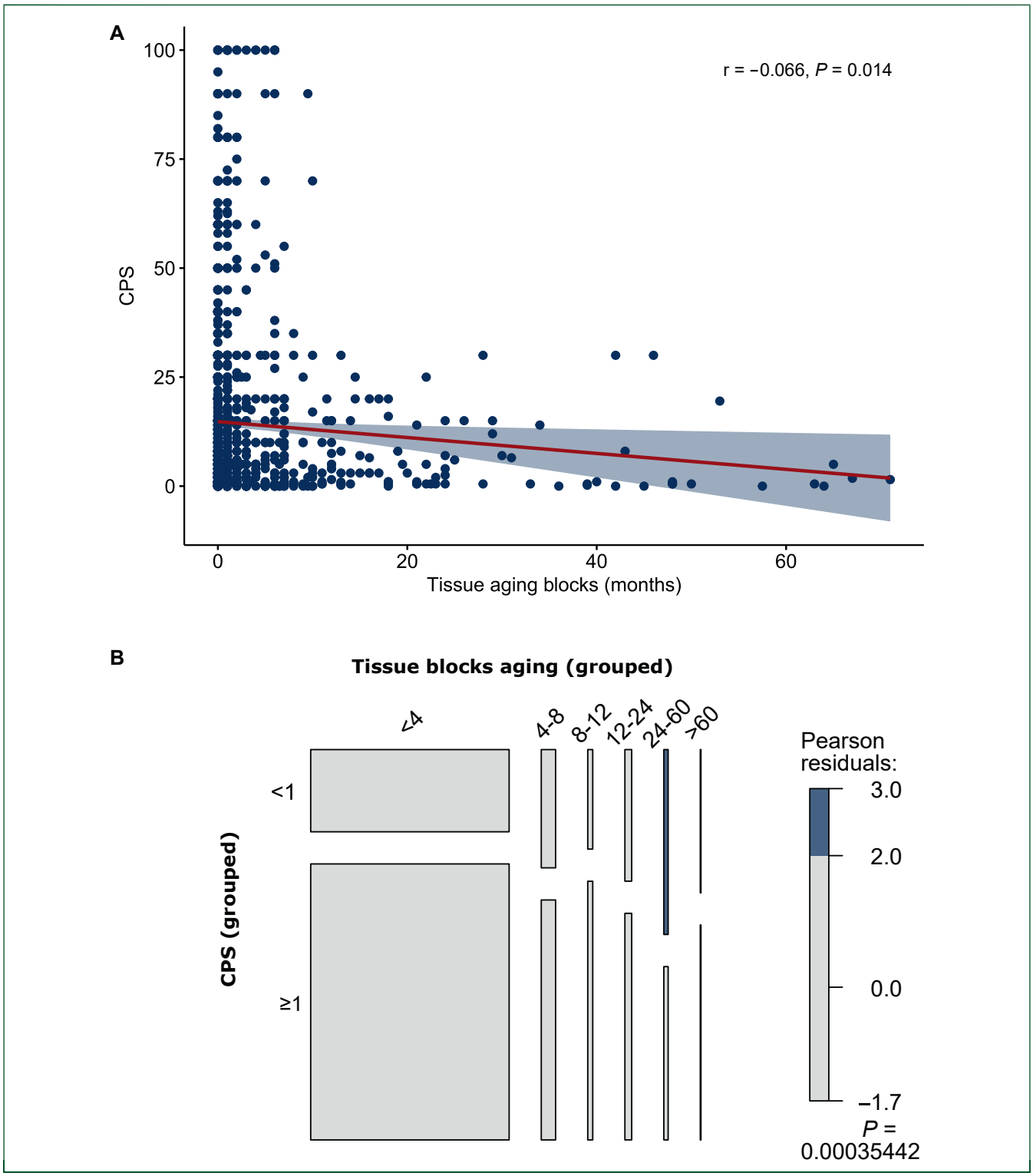

Figure 2. The older the tissue block the lower the combined positive score (CPS) values. (A) Scatter plot illustrating the statistically significant ($P < 0.05$) correlation between CPS and tissue block aging across tissue samples. The plot includes a linear regression line (red line) with 95% confidence interval (gray shaded). Pearson correlation coefficient (r) = −0.066, $P = 0.014$. (B) Mosaic plot visualizing the statistically significant ($P < 0.001$) association between CPS (grouped as <1 and ≥1) and tissue blocks aging (grouped as <4 months, 4-8 months, 8-12 months, 12-24 months, 24-60 months, >60 months). Tile area corresponds to the proportion of observations. Color intensity reflects Pearson residuals from a chi-square test, with blue indicating observed values greater than expected and gray indicating nonsignificant residuals or expected values

0.03). PD-L1 IHC yield was also associated with tissue block aging as a semiquantitative variable ($P < 0.05$) with a critical time window identified at 4-8 months. PD-L1 IHC case workload and aging of tissue slides were not associated with CPS results (Supplementary Figure S8, available at https://doi.org/10.1016/j.esmoop.2025.105846). None

of the assessed variables was associated with TPS values (Supplementary Figure S9, available at https://doi.org/10.1016/j.esmoop.2025.105846). Due to the limited number of variables identified and small sample size, we did not perform a multivariate model with the SCC subgroup.

DISCUSSION

Our study represents the first multicentric, Italian, real-world analysis of PD-L1 workflow and histopathological assessment in GEC. We provide a comprehensive representation of how PD-L1 testing is carried out in routine clinical practice across 28 Italian institutions, and harness data from a large cohort of GEC samples ($n = 1936$), evaluating the distribution of expression according to different scoring systems (CPS, TPS, and TAP), thereby exploring potential sources of variability. Such real-world data is crucial for clinicians to understand the variability they may encounter in PD-L1 results and to contextualize clinical trial findings within their routine clinical practice. As documented in the literature and supported by our data, CPS is the most used scoring system for PD-L1 in the ADCA subgroup of GEC. TAP has recently emerged as a novel and intuitive scoring system that may also be employed for GEC-ADCA, and may play a major role for other ICIs in the near future.[26,27] Most recent guidelines on advanced esophageal SCC have validated CPS (≥10), TPS (≥1%), and TAP score (≥5%) for patient selection for ICIs.[48]

Overall, results from PD-L1 assessment in our study were concordant with data reported in clinical trials, in terms of percentage of PD-L1-positive cases, for both ADCA and SCC. This was especially evident for CPS, considering that 76.1% of ADCA collected via our survey presented a CPS ≥1, which is in line with the range of clinical trial data.[28-31] Of note, the percentage of positive cases according to TPS and TAP was higher than expected in our series. As this finding may be related to several variables (e.g. differences in the scoring system's structure, clone utilized), a dedicated study with more granular analyses is required to identify potential causes of this discrepancy. We observed a relevant difference in TPS positivity between ADCA and SCC subgroups, with SCC exhibiting a higher proportion of TPS ≥1% cases. As the TPS focused on PD-L1 expression of tumor cells rather than immune cells, this observation is well-documented in the literature,[7,32] and further confirms the different biological and clinical behavior of SCC and ADCA.

A deeper analysis of our data has revealed several areas for improvement that warrant further investigation. Firstly, our study revealed a non-negligible inter-institutional heterogeneity in PD-L1 IHC results, with some centers reporting, on average, higher values of PD-L1 than others. This heterogeneity was especially evident within the ADCA subgroup and using the CPS scoring system. Considering the relevance of this variability and related implications on subsequent access to ICI protocols, we explored potential causes. Several variables were identified as significant factors associated with PD-L1 IHC results, including the PD-L1 clone used for IHC staining, tissue aging (both as tissue slides and block aging), specimen origin, and PD-L1 IHC workload. We observed that PD-L1 clones were associated with IHC results, with the SP263 clone mostly showing lower values compared with 22C3. This finding was observed in both ADCA and SCC subgroups, but contrasts with some data in the literature.[26,43-45] Most centers used the SP263 clone as opposed to the 22C3 or 28-8 clones, and this is probably related to the wider use of the Ventana immunostainer in Italy. It is worth remembering that this analysis considered all cases independently, thus referring to the PD-L1 IHC values obtained from different cases. As this is a real-life series, none of the cases were stained with both clones to evaluate and compare the specific performances, and cases were evaluated by numerous different observers. These caveats certainly influence the significance of this result, thereby reducing its impact.

The finding that prolonged tissue aging was associated with lower PD-L1 expression further emphasizes the need for timely and adequate IHC staining. Interestingly, our data suggests that aging of tissue blocks may have a more significant impact on PD-L1 inter-institution variability than tissue slide aging. This implies that information concerning block and slide aging must be included in the pathology report and should be considered by oncologists. Unfortunately, additional data on storage conditions were not collected with the survey. Furthermore, we observed that both tissue slide and block aging were inversely related to the center's workload. We also observed a trend toward higher PD-L1 IHC values in institutions with lower PD-L1 testing volumes. Finally, <70% of centers had a gastrointestinal pathology expert assessing PD-L1 IHC, and only seven centers reviewed complex cases. Overall, these data suggest that the institution's experience/workload may play a crucial role in PD-L1 assessment.

Our study is the second published to date that evaluates potential causes of PD-L1 variability on a nationwide scale. It is important to acknowledge the study of Abe et al., which is a survey on HER2 and PD-L1 testing practices in gastric cancer across Japan, and shares several similarities with our work.[40] Although the Japanese study focused more on the broader landscape of HER2 and PD-L1 testing, our study provides a more in-depth analysis of different scoring systems, preanalytical variables, and their impact on PD-L1 scoring.

Our study documents an overall agreement between real-life PD-L1 evaluation and clinical trials and identifies critical aspects of PD-L1 testing in GEC. However, the retrospective nature and limited sample size in the SCC subgroup may limit the generalization of our findings. Limited response rate (28%; 28/100) of institutions to the survey invitation, a 1-year window for data collection (which may not capture seasonal variations or changes in practice patterns over time), and lack of centralized review and related observer bias are additional limitations of our study. Of note, we did not collect information on whether centers validated their IHC protocols when not using the Companion Diagnostic platform, or whether prior

treatment (i.e. neoadjuvant therapy) had been carried out, especially for surgical samples. Although several studies proved the interchangeability of PD-L1 IHC assays for GEC samples,[26,49,50] collecting data on center participation in external quality control assessment schemes and prior treatment may be useful to limit a potential source of IHC-stain variability. Future studies based on prospective cohorts, higher institutional adherence, or focusing on specific critical areas are warranted to further refine the understanding of factors influencing PD-L1 IHC expression in GEC.

Another key consideration is the pivotal role of PD-L1 expression in guiding therapeutic strategies.[10,48,51] Currently, regulatory approvals and clinical indications for ICIs are closely linked with specific PD-L1 (CPS, TPS, and TAP) expression thresholds. These thresholds not only determine drug eligibility but also aim to predict the expected benefit of combining immunotherapy with chemotherapy. However, recent evidence suggests that the clinical benefit of ICIs in ADCA patients with low PD-L1 expression, particularly those with CPS values <1 or between 1 and 4, remains questionable.[28,52] In such borderline cases where the PD-L1 assessment and reliability are even more challenging, alternative targeted approaches may deserve consideration. For instance, patients whose tumors express claudin 18.2 might benefit more from claudin-targeted therapies, which are currently gaining clinical traction.[53,54] Furthermore, the therapeutic landscape is likely to become even more complex with the expected introduction of fibroblast growth factor receptor (FGFR) inhibitors, particularly for tumors harboring FGFR2 alterations.[55] These new treatment options will need to be carefully weighed against the potential, but somehow limited, benefit of immunotherapy in patients with low or intermediate PD-L1 expression. Altogether, these considerations underscore the pressing need for robust, reproducible, and clinically reliable PD-L1 assessment to ensure optimal patient selection in an increasingly crowded therapeutic landscape.

In conclusion, this multicentric, real-world Italian study provides valuable insights into the current state of PD-L1 IHC assessment in GEC. Our data confirm the alignment of real-world PD-L1 expression with clinical trials, but the observed inter-institutional heterogeneity and the influence of pre-analytical variables, particularly tissue aging and PD-L1 clone, still pose significant controversies in real-world practice. By addressing these issues, we can improve the reliability of PD-L1 testing and ensure that patients with GEC receive the most appropriate and effective treatment.

FUNDING

This study was supported by the Italian Ministry of Health (Ricerca Corrente).

DISCLOSURE

FG: consulting or advisory role: MSD, GSK, Beigene; invited speaker: Pierre Fabbre, MSD, GSK, Servier, Astellas, Incyte, BMS, AstraZeneca, BeiGene, Daiichi-Sankyo, Amgen. PP: consulting or advisory role: Amgen, Astellas, AstraZeneca, BMS, BeOne, Incyte, Servier, Daiichi-Sankyo, GSK, MSD, Diaceutics. AV: consulting or advisory role (with honoraria): Amgen, MSD Italia, Sanofi, BeOne. AC: research support (to institution): Roche Diagnostics Italia SpA; consulting or advisory role: Roche Diagnostics Italia SpA; invited speaker: HOLOGIC Italia. MP: consulting: Astrazeneca. AP: research funding: Amgen, Astra Zeneca; travel, accommodations, expenses: BeOne, Pierre Fabre, Amgen, Merck Serono; consulting or advisory role: BeOne, Pierre Fabre, Amgen, Merck Serono, Amgen, Astra Zeneca, Daiichi, Nordic Pharma, Servier, Bayer, GSK, MSD, BMS; principal investigator: Amgen, Merck Serono, GSK; invited speaker: BeOne, Pierre Fabre, Amgen, Merck Serono, Amgen, Astra Zeneca, Daiichi, Servier, Bayer, GSK, MSD, BMS. MF: Amgen, Astellas, BMS, Diaceutics, Diapath, Eli Lilly, GSK, Incyte, IQvia, MSD, Novartis, Sanofi, Thermofisher, Roche. LM: consulting or advisory role: MSD Italia, BeOne, Sanofi, Astrazeneca; invited speaker: MSD, Astrazeneca, Daiichi-Sankyo. All other authors declare no conflicts of interest.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.
2. Siewert JR, Bottcher K, Stein HJ, Roder JD. Relevant prognostic factors in gastric cancer: ten-year results of the German Gastric Cancer Study. *Ann Surg*. 1998;228(4):449-461.
3. Hagen JA, DeMeester SR, Peters JH, Chandrasoma P, DeMeester TR. Curative resection for esophageal adenocarcinoma: analysis of 100 en bloc esophagectomies. *Ann Surg*. 2001;234(4):520-530. ; discussion 530-521.
4. Ronellenfitsch U, Schwarzbach M, Hofheinz R, et al. Preoperative chemo(radio)therapy versus primary surgery for gastroesophageal adenocarcinoma: systematic review with meta-analysis combining individual patient and aggregate data. *Eur J Cancer*. 2013;49(15):3149-3158.
5. Kang YK, Boku N, Satoh T, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390(10111):2461-2471.
6. Marabelle A, Le DT, Ascierto PA, et al. Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J Clin Oncol*. 2020;38(1):1-10.
7. Janjigian YY, Shitara K, Moehler M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet*. 2021;398(10294):27-40.
8. Sun JM, Shen L, Shah MA, et al. Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. *Lancet*. 2021;398(10302):759-771.
9. Rha SY, Wyrwicz LS, Weber PEY, et al. VP1-2023: Pembrolizumab (pembro) plus chemotherapy (chemo) as first-line therapy for advanced HER2-negative gastric or gastroesophageal junction (G/GEJ) cancer: phase III KEYNOTE-859 study. *Ann Oncol*. 2023;34(3):319-320.
10. Lordick F, Carneiro F, Cascinu S, et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol*. 2022;33(10):1005-1020.

11. Janjigian YY, Al-Batran SE, Wainberg ZA, et al. Perioperative durvalumab in gastric and gastroesophageal junction cancer. *N Engl J Med*. 2025;393(3):217-230.
12. Matias-Guiu X, Raspollini MR, Kulka J, et al. The role of the pathologist in the design and conducting of biomarker-driven clinical trials in cancer: position paper of the European Society of Pathology. *Virchows Arch*. 2025;486(2):207-214.
13. Angerilli V, Fassan M, Parente P, et al. A practical approach for PD-L1 evaluation in gastroesophageal cancer. *Pathologica*. 2023;115(2):57-70.
14. Mastracci L, Grillo F, Parente P, et al. PD-L1 evaluation in the gastrointestinal tract: from biological rationale to its clinical application. *Pathologica*. 2022;114(5):352-364.
15. Parente P, Grillo F, Vanoli A, Mastracci L, Fassan M. A charming molecular profile in the wrong tissue: the pathologist in the era of molecular diagnostics in the gastrointestinal tract. *Expert Rev Mol Diagn*. 2023;23(12):1045-1047.
16. Fassan M, Scarpa A, Remo A, et al. Current prognostic and predictive biomarkers for gastrointestinal tumors in clinical practice. *Pathologica*. 2020;112(3):248-259.
17. Doroshow DB, Bhalla S, Beasley MB, et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol*. 2021;18(6):345-362.
18. Zhang M, Dong Y, Liu H, et al. The clinicopathological and prognostic significance of PD-L1 expression in gastric cancer: a meta-analysis of 10 studies with 1,901 patients. *Sci Rep*. 2016;6:37933.
19. Tumeh PC, Harview CL, Yearley JH, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014;515(7528):568-571.
20. Herbst RS, Soria JC, Kowanetz M, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014;515(7528):563-567.
21. Badve SS, Kumar GL, Lang T, et al. AI based PD-L1 CPS quantifier software to identify more patients for checkpoint therapy in gastric cancer at pathologist-level interobserver concordance. *J Clin Oncol*. 2024;42(suppl 16):2633. 2633.
22. Kim HN, Jang J, Heo YJ, et al. PD-L1 expression in gastric cancer determined by digital image analyses: pitfalls and correlation with pathologist interpretation. *Virchows Arch*. 2020;476(2):243-250.
23. Dai L, Yin J, Xin X, et al. An interpretable machine learning model based on computed tomography radiomics for predicting programmed death ligand 1 expression status in gastric cancer. *Cancer Imaging*. 2025;25(1):31.
24. Lee T, Cho SI, Choi S, et al. Performance validation of an artificial intelligence-powered PD-L1 combined positive score analyzer in six cancer types. *J Clin Oncol*. 2023;41(suppl 16):e13553. e13553.
25. Caputo A, Angerilli V, Gambella A, et al. Immunohistochemical biomarker scoring in gastroesophageal cancers: can computers help us? *Pathol Res Pract*. 2025;272:156068.
26. Klempner SJ, Cowden ES, Cytryn SL, et al. PD-L1 immunohistochemistry in gastric cancer: comparison of combined positive score and tumor area positivity across 28-8, 22C3, and SP263 assays. *JCO Precis Oncol*. 2024;8:e2400230.
27. Liu C, Fang F, Kong Y, ElGabry EA. Tumor area positivity (TAP) score of programmed death-ligand 1 (PD-L1): a novel visual estimation method for combined tumor cell and immune cell scoring. *Diagn Pathol*. 2023;18(1):48.
28. Leone AG, Mai AS, Fong KY, et al. Immune checkpoint inhibitors in advanced gastroesophageal adenocarcinoma: a series of patient-level meta-analyses in different programmed death-ligand 1 subgroups. *ESMO Open*. 2024;9(11):103962.
29. Rha SY, Oh DY, Yañez P, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for HER2-negative advanced gastric cancer (KEYNOTE-859): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2023;24(11):1181-1195.
30. Janjigian YY, Kawazoe A, Bai Y, et al. Pembrolizumab plus trastuzumab and chemotherapy for HER2-positive gastric or gastro-oesophageal junction adenocarcinoma: interim analyses from the phase 3 KEYNOTE-811 randomised placebo-controlled trial. *Lancet*. 2023;402(10418):2197-2208.
31. Qiu MZ, Oh DY, Kato K, et al. Tislelizumab plus chemotherapy versus placebo plus chemotherapy as first line treatment for advanced gastric or gastro-oesophageal junction adenocarcinoma: RATIONALE-305 randomised, double blind, phase 3 trial. *BMJ*. 2024;385:e078876.
32. Kang YK, Chen LT, Ryu MH, et al. Nivolumab plus chemotherapy versus placebo plus chemotherapy in patients with HER2-negative, untreated, unresectable advanced or recurrent gastric or gastro-oesophageal junction cancer (ATTRACTION-4): a randomised, multicentre, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2022;23(2):234-247.
33. Kojima T, Shah MA, Muro K, et al. Randomized phase III KEYNOTE-181 study of pembrolizumab versus chemotherapy in advanced esophageal cancer. *J Clin Oncol*. 2020;38(35):4138-4148.
34. Kato K, Doki Y, Chau I, et al. Nivolumab plus chemotherapy or ipilimumab versus chemotherapy in patients with advanced esophageal squamous cell carcinoma (CheckMate 648): 29-month follow-up from a randomized, open-label, phase III trial. *Cancer Med*. 2024;13(9):e7235.
35. Shen L, Kato K, Kim SB, et al. Tislelizumab versus chemotherapy as second-line treatment for advanced or metastatic esophageal squamous cell carcinoma (RATIONALE-302): a randomized phase III study. *J Clin Oncol*. 2022;40(26):3065-3076.
36. Xu J, Kato K, Raymond E, et al. Tislelizumab plus chemotherapy versus placebo plus chemotherapy as first-line treatment for advanced or metastatic oesophageal squamous cell carcinoma (RATIONALE-306): a global, randomised, placebo-controlled, phase 3 study. *Lancet Oncol*. 2023;24(5):483-495.
37. Angerilli V, Ghelardi F, Nappo F, et al. Claudin-18.2 testing and its impact in the therapeutic management of patients with gastric and gastroesophageal adenocarcinomas: a literature review with expert opinion. *Pathol Res Pract*. 2024;254:155145.
38. Yamada A, Takeuchi M, Komatsu K, Bonate PL, Poondru S, Yang J. Population PK and exposure-response analyses of zolbetuximab in patients with locally advanced unresectable or metastatic G/GEJ adenocarcinoma. *Clin Transl Sci*. 2025;18(7):e70280.
39. Ajani JA, D'Amico TA, Bentrem DJ, et al. Gastric cancer, version 2. 2025, NCCN Clinical Practice Guidelines In Oncology. *J Natl Compr Canc Netw*. 2025;23(5):169-191.
40. Abe H, Kuwata T, Kushima R, Ushiku T. Nationwide survey on HER2 and PD-L1 testing practices in gastric cancer across Japan. *Gastric Cancer*. 2025;28(2):294-300.
41. Gambella A, Porro L, Pigozzi S, Fiocca R, Grillo F, Mastracci L. Section detachment in immunohistochemistry: causes, troubleshooting, and problem-solving. *Histochem Cell Biol*. 2017;148(1):95-101.
42. Grillo F, Pigozzi S, Ceriolo P, Calamaro P, Fiocca R, Mastracci L. Factors affecting immunoreactivity in long-term storage of formalin-fixed paraffin-embedded tissue sections. *Histochem Cell Biol*. 2015;144(1):93-99.
43. Barbera A, Gonzalez J, Martin M, et al. Impact of prolonged ischemia on the immunohistochemical expression of programmed death ligand 1 (PD-L1). *Appl Immunohistochem Mol Morphol*. 2023;31(9):607-612.
44. Barbera A, Marginet Flinch R, Martin M, et al. The immunohistochemical expression of programmed death ligand 1 (PD-L1) is affected by sample overfixation. *Appl Immunohistochem Mol Morphol*. 2021;29(1):76-81.
45. Lawson NL, Scorer PW, Williams GH, Vandenberghe ME, Ratcliffe MJ, Barker C. Impact of decalcification, cold ischemia, and deglycosylation on performance of programmed cell death ligand-1 antibodies with different binding epitopes: comparison of 7 clones. *Mod Pathol*. 2023;36(9):100220.
46. Bragoni A, Gambella A, Pigozzi S, et al. Quality control in diagnostic immunohistochemistry: integrated on-slide positive controls. *Histochem Cell Biol*. 2017;148(5):569-573.
47. Xu J, Jiang H, Pan Y, et al. Sintilimab plus chemotherapy for unresectable gastric or gastroesophageal junction cancer: the ORIENT-16 randomized clinical trial. *JAMA*. 2023;330(21):2064-2074.

48. Obermannova RL, Leong T, ESMO Guidelines Committee. ESMO Clinical Practice Guideline interim update on the treatment of locally advanced oesophageal and oesophagogastric junction adenocarcinoma and metastatic squamous-cell carcinoma. *ESMO Open*. 2025;10 (2):104134.
49. Ahn S, Kim KM. PD-L1 expression in gastric cancer: interchangeability of 22C3 and 28-8 pharmDx assays for responses to immunotherapy. *Mod Pathol*. 2021;34(9):1719-1727.
50. Park Y, Koh J, Na HY, et al. PD-L1 testing in gastric cancer by the combined positive score of the 22C3 pharmDx and SP263 assay with clinically relevant cut-offs. *Cancer Res Treat*. 2020;52(3):661-670.
51. Lordick F, Candia Montero L, Castelo-Branco L, et al. ESMO Gastric Cancer Living Guideline v1.4. 2024. *Ann Oncol*. 2022;33(10):1005-1020.
52. Formica V, Morelli C, Fornaro L, et al. PD-L1 thresholds predict efficacy of immune checkpoint inhibition in first-line treatment of advanced gastroesophageal adenocarcinoma. a systematic review and meta-analysis of seven phase III randomized trials. *ESMO Open*. 2024;9 (11):103967.
53. Shah MA, Shitara K, Ajani JA, et al. Zolbetuximab plus CAPOX in CLDN18.2-positive gastric or gastroesophageal junction adenocarcinoma: the randomized, phase 3 GLOW trial. *Nat Med*. 2023;29(8): 2133-2141.
54. Shitara K, Lordick F, Bang Y-J, et al. Zolbetuximab plus mFOLFOX6 in patients with CLDN18.2-positive, HER2-negative, untreated, locally advanced unresectable or metastatic gastric or gastro-oesophageal junction adenocarcinoma (SPOTLIGHT): a multicentre, randomised, double-blind, phase 3 trial. *The Lancet*. 2023;401(10389):1655-1668.
55. Wainberg ZA, Kang YK, Lee KW, et al. Bemarituzumab as first-line treatment for locally advanced or metastatic gastric/gastroesophageal junction adenocarcinoma: final analysis of the randomized phase 2 FIGHT trial. *Gastric Cancer.* 2024;27(3):558-570.

APPENDIX A

On behalf of the RELIABL Study Group

The RELIABL Study Group: L. Albarello[13], M. R. Ambrosio[14], M. R. Aprile[15], I. Banchelli[15], V. Barresi[16], P. R. Billo[17], S. Bruno[18], P. Cassoni[19,20], B. Cerbelli[21], M. D'Armiento[22], G. De Lisi[4], E. C. Falco[19], D. Fiorini[16], M. Fortunato[23], S. Lazzi[24], S. La Rosa[17], G. Locci[25], M. C. Macciomei[26], M. Martini[27], L. Melocchi[28], P. Morbini[29], V. Nardini[30], I. Panarese[31], B. G. Pozzi[32], A. Remo[33], L. Savino[34], P. Spaggiari[35] & F. Sommavilla[13]

[13]Università Vita Salute — IRCCS San Raffaele Milano, Milano; [14]University of Calabria (UNICAL), Azienda Ospedaliera Universitaria "SS. Annunziata", Cosenza; [15]ASL Romagna, Ravenna; [16]Dipartimento di Diagnostica e Sanità Pubblica, Università di Verona, Verona; [17]Anatomia Patologica, ASST dei Sette Laghi, Varese; [18]Pathology Unit, ASL2, Savona; [19]Città della Salute e della Scienza University Hospital, Turin; [20]Pathology Unit, Department of Medical Sciences, University of Turin, Turin; [21]Department of Radiological, Oncological and Pathological Sciences, Sapienza University of Rome, Policlinico Umberto I, Rome; [22]AOU "Federico II" Naples University, Napoli; [23]Azienda Ospedaliera Santa Croce e Carle, Cuneo; [24]Institute of Pathology, Azienda Ospedaliera Universitaria Senese, Policlinico Le Scotte, Siena; [25]ARNAS G. Brotzu, Piazzale A. Ricchi, Cagliari; [26]Azienda Ospedaliera San Camillo Forlanini, Roma; [27]AOU Policlinico "G. Martino", Messina; [28]Pathology Unit, Fondazione Poliambulanza, Istituto Ospedaliero, Brescia; [29]Anatomia patologica, E.O. Ospedali galliera, Genova; [30]Ospedale Cisanello, Azienda Ospedaliero-Universitaria Pisana, Pisa; [31]A.O.U. "Luigi Vanvitelli" Napoli; [32]S.C Anatomia Patologica, ASST Lecco, Lecco; [33]Pathology Unit, ULSS9 "Scaligera", Verona; [34]Policlinico Tor Vergata, Roma; [35]Unit of Anatomic Pathology, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy