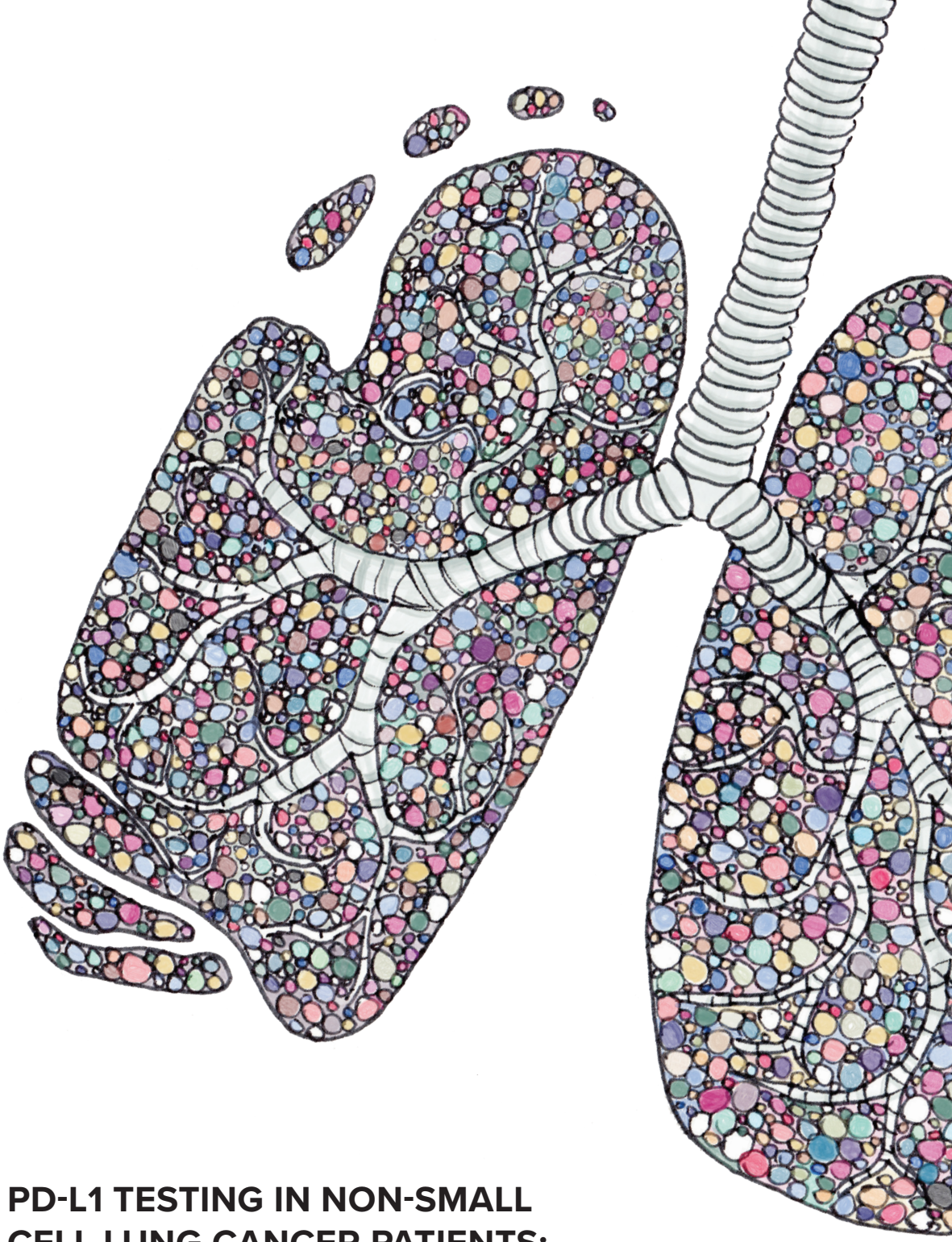


PD-L1 TESTING IN NON-SMALL CELL LUNG CANCER PATIENTS: A STUDY ON INTERLABORATORY VARIATION

Bregje M. Koomen

PD-L1 testing in non-small cell lung cancer patients: a study on interlaboratory variation

Bregje M. Koomen

PD-L1 testing in non-small cell lung cancer patients: a study on interlaboratory variation
Bregje M. Koomen
ISBN: 978-94-6496-072-3
DOI: https://doi.org/10.33540/2202

Cover: Arjan Helmer
Layout: Ilse Modder – www.ilsemodder.nl
Printing: Gildeprint – www.gildeprint.nl

Part of the research described in this thesis was financially supported by AstraZeneca, MSD, and Roche.

PD-L1 testing in non-small cell lung cancer patients: a study on interlaboratory variation

PD-L1 bepaling bij patiënten met niet-kleincellig longkanker: een studie naar interlaboratoriumvariatie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 2 mei 2024 des middags te 4.15 uur

door

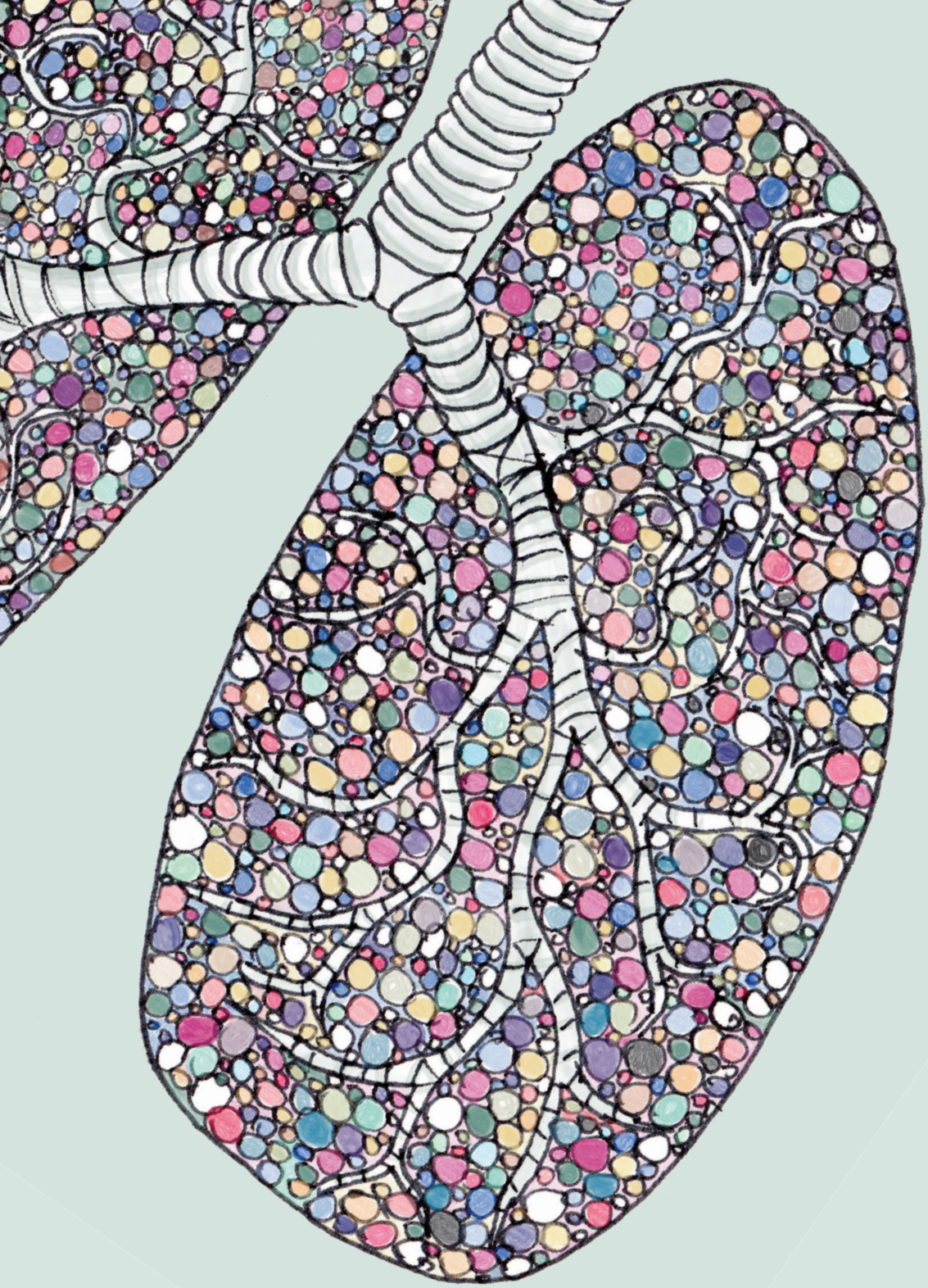
Bregje Maria Koomen

geboren op 10 november 1990
te Veghel

Promotoren:
Prof. dr. M.R. van Dijk
Prof. dr. S.M. Willems

Beoordelingscommissie:
Prof. dr. P.J. van Diest
Prof. dr. C.H. van Gils
Prof. dr. E.F. Smit
Prof. dr. K.P.M. Suijkerbuijk
Dr. J.H. von der Thüsen

TABLE OF CONTENTS



1

GENERAL INTRODUCTION AND THESIS OUTLINE

GENERAL INTRODUCTION

Lung cancer is the second most commonly diagnosed type of cancer worldwide and accounts for the most cancer related deaths in men and women combined[1]. Two main histopathological types of lung cancer exist, i.e. small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), with the latter accounting for 85% of lung cancer cases[2]. In the Netherlands, around 10.000 patients are diagnosed with NSCLC each year[3]. At diagnosis, almost half (48%) of all newly diagnosed patients in the Netherlands have stage IV disease, indicating the presence of metastases[3]. Patients with stage IV disease without actionable mutations in driver genes, such as *EGFR* or *ALK*, who were treated with chemotherapy showed a 5-year survival rate of only 6% between 2017 and 2021[4]. Introduction of immunotherapy and combinations of chemotherapy and immunotherapy led to improved survival, with 5-year survival rates now being 25% for Dutch patients treated with immunotherapy and 17% for patients treated with a combination of chemotherapy and immunotherapy[4].

Immunotherapy in NSCLC

Immunotherapy for patients with NSCLC, which was first introduced in clinical practice in 2015[5], is based on blockade of the programmed death-1 (PD-1) receptor, an immune checkpoint protein primarily expressed on activated T-cells[6], or one of its ligands programmed death-ligand 1 (PD-L1). Binding of PD-1 to PD-L1 results in suppression of T-cell activity[6, 7]. PD-L1 is expressed in both lymphoid and non-lymphoid tissues[8, 9], and is involved in preventing immune-mediated tissue damage by limiting excessive T-cell activation[10]. PD-L1 expression is also found in various cancers, resulting in resistance of tumor cells to antitumor immune response of T-cells[7, 8]. Preventing PD-1 from binding to PD-L1 may restore T-cell function and, thus, aid in destruction of tumor cells[10].

The first PD-1/PD-L1 blockade agent used in treatment of NSCLC patients was nivolumab, an anti-PD-1 antibody, which was approved for second line treatment of advanced NSCLC based on clinical trials showing improved survival in comparison to chemotherapy[11, 12]. Other PD-1/PD-L1 blockade agents, i.e. pembrolizumab (an anti-PD-1 antibody) and atezolizumab (an anti-PD-L1 antibody) showed beneficial results as second line treatment agents in comparison to chemotherapy, too[13, 14]. Importantly, these drugs also showed a favorable safety profile compared to chemotherapy[15]. Subsequently, the efficacy of PD-1/PD-L1 blockade agents as first-line treatment was assessed. Pembrolizumab showed significant improvement of both overall and progression free survival compared with chemotherapy in patients with advanced NSCLC and expression of PD-L1 on at least 50% of tumor cells[16]. More recently, monotherapy with cemiplimab, a relatively new anti-PD-1 antibody, and with atezolizumab showed similar results in this patient group[17,18]. Other trials showed improved survival among patients treated with chemotherapy combined with immunotherapy in

comparison with chemotherapy alone, regardless of PD-L1 expression[19-22]. The efficacy of another anti-PD-L1 antibody called durvalumab was tested in patients with unresectable locally advanced (stage III) NSCLC, who received durvalumab as consolidation treatment after chemoradiotherapy. This treatment resulted in significantly longer progression free survival and overall survival compared with placebo[23, 24].

Currently, treatment guidelines recommend first-line prescription of pembrolizumab, cemiplimab or atezolizumab as monotherapy to patients with metastatic NSCLC with ≥50% PD-L1 expression on tumor cells[25, 26]. For patients whose tumors express PD-L1 in <50% of tumor cells, combinations of immunotherapy and chemotherapy are recommended. Furthermore, a PD-L1 expression of ≥1% is required for second line treatment with pembrolizumab[25, 27]. Use of durvalumab is recommended as consolidation treatment after chemoradiotherapy for patients with unresectable stage III disease, although the European Medicines Agency (EMA) recommends that these patients should have an expression of ≥1% on tumor cells[28]. Thus, as becomes clear from these guidelines, treatment decisions are guided by PD-L1 expression on tumor cells.

PD-L1 immunohistochemistry

Assessment of PD-L1 expression on tumor cells is performed by pathologists through immunohistochemistry (IHC), which visualizes PD-L1 present on tumor cells by binding of labelled antibodies directed against the PD-L1 protein. Pathologists score the level of PD-L1 expression by determining the percentage of all tumor cells that express PD-L1, which is called the tumor proportion score (TPS). The TPS is then used by clinicians to determine the right course of treatment. Hence, it is very important that PD-L1 expression is stained and scored accurately. This accuracy is influenced by pre-analytic, analytic and post-analytic factors.

Analytic factors

One important analytic factor in PD-L1 IHC is the antibody that is used in the immunostaining protocol. Various IHC antibodies directed at PD-L1 were developed. In fact, with each of the initial PD-1/PD-L1 blockade agents, a separate PD-L1 IHC antibody was introduced: antibody clone 28-8 (Dako) for nivolumab[29], antibody clone 22C3 (Dako) for pembrolizumab[30], antibody clone SP142 (Ventana) for atezolizumab[31], and antibody clone SP263 (Ventana) for durvalumab[32]. The use of multiple antibodies to test for PD-L1 expression in one laboratory is generally not feasible, since this would be too expensive and time consuming, and the number of tests that can be performed is often restricted due to limited tissue availability[33]. Consequently, laboratories will mostly use only one of the available PD-L1 IHC antibodies, and may differ in their choice of antibody. Additionally, instead of using one of the available commercial assays, some laboratories may choose to use a laboratory-developed test (LDT), which is usually a lot less costly[34]. It is unsure

whether these various commercial assays and LDTs can be used interchangeably in clinical practice, or whether this would in fact result in a high amount of variation in PD-L1 positivity between laboratories.

Post-analytic factors

When it comes to post-analytic factors that influence PD-L1 IHC accuracy, obviously scoring of PD-L1 expression by pathologists has a major impact. Since scoring entails estimating the percentage of tumor cells that express PD-L1, it is susceptible to interobserver variability[35]. Considering the crucial role that PD-L1 IHC plays as predictive biomarker for response to anti-PD-1/PD-L1 immunotherapy, it is important that a high amount of both intraobserver and interobserver variability is avoided.

Pre-analytic factors

Finally, various pre-analytic factors may influence PD-L1 IHC results, such as the sample site of the material that is tested (e.g. primary tumor vs. metastasis) or the specimen size (e.g. biopsy vs. resection)[35, 36]. Another possible issue is the performance of PD-L1 tests on cytologic specimens instead of on histologic specimens. Since NSCLC is frequently diagnosed when the disease is already advanced, minimally invasive techniques such as fine needle aspiration (FNA) are often used to collect diagnostic material[35, 37]. In those instances, PD-L1 testing may have to be performed on these cytologic specimens as well, even though PD-L1 immunostaining on cytology was not validated in clinical trials[35]. Various ways of processing cytologic specimens exist, and some of these incorporate non-formaldehyde based fixatives[38, 39]. It has been demonstrated that alcohol-based fixatives may have a detrimental effect on the immunoreactivity of several IHC antibodies[40-42]. If these negative effects of alcohol-based fixatives are also seen when using PD-L1 IHC antibodies, this could impact treatment decisions in clinical practice.

Interlaboratory variation in clinical practice

Variation between laboratories in pre-analytic, analytic and post-analytic variables involved in PD-L1 IHC could potentially result in a high degree of interlaboratory variation in PD-L1 positivity. This would be problematic, since it could lead to patients being denied effective treatment options with a PD-1/PD-L1 blockade agent or being exposed to unnecessary toxicity of added chemotherapy. As an example, appreciable variation in hormone- and HER2-receptor assessment for breast cancer patients has been demonstrated between pathology laboratories in a real-world setting[43, 44]. This raises the question how much variation actually exists between laboratories that perform PD-L1 IHC in clinical practice.

THESIS OUTLINE

The first part of this thesis evaluates the influence of various pre-analytic, analytic and post-analytic factors on immunohistochemical PD-L1 expression. In **chapter 2** we systematically review the available literature on comparability of PD-L1 IHC commercial assays and LDTs, focusing on interassay concordance, interobserver concordance and interlaboratory concordance. **Chapter 3** assesses concordance of PD-L1 immunostaining between cytology cell blocks and their histologic counterparts, using FNA samples and histologic tissue from the same resected lung tumor. Furthermore, the effects of various fixatives and fixation times on PD-L1 immunostaining were studied in PD-L1–expressing cell lines. In **chapter 4** we study the effect of ethanol pre-fixation on PD-L1 immunostaining using paired ethanol-fixed and formalin-fixed endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) samples from NSCLC patients.

We then move on to assessing variation in PD-L1 IHC between pathology laboratories in clinical practice, using real-world clinical pathology data. **Chapter 5** assesses interlaboratory variation in PD-L1 positivity rates on a nationwide level at both the 1% cutoff and the 50% cutoff. Moreover, analysis of interlaboratory variation is performed for PD-L1 positivity rates based on histological material and PD-L1 positivity rates based on cytological material, separately. In **chapter 6** we evaluate the amount of variation in fixation and cell block processing of cytology samples between pathology laboratories in the Netherlands. Subsequently, we assess whether differences in fixation and cell block processing of cytologic specimens influence interlaboratory variation in PD-L1 positivity. The results of this thesis and future perspectives are discussed in **chapter 7**.

REFERENCES

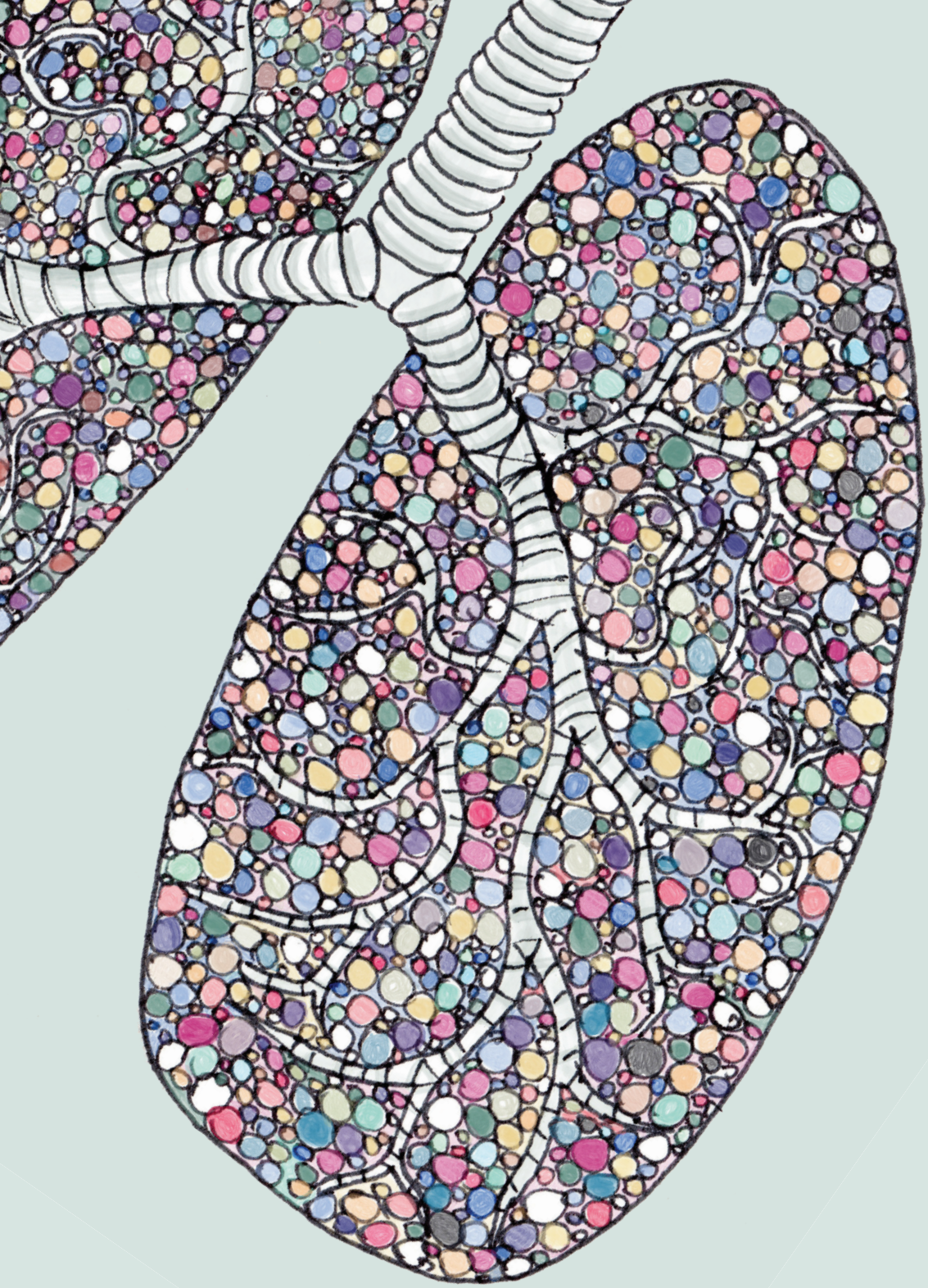
1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.
2. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. Cancer Manag Res. 2019;11:943-953.
3. Incidence Lung Cancer. Netherlands Cancer Registry, Integraal Kankercentrum Nederland (IKNL), 2023. Available from: https://iknl.nl/kankersoorten/longkanker/registratie/incidentie (Accessed 7 November 2023).
4. Survival Lung Cancer. Netherlands Cancer Registry, Integraal Kankercentrum Nederland (IKNL), 2023. Available from: https://iknl.nl/kankersoorten/longkanker/registratie/overleving (Accessed 7 November 2023).
5. Kazandjian D, Suzman DL, Blumenthal G, et al. FDA Approval Summary: Nivolumab for the Treatment of Metastatic Non-Small Cell Lung Cancer With Progression On or After Platinum-Based Chemotherapy. Oncologist. 2016;21(5):634-642.
6. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. Annu Rev Immunol. 2008;26:677-704.
7. Xia L, Liu Y, Wang Y. PD-1/PD-L1 Blockade Therapy in Advanced Non-Small-Cell Lung Cancer: Current Status and Future Directions. Oncologist. 2019;24(Suppl 1):S31-S41.
8. Freeman GJ, Long AJ, Iwai Y, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J Exp Med. 2000;192(7):1027-1034.
9. Okazaki T, Honjo T. PD-1 and PD-1 ligands: from discovery to clinical application. Int Immunol. 2007;19(7):813-824.
10. Lee J, Ahn E, Kissick HT, Ahmed R. Reinvigorating Exhausted T Cells by Blockade of the PD-1 Pathway. For Immunopathol Dis Therap. 2015;6(1-2):7-17.
11. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373(17):1627-1639.
12. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373(2):123-135.
13. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387(10027):1540-1550.
14. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 2017;389(10066):255-265.
15. Ramos-Esquivel A, van der Laat A, Rojas-Vigott R, Juarez M, Corrales-Rodriguez L. Anti-PD-1/anti-PD-L1 immunotherapy versus docetaxel for previously treated advanced non-small cell lung cancer: a systematic review and meta-analysis of randomised clinical trials. ESMO Open. 2017;2(3):e000236.
16. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2016;375(19):1823-1833.
17. Sezer A, Kilickap S, Gümüş M, et al. Cemiplimab monotherapy for first-line treatment of advanced non-small-cell lung cancer with PD-L1 of at least 50%: a multicentre, open-label, global, phase 3, randomised, controlled

trial. The Lancet. 2021;397(10274):592-604.

18. Jassem J, de Marinis F, Giaccone G, et al. Updated Overall Survival Analysis From IMpower110: Atezolizumab Versus Platinum-Based Chemotherapy in Treatment-Naive Programmed Death-Ligand 1–Selected NSCLC. J Thorac Oncol. 2021;16(11):1872-1882.
19. Garassino MC, Gadgeel S, Esteban E, et al. Patient-reported outcomes following pembrolizumab or placebo plus pemetrexed and platinum in patients with previously untreated, metastatic, non-squamous non-small-cell lung cancer (KEYNOTE-189): a multicentre, double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol. 2020;21(3):387-397.
20. Socinski MA, Jotte RM, Cappuzzo F, et al. Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC. N Engl J Med. 2018;378(24):2288-2301.
21. West H, McCleod M, Hussein M, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2019;20(7):924-937.
22. Paz-Ares L, Luft A, Vicente D, et al. Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. N Engl J Med. 2018;379(21):2040-2051.
23. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. N Engl J Med. 2017;377(20):1919-1929.
24. Antonia SJ, Villegas A, Daniel D, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. N Engl J Med. 2018;379(24):2342-2350.
25. Hendriks LE, Kerr KM, Menis J, et al. Non-oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023;34(4):358-376.
26. Akinboro O, Larkins E, Pai-Scherf L, et al. FDA Approval Summary: Pembrolizumab, Atezolizumab, and Cemiplimab-rwlc as Single Agents for First-Line Treatment of Advanced/Metastatic PD-L1–High NSCLC. Clin Cancer Res. 2022;28(11):2221-2228.
27. Sul J, Blumenthal GM, Jiang X, He K, Keegan P, Pazdur R. FDA Approval Summary: Pembrolizumab for the Treatment of Patients With Metastatic Non-Small Cell Lung Cancer Whose Tumors Express Programmed Death-Ligand 1. Oncologist. 2016;21(5):643-650.
28. Remon J, Soria JC, Peters S, Esmo Guidelines Committee. Early and locally advanced non-small-cell lung cancer: an update of the ESMO Clinical Practice Guidelines focusing on diagnosis, staging, systemic and local therapy. Ann Oncol. 2021;32(12):1637-1642.
29. PD-L1 IHC 28-8 pharmDx Overview. Agilent Dako, 2021. Available from: https://www.agilent.com/en/product/pharmdx/pd-l1-ihc-28-8-overview (Accessed 26 November 2021).
30. PD-L1 IHC 22C3 pharmDx Overview. Agilent Dako, 2021. Available from: https://www.agilent.com/en/product/pharmdx/pd-l1-ihc-22c3-pharmdx-overview (Accessed 26 November 2021).
31. VENTANA PD-L1 (SP142) Assay (CE IVD). Roche Diagnostics, 2021. Available from: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp142-assay2.html (Accessed 26 November 2021).
32. VENTANA PD-L1 (SP263) Assay (CE IVD). Roche Diagnostics, 2021. Available from: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp263-assay2.html (Accessed 26 November 2021).
33. Kerr KM, Hirsch FR. Programmed Death Ligand-1 Immunohistochemistry: Friend or Foe? Arch Pathol Lab Med.

2016;140(4):326-331.

34. Savic Prince S, Bubendorf L. Predictive potential and need for standardization of PD-L1 immunohistochemistry. Virchows Arch. 2019;474(4):475-484.
35. Lantuejoul S, Sound-Tsao M, Cooper WA, et al. PD-L1 Testing for Lung Cancer in 2019: Perspective From the IASLC Pathology Committee. J Thorac Oncol. 2020;15(4):499-519.
36. Thunnissen E, Kerr KM, Dafni U, et al. Programmed death-ligand 1 expression influenced by tissue sample size. Scoring based on tissue microarrays' and cross-validation with resections, in patients with, stage I-III, non-small cell lung carcinoma of the European Thoracic Oncology Platform Lungscape cohort. Mod Pathol. 2020;33(5):792-801.
37. Crapanzano JP, Heymann JJ, Monaco S, Nassar A, Saqi A. The state of cell block variation and satisfaction in the era of molecular diagnostics and personalized medicine. Cytojournal. 2014;11:7.
38. Nambirajan A, Jain D. Cell blocks in cytopathology: An update. Cytopathology. 2018;29(6):505-524.
39. Krogerus L, Kholova I. Cell Block in Cytological Diagnostics: Review of Preparatory Techniques. Acta Cytol. 2018;62(4):237-243.
40. Sauter JL, Grogg KL, Vrana JA, Law ME, Halvorson JL, Henry MR. Young investigator challenge: Validation and optimization of immunohistochemistry protocols for use on cellient cell block specimens. Cancer Cytopathol. 2016;124(2):89-100.
41. Gruchy JR, Barnes PJ, Dakin Hache KA. CytoLyt(R) fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. Appl Immunohistochem Mol Morphol. 2015;23(4):297-302.
42. Kruger AM, Stevens MW, Kerley KJ, Carter CD. Comparison of the Cellient() automated cell block system and agar cell block method. Cytopathology. 2014;25(6):381-388.
43. van Dooijeweert C, Deckers IAG, Baas IO, van der Wall E, van Diest PJ. Hormone- and HER2-receptor assessment in 33,046 breast cancer patients: a nationwide comparison of positivity rates between pathology laboratories in the Netherlands. Breast Cancer Res Treat. 2019;175(2):487-497.
44. Paik S, Bryant J, Tan-Chiu E, et al. Real-world performance of HER2 testing--National Surgical Adjuvant Breast and Bowel Project experience. J Natl Cancer Inst. 2002;94(11):852-854.



2

COMPARABILITY OF PD-L1 IMMUNOHISTOCHEMISTRY ASSAYS FOR NON-SMALL CELL LUNG CANCER: A SYSTEMATIC REVIEW

Koomen BM
Badrising SK
van den Heuvel MM
Willems SM

Histopathology. 2020;76(6):793-802

ABSTRACT

Programmed death-ligand 1 (PD-L1) immunohistochemistry is used to determine which patients with advanced non-small-cell lung cancer (NSCLC) respond best to treatment with PD-L1 inhibitors. For each inhibitor, a unique immunohistochemical assay was developed. This systematic review gives an up-to-date insight into the comparability of standardized immunohistochemical assays and laboratory-developed tests (LDTs), focusing specifically on tumor cell (TC) staining and scoring. A systematic search was performed identifying publications that assessed interassay, interobserver and/or interlaboratory concordance of PD-L1 assays and LDTs in tissue of NSCLC patients. Of 4,294 publications identified through the systematic search, 27 fulfilled the inclusion criteria and were of sufficient methodological quality. Studies assessing interassay concordance found high agreement between assays 22C3, 28-8 and SP263 and properly validated LDTs, and lower concordance for comparisons involving SP142. A decrease in concordance, however, is seen with use of cutoffs, which hampers interchangeability of PD-L1 immunohistochemistry assays and LDTs. Studies assessing interobserver concordance found high agreement for all assays and LDTs, but lower agreement with use of a 1% cutoff. This may be problematic in clinical practice, as discordance between pathologists at this cutoff may result in some patients being denied valuable treatment options. Finally, five studies assessed interlaboratory concordance and found moderate to high agreement levels for various assays and LDTs. However, to assess the actual existence of interlaboratory variation in PD-L1 testing and PD-L1 positivity in clinical practice, studies using real-world clinical pathology data are needed.

INTRODUCTION

Since the approval of the first immune check-point inhibitor in 2011[1,2,3], immunotherapy has become an important part of treatment for several forms of cancer. In patients with advanced non-small cell lung cancer (NSCLC), treatment with programmed death-1 (PD-1) or programmed death-ligand 1 (PD-L1) inhibitors has become part of standard care. These patients may be treated with nivolumab or pembrolizumab, both anti-PD-1 checkpoint inhibitors, or with an anti-PD-L1 checkpoint inhibitor, i.e. atezolizumab or durvalumab[4,5,6,7,8,9]. Some of these drugs may only be prescribed to patients who show PD-L1 expression in at least 1% or 50% of tumor cells, measured with immunohistochemistry (IHC)[10,11,12]. Immunohistochemical PD-L1 testing thus aids clinicians in treatment decision-making.

For each immune checkpoint inhibitor, however, a separate PD-L1 IHC assay has been developed. The PD-L1 IHC 22C3 PharmDx assay was used in clinical trials assessing efficacy of pembrolizumab, and is therefore Food and Drug Administration (FDA)-approved and Conformité Européenne (CE)-marked as a companion diagnostic for prescription of this drug[8,13,14]. In a similar fashion, the PD-L1 IHC 28-8 PharmDx assay was FDA-approved and CE-marked as a complementary diagnostic for nivolumab[15,16], while the PD-L1 IHC SP142 assay became a complementary diagnostic for atezolizumab[17,18]. Finally, the PD-L1 IHC SP263 assay was developed for durvalumab, but it has also received CE marking for identification of patients eligible for treatment with pembrolizumab and of patients most likely to benefit from treatment with nivolumab[19,20].

Using all these different assays to test for PD-L1 expression in one pathology laboratory is not feasible. Not only would it be expensive and time-consuming to run so many different tests for each patient, most laboratories will not have both staining platforms (i.e. Dako and Ventana/Roche) needed for these tests at their disposal. Furthermore, the number of tests that can be performed is restricted due to limited tissue availability[21]. It is thus important to assess whether results from different assays are interchangeable. In addition, it should be assessed if laboratory-developed tests (LDTs) can be used instead of the standardized PD-L1 assays. In recent years, a multitude of studies examining these issues has been published, such as the Blueprint PD-L1 IHC Assay Comparison Project[22] or the harmonization studies by Ratcliffe et al.[23], Rimm et al.[24] and Scheel et al.[25] Others, such as Büttner et al.[19], have reviewed the analytical performance of PD-L1 IHC assays previously. Considering the abundance of studies that have been published on the subject, however, there is need for a systematic, comprehensive and up-to-date overview of the literature, which not only focuses on interassay and interobserver concordance, but also includes a review of interlaboratory concordance. Hence, the aim of this study was to systematically review all studies that assessed interassay, interobserver and/or interlaboratory concordance of PD-L1 IHC assays and LDTs, and in so doing provide an updated insight

into the comparability of these standardized assays and LDTs.

MATERIALS AND METHODS

Search Strategy

A systematic search of PubMed, Embase and Cochrane Library was performed, using the search terms 'lung cancer' and 'PD-L1' with all relevant synonyms (see Table S1). Only these two terms were used to ensure that no relevant articles would be missed. Adding another term, such as 'immunohistochemistry', might have made the search more specific, but would also have increased the risk of eliminating relevant titles. After removal of duplicates, titles and abstracts were screened by two researchers independently (B.K. and S.B.) based on predefined inclusion and exclusion criteria (see Table S2). Remaining articles were read in full, and a further selection was made based on the relevance of these full texts. Discrepancies between the two researchers were discussed and resolved by consensus.

Inclusion Criteria

Studies were included if they evaluated interassay, interobserver and/or interlaboratory concordance of at least two PD-L1 IHC assays and/or LDTs used on tissue from NSCLC patients in clinical practice. Studies examining interobserver and/or interlaboratory concordance in only one assay were also included. In order for studies to qualify, determination of PD-L1 expression had to be performed on histological tissue from NSCLC patients and appropriate scoring methods had to be used (i.e. assessment of membranous staining of tumor cells by at least one pathologist). Since PD-L1 IHC was validated in histological specimens, studies examining cytological specimens only were excluded. Studies that did not perform adequate statistical analysis to compare assays (i.e. overall percentage of agreement should at least be given) were also excluded. Only articles written in English and containing original published data were eligible for inclusion.

Quality Assessment

Methodological quality of all articles remaining after full text reading were appraised by using a revised form of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool for assessing risk of bias[26]. Originally, this tool consists of four domains, i.e. patient selection, index test, reference standard and flow and timing. As individual PD-L1 IHC assays were not compared to a reference standard in the included studies, but rather to each other, the reference standard domain was excluded from the QUADAS-2 tool for this review. Instead, another domain was added based on the Quality in Prognosis Studies (QUIPS) tool, i.e. statistical analysis and reporting[27]. Risk of bias was scored as low, moderate or high for each domain of the revised QUADAS-2 tool and points were

awarded accordingly (1 point for low risk of bias, 0.5 points for moderate risk of bias and 0 points for high risk of bias). Based on the sum of the scores given to each individual domain, overall scores of low, moderate or high risk of bias were awarded to studies using the following scoring system: low risk of bias for studies with ≥3.5 points, moderate risk of bias for studies with ≥2.5 and <3.5 points and high risk of bias for studies with <2.5 points. Appraisal of methodological quality was performed independently by two researchers (B.K. and S.B.) and differences were resolved through discussion. Studies with high risk of bias were excluded from data extraction and further analysis.

Data Extraction And Synthesis Of Results

The following data were extracted from each study included after appraisal of methodological quality: first author's name, year of publication, sample size, type of cancer of included patients, type of material used for PD-L1 testing, type of standardized assay and/or LDT used for PD-L1 testing, scoring method, cutoff values, number of observers scoring PD-L1, type of statistical analysis and results from comparison between assays, observers and/or laboratories. This review focuses on concordance of tumor cell (TC) staining and scoring, as treatment decisions for NSCLC patients are based on scoring of PD-L1 expression on TCs in clinical practice. However, as scoring of PD-L1 expression on immune cells (IC) could become relevant to clinical practice in the future, we also extracted data on concordance of IC staining and scoring and included this as Data S1 and Table S8. Due to heterogeneity between included studies, such as differences in antibodies tested, number of pathologists scoring and statistical methods applied, results could not be quantitatively pooled and a meta-analysis could not be performed.

RESULTS

Systematic Search And Study Selection

The search in PubMed, Embase and Cochrane Library yielded 4,294 unique hits after removal of duplicates (see Figure 1). Fifty-nine records remained after screening of titles and abstracts. Of these, one full text was unavailable. Therefore, 58 full text articles were evaluated in detail, of which 41 articles met the inclusion criteria. All selected articles studied interassay, interobserver and/or interlaboratory concordance of at least one PD-L1 IHC assay, using material from NSCLC patients. Most studies included multiple subtypes of NSCLC, with adenocarcinoma and squamous cell carcinoma being studied most frequently. Some studies also included patients with other types of lung cancer, such as small cell lung cancer (SCLC)[28, 29] and mesothelioma[30]. Sample sizes ranged from 15 to 713 tissue specimens. All studies used statistical analysis to measure concordance. The statistical methods used, however, varied. The kappa statistic (κ) was used most, but some studies used intraclass correlation coefficient (ICC), Pearson's/Spearman's correlation or

calculation of percentage agreement.

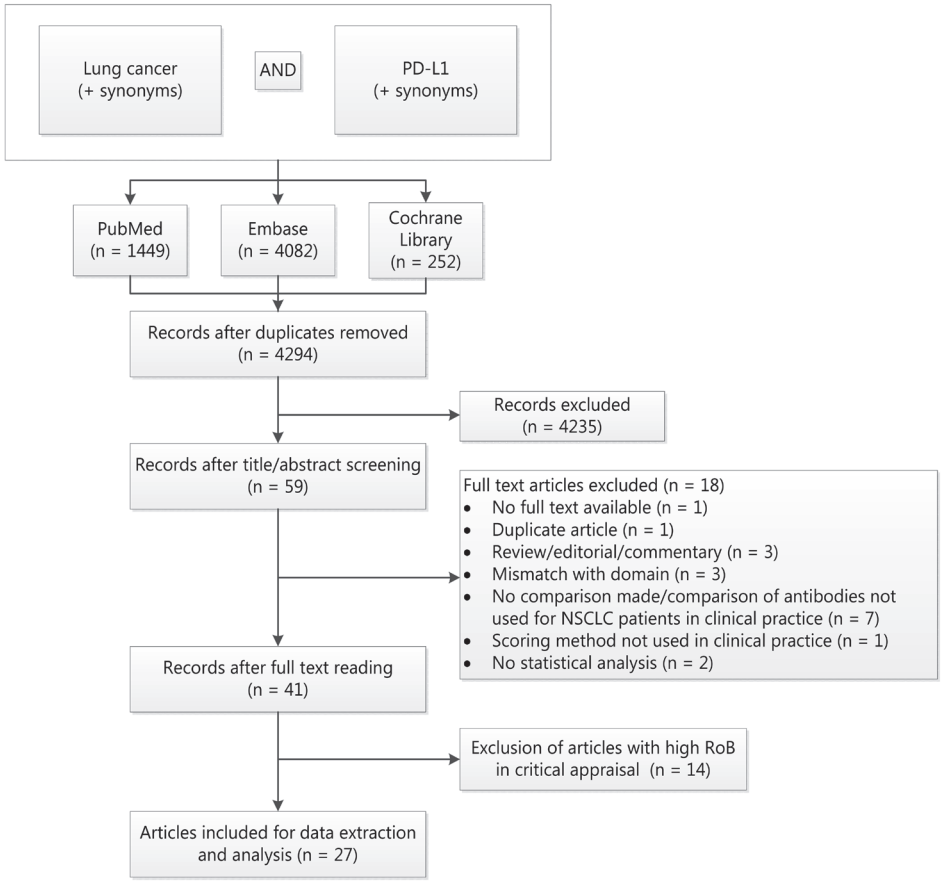


Figure 1. Flowchart of study selection process (date of search: 27 June 2018). PD-L1, programmed death-ligand 1; RoB, risk of bias.

Quality Assessment

The 41 articles selected through full text reading were critically appraised on methodological quality. Based on scoring with the revised QUADAS-2 tool, studies ranged from low to high risk of bias (see Table S3). Studies with high risk of bias were often unclear concerning their method of patient selection and reasons for patient exclusion, about blinding of pathologists for each other's results and for the specific antibody used, about the use of staining platform and staining protocol, or about the scoring method used. Also, some studies did not provide sufficient information on the use of statistical methods or did not present all data, prohibiting assessment of adequacy of analytical strategy. Five studies

were judged as having low risk of bias[29, 31, 32, 33, 34], 22 studies as having moderate risk of bias[22, 23, 24, 25, 28, 30, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50] and 14 studies as having high risk of bias[51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64]. The 14 studies with high risk of bias were excluded, which left 27 articles for data extraction and further analysis. An overview of study characteristics of all included studies can be found in Table S4.

Interassay Concordance

Of the 27 included articles, 22 reported on interassay concordance of TC staining between PD-L1 IHC assays. A summary of results from all 22 studies can be found in Table 1, while a more detailed presentation of results from each study can be found in Table S5. Many studies compared the standardized assays 22C3, 28-8, SP263 and SP142. Overall, moderate to strong concordance was seen between 22C3, 28-8 and SP263[22, 23, 28, 31, 32, 35, 38, 43, 48] and lower concordance between SP142 and the other assays[22, 24, 28, 31, 32, 35, 38, 39, 48, 50]. Concordance was often highest between assays 22C3 and 28-8[22, 30, 32, 35, 48], such as demonstrated by Brunnström et al.[32], who found a weighted κ value of 0.891 (0.82–0.96) for comparison of these two assays. Two studies by Scheel et al.[25, 46] showed somewhat lower concordance values between 22C3, 28-8 and SP263 than the other studies, but these results may have been affected by interobserver variation and by the low sample size in both studies. Several studies described a higher proportion of stained TCs with use of antibody SP263 when compared to antibody 22C3 and/or 28-8[25, 28, 35, 44, 46, 48]. According to Munari et al.[44] this difference in staining led to a significantly lower proportion of positive cases with assay 22C3 compared to assay SP263 for both the 1% and 50% cutoffs. Similarly, other studies also assessed concordance with deployment of clinically relevant cutoffs. Some of these studies showed diminished concordance rates when cutoffs were used. Hendry et al.[28] showed only moderate agreement between 22C3, 28-8 and SP263 when cutoffs were used (Cohen's κ range = 0.433–0.631), while good agreement was found for PD-L1 expression on a continuous scale (ICC range = 0.726–0.812). In the Blueprint Phase 1 study, agreement with the reference assay ranged from 86.8% to 94.7% for comparisons between antibody 22C3, 28-8 and SP263 when different cutoffs were used, meaning that in some cases almost 15% of patients in the study would not have been assigned a treatment if an alternative to the reference assay had been used[22]. Other studies showed lower agreement for the 1% than for the 50% cutoff[35, 38, 43]. Two studies showed good concordance between assays for any cutoff used[23, 30], but these studies only calculated percentage agreement, which may overestimate true agreement[65, 66].

Table 1. Summary of results from studies assessing interassay concordance of TC staining

Type of test	Comparison	Interassay concordance
Standardized assays	22C3, 28-8 and SP263	• Moderate to high concordance for all comparisons[22, 23, 28, 31, 32, 35, 38, 43, 48]; • Highest concordance between 22C3 and 28-8[22, 30, 32, 35, 48]; • Lower concordance rates with use of cutoffs[22, 28, 44], especially using the 1% cutoff[35, 38, 43].
	SP142 vs. all other assays	• Lower concordance levels compared to comparisons between all other assays[22, 24, 28, 31, 32, 35, 38, 39, 48, 50].
LDTs	Various LDTs vs. standardized assays	• High concordance for some LDTs, only if appropriate protocol used[31, 32, 47].
	22C3 LDT vs. 22C3 standardized assay	• High correlation[28, 40, 41, 44]; • In some studies higher correlation than between two different standardized assays[28, 44].
	E1L3N vs. all standardized assays	• High concordance between E1L3N and 22C3, 28-8 and SP263[24, 37, 47]; • Lower concordance between E1L3N and SP142[24, 42].

Abbreviations: LDT=laboratory-developed test; TC=tumor cell.

Comparisons of TC staining were also made between the 22C3, 28-8, SP263 and SP142 antibodies being used with their standard protocols and being used in LDTs. A study by Adam et al.[31] demonstrated that 14 of 27 LDTs were concordant (defined as weighted κ value ≥0.75) with one of the pre-specified reference assays. The lowest κ value was seen for the SP142 LDT compared to the SP263 reference assay (weighted κ = 0.38). Two studies by Ilie et al.[40, 41] showed high correlation between two different 22C3 LDTs and the 22C3 standardized assay. Another study also showed excellent agreement between the 22C3 standardized assay and LDT, with an ICC of 0.921 and Cohen's κ of 0.897 for the 50% cut-off[28]. In this study, discrepancies were actually much greater between two different antibodies used on the same platform (22C3 and 28-8) than between the same antibody (22C3) used on different platforms. A similar finding was reported by Munari et al.[44] Other studies also compared one or more of the aforementioned standardized assays with antibody E1L3N, which is used as an LDT by some laboratories in clinical practice. Good correlation was seen between E1L3N and assays SP263, 28-8 and 22C3[24, 37, 47], while comparison with SP142 again showed lower concordance values[24, 42]. One study[36] showed higher sensitivity in staining of PD-L1 using 28-8 compared to E1L3N. Finally, in a study by Soo et al.[47], which used the SP142 antibody as an LDT, changes in the SP142 protocol led to a higher intensity of staining compared to the original protocol, demonstrating how the IHC protocol can influence the apparent level of PD-L1 expression.

Interobserver Concordance

Sixteen of the 27 included studies examined interobserver concordance (see Table 2; Table S6). All these studies assessed concordance between pathologists scoring TC staining and most found moderate to almost perfect agreement for all assays[23, 24, 29, 32, 33, 34, 35, 37

[39, 40, 45, 48, 49]. Only Scheel et al.[25] found somewhat lower concordance values for E1L3N and SP142 LDTs and 22C3, 28-8, SP263 and SP142 standardized assays when a scoring system applying five cutoffs was used (Light's κ range = 0.47–0.50). However, the sample size in this study was very small (n = 15), and classifying the cases by the dichotomous cutoff criteria included in the scoring system resulted in higher concordance levels for all antibodies (Light's κ range = 0.59–0.80). Other studies also assessed interobserver concordance for multiple cutoffs, and many found concordance levels to be lower for the 1% cutoff compared to the 50% cutoff[23, 24, 29, 32, 35, 43, 48] and the 5%, 10% or 25% cutoff[23, 32, 48]. The Blueprint Phase 2 study also assessed the 80% cutoff and found interpathologist agreement to be slightly diminished for this cutoff compared to the 5%, 10%, 25% and 50% cutoffs[48]. A study by Cooper et al.[33] actually reported lower concordance levels for the 50% cutoff than for the 1% cutoff for assay 22C3 (overall percentage agreement (OPA) 81.9% and κ = 0.58 versus OPA 84.2% and κ = 0.69, respectively). However, this study reported prevalence bias to have influenced the κ magnitude for the 50% cutoff. These results therefore have to be interpreted with caution.

Table 2. Summary of results from studies assessing interobserver concordance of TC scoring

Type of test	Overall	Use of cutoffs
Standardized assays	• Good concordance for all standardized assays[23, 24, 29, 32-35, 37, 39, 40, 43-45, 48, 49]; • One study showing only moderate agreement[25].	• Lower concordance levels for 1% cutoff compared to 50% cutoff[23, 24, 29, 32, 35, 43, 48]; • Lower concordance levels for 1% cutoff compared to 5%, 10% and 25% cutoffs[23, 32, 48]; • Lower concordance levels for 80% cutoff compared to other cutoffs[48].
LDTs	Good concordance for various LDTs[24, 29, 32, 37, 40, 44, 45].	Lower concordance levels for 1% cutoff compared to other cutoffs[24, 29, 32].

Abbreviations: LDT=laboratory-developed test; TC=tumor cell.

Interlaboratory Concordance

Interlaboratory concordance of TC staining was assessed by five of the 27 included studies (see Table 3; Table S7). Two of these[34, 49] assessed only one antibody (22C3 and SP142, respectively). Both studies found high interlaboratory agreement. Adam et al.[31], who assessed interlaboratory concordance for 22C3, 28-8 and SP263, found very high agreement between participating centers for each of these assays. Marchetti et al.[43] found similar results for the assays 22C3 and SP263. Scheel et al.[46] assessed interlaboratory concordance for standardized assays 22C3, 28-8, SP263 and SP142 and for 22C3, 28-8, SP263 and E1L3N used in LDTs, performed in 10 different sites. Concordance values ranged from Light's κ = 0.63–0.69 for the standardized assays when five cutoffs were used. κ was 0.49 for all the LDTs grouped together. When only a 1% and 50% cutoff were used, concordance values improved to κ = 0.73–0.89 for the standardized assays and κ = 0.5 for the LDTs.

Table 3. Summary of results from studies assessing interlaboratory concordance of TC scoring

Type of test	Interlaboratory concordance
Standardized assays	• 22C3: substantial to near perfect concordance[31, 34, 43, 46]; • 28-8: substantial to near perfect concordance[31, 46]; • SP263: substantial to near perfect concordance[31, 43, 46]; • SP142: high inter-site percent agreement[49].
LDTs	Only moderate concordance levels compared to standardized assays[46].

Abbreviations: LDT=laboratory-developed test; TC=tumor cell.

Concordance Of Immune Cell Staining And Scoring

A short analysis of concordance of IC staining and scoring can be found as Data S1 and Table S8.

DISCUSSION

Ever since the approval of PD-1 and PD-L1 inhibitors as treatment options for patients with advanced NSCLC, various studies have been published assessing the comparability of different PD-L1 IHC assays. In this systematic review, interassay, interobserver and interlaboratory concordance of these PD-L1 IHC assays and LDTs were investigated by reviewing all currently available literature.

Overall, interassay agreement of TC staining is high between standardized assays 22C3, 28-8 and SP263, while assay SP142 frequently shows lower staining of TCs. Agreement between LDTs and their reference assay may also be high, depending on the protocol that is used, with some studies even showing greater agreement between LDTs and their reference assays than between different standardized assays[28, 44]. These data seem to suggest that the assays 22C3, 28-8 and SP263 and properly validated LDTs could be used interchangeably on histological specimens of NSCLC patients. However, some studies have shown lower concordance levels with the use of clinically relevant cutoffs[22, 28, 44]. The 1% cutoff especially may lead to higher disagreement compared to the 50% cutoff[35, 38, 43], although this could perhaps be attributed to lower interrater agreement levels at this cutoff[43]. Based on the lower concordance levels found when using various cutoffs, it would be too premature to draw the conclusion that assays and LDTs can be used interchangeably without any consequences. Notably, in a recent meta-analysis of diagnostic accuracy of PD-L1 IHC assays, Torlakovic et al.[67] demonstrated that none of the standardized PD-L1 assays could be deemed as interchangeable, when interchangeability is defined as achieving ≥90% sensitivity and specificity for both the 1 and 50% cutoffs. Because discordance may exist between assays at clinically relevant cutoffs, simply interchanging one assay with another may potentially lead to patients being wrongfully denied valuable treatment options in clinical practice.

Assessment of interobserver concordance of TC scoring showed that agreement between pathologists is moderate to high for all assays and LDTs. Markedly, agreement is often found to be lowest for the 1% cutoff compared to other cutoffs. This is problematic, especially now that the European Medicines Agency (EMA) has only approved durvalumab as consolidation treatment in stage III NSCLC patients whose tumors show PD-L1 expression of ≥1%[12]. One could question if the use of this cutoff provides results that are reliable enough to aid clinicians in making treatment decisions. Agreement is likely to be higher between more experienced pathologists[43], yet still leaves room for improvement. One study assessing training of already experienced pathologists showed no or only little improvement of interobserver agreement[33]. This study, however, employed a 1-h training session consisting of a presentation only. Alternative training initiatives, preferably including a more practical element during which trainees have to perform PD-L1 scoring on multiple specimens, might prove to be more effective. A recent study assessing interpathologist concordance of PD-L1 scoring using real-world data showed that training for PD-L1 scoring and experience in routine pathology practice correlated with higher concordance[68]. The effect of training on interobserver concordance should thus be studied more extensively. Other solutions also deserve more attention, especially the use of digital image analysis for PD-L1 scoring, as this has been shown to reduce interobserver variability[69].

Finally, we assessed concordance of PD-L1 IHC assays and LDTs between laboratories. Only a limited number of studies assessed this type of concordance, especially compared to the large number of studies assessing interassay and/or interobserver concordance. Most of the studies assessing interlaboratory concordance found high agreement for all standardized assays, while one study found lower agreement for LDTs[46]. However, not all these studies used the right study protocol and the right outcome measure to properly assess interlaboratory concordance. Two studies[34, 49] used percentage of agreement as outcome measure, which does not account for random agreement and may thus overestimate true agreement[65, 66]. Two other studies[43, 46] used study designs that did not allow for separate analysis of interobserver and interlaboratory concordance. Moreover, none of the study designs allowed for assessment of the influence of pre-analytical variables on PD-L1 immunostaining, while in clinical practice pre-analytical processing of samples may actually differ considerably between laboratories and may influence IHC staining results[70, 71, 72]. Therefore, studies assessing interlaboratory variation in PD-L1 expression are needed, using real-world data and thereby taking into account these possible differences in pre-analytical variables in clinical practice.

This systematic review has some limitations. Most importantly, there is significant heterogeneity between the studies included, especially in the choice of antibodies tested and the statistical methods used to analyze concordance. This prohibits pooling of data and complicates proper comparison of results between studies. Most studies, however,

used similar samples for PD-L1 testing, i.e. formalin-fixed paraffin-embedded material from tumor resections or biopsies from NSCLC patients. This supports comparability between studies. Conversely, this also provides a disadvantage: it only allows for comparison of PD-L1 IHC assays and LDTs in histology, while in clinical practice PD-L1 immunostaining is frequently performed on cytological specimens. Comparison of PD-L1 IHC assays and LDTs in cytological NSCLC specimens falls beyond the scope of this review, but would be worth evaluation in a separate study. Finally, many of the included studies were not of high methodological quality, with only five studies being judged as having low risk of bias. Excluding the studies with the highest level of risk of bias, however, has improved the overall quality of this review.

To conclude, this systematic review has shown that interassay concordance of TC staining is generally high between the standardized assays 22C3, 28-8 and SP263 and properly developed and validated LDTs. Nevertheless, the use of clinically relevant cutoffs may lead to lower levels of interassay concordance, indicating that these assays and LDTs cannot simply be interchanged. Interobserver agreement, moreover, is generally high for all assays and LDTs, but decreases with use of the 1% cutoff. Lastly, interlaboratory concordance seems to be high for standardized assays and moderate for LDTs, but has not been studied sufficiently to draw definitive conclusions. Studies using real-world clinical pathology data are necessary to assess whether use of different PD-L1 IHC assays and LDTs, scoring by different pathologists and use of different pre-analytical variables actually lead to differences in PD-L1 positivity between laboratories in clinical practice.

REFERENCES

1. Li Y, Li F, Jiang F, et al. A mini-review for cancer immunotherapy: molecular understanding of PD-1/PD-L1 pathway & translational blockade of immune checkpoints. Int. J. Mol. Sci. 2016;17:pii:E1151.
2. US Food and Drug Administration . YERVOY (ipilimumab) [highlights of prescribing information]. 2018. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/125377s094lbl.pdf (Accessed 30 October 2019).
3. European Medicines Agency (EMA). Yervoy (ipilimumab). 2019. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/yervoy (Accessed 30 October 2019).
4. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N. Engl. J. Med. 2015;373:123–135.
5. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N. Engl. J. Med. 2015;373:1627–1639.
6. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet 2017;389:255–265.
7. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in Stage III non-small-cell lung cancer. N. Engl. J. Med. 2017;377:1919–1929.
8. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N. Engl. J. Med. 2016;375:1823–1833.
9. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016;387:1540–1550.
10. Planchard D, Popat S, Kerr K, et al. Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. 2018;29(Suppl 4):iv192–iv237.
11. US Food and Drug Administration. KEYTRUDA (pembrolizumab) [prescribing information]. 2018. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/125514s034lbl.pdf (Accessed 13 February 2019).
12. European Medicines Agency (EMA). Imfinzi (durvalumab). 2018. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/imfinzi (Accessed 9 April 2019).
13. US Food and Drug Administration. Dako PD-L1 IHC 22C3 pharmDx. 2015. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150013c.pdf (Accessed 13 February 2019).
14. Dako Agilent Pathology Solutions. PD-L1 IHC 22C3 pharmDx is CE-IVD-marked for in vitro diagnostic use. Available from: https://www.agilent.com/cs/library/usermanuals/public/29171_22C3-ihc-pharmdx-interpretation-manual-eu.pdf (Accessed 30 October 2019).
15. US Food and Drug Administration. Dako PD-L1 IHC 28–8 pharmDx. 2015. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150025c.pdf (accessed 13 February 2019).
16. Dako Agilent Pathology Solutions. PD-L1 IHC 28–8 pharmDx product information. Available from: https://www.agilent.com/cs/library/brochures/29125_pd-l1-ihc-28-8-pharmdx-brochure_row.pdf (Accessed 30 October 2019).

17. Ventana Medical Systems Inc. VENTANA PD-L1 (SP142) assay. 2019. Available from: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp142-assay2.html (Accessed 13 February 2019).
18. US Food and Drug Administration. VENTANA PD-L1 (SP142) assay. 2016. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160002c.pdf (Accessed 30 October 2019).
19. Buttner R, Gosney JR, Skov BG, et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer. J. Clin. Oncol. 2017;35:3867–3876.
20. Ventana Medical Systems, Inc. VENTANA PD-L1 (SP263) assay. 2019. Available from: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp263-assay2.html (Accessed 13 February 2019).
21. Kerr KM, Hirsch FR. Programmed death ligand-1 immunohistochemistry: friend or foe? Arch. Pathol. Lab. Med. 2016;140:326–331.
22. Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 immunohistochemistry assays for lung cancer: results from Phase 1 of the blueprint PD-L1 IHC assay comparison project. J. Thorac. Oncol. 2017;12:208–222.
23. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between programmed cell death ligand-1 diagnostic assays across multiple protein expression cutoffs in non-small-cell lung cancer. Clin. Cancer Res. 2017;23:3585–3591.
24. Rimm DL, Han G, Taube JM, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small-cell lung cancer. JAMA Oncol. 2017;3:1051–1058.
25. Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. Mod. Pathol. 2016;29:1165–1172.
26. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 2011;155:529–536.
27. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann. Intern. Med. 2013;158:280–286.
28. Hendry S, Byrne DJ, Wright GM, et al. Comparison of four PD-L1 immunohistochemical assays in lung cancer. J. Thorac. Oncol. 2018;13:367–376.
29. Russell-Goldman E, Kravets S, Dahlberg SE, Sholl LM, Vivero M. Cytologic–histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. Cancer Cytopathol. 2018;126:253–263.
30. Skov BG, Skov T. Paired comparison of PD-L1 expression on cytologic and histologic specimens from malignancies in the lung assessed with PD-L1 IHC 28–8pharmDx and PD-L1 IHC 22C3pharmDx. Appl. Immunohistochem. Mol. Morphol. 2017;25:453–459.
31. Adam J, Le Stang N, Rouquette I, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. Ann. Oncol. 2018;29:953–958.
32. Brunnström H, Johansson A, Westbom-Fremer S, et al. PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability. Mod. Pathol. 2017;30:1411–1421.
33. Cooper WA, Russell PA, Cherian M, et al. Intra- and interobserver reproducibility assessment of PD-L1 biomarker in non-small-cell lung cancer. Clin. Cancer Res. 2017;23:4569–4577.
34. Roach C, Zhang N, Corigliano E, et al. Development of a companion diagnostic PD-L1 immunohistochemistry assay for pembrolizumab therapy in non-small-cell lung cancer. Appl. Immunohistochem. Mol. Morphol. 2016;24:392–397.

35. Chan AWH, Tong JHM, Kwan JSH, et al. Assessment of programmed cell death ligand-1 expression by 4 diagnostic assays and its clinicopathological correlation in a large cohort of surgical resected non-small-cell lung carcinoma. Mod. Pathol. 2018;31:1381–1390.
36. Cogswell J, Inzunza HD, Wu Q, et al. An analytical comparison of Dako 28–8 PharmDx assay and an E1L3N laboratory-developed test in the immunohistochemical detection of programmed death-ligand 1. Mol. Diagn. Ther. 2017;21:85–93.
37. Conde E, Caminoa A, Dominguez C, et al. Aligning digital CD8(+) scoring and targeted next-generation sequencing with programmed death ligand 1 expression: a pragmatic approach in early-stage squamous cell lung carcinoma. Histopathology 2018;72:270–284.
38. Fujimoto D, Sato Y, Uehara K, et al. Predictive performance of four programmed cell death ligand 1 assay systems on nivolumab response in previously treated patients with non-small-cell lung cancer. J. Thorac. Oncol. 2018;13:377–386.
39. Ilie M, Falk AT, Butori C, et al. PD-L1 expression in basaloid squamous cell lung carcinoma: relationship to PD-1(+) and CD8(+) tumor-infiltrating T cells and outcome. Mod. Pathol. 2016;29:1552–1564.
40. Ilie M, Khambata-Ford S, Copie-Bergman C, et al. Use of the 22C3 anti-PD-L1 antibody to determine PD-L1 expression in multiple automated immunohistochemistry platforms. PLoS One 2017;12:e0183023.
41. Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death ligand 1 antibody to determine programmed death ligand 1 expression in cytology samples obtained from non-small-cell lung cancer patients. Cancer Cytopathol. 2018;126:264–274.
42. Keller MD, Neppl C, Irmak Y, et al. Adverse prognostic value of PD-L1 expression in primary resected pulmonary squamous cell carcinomas and paired mediastinal lymph node metastases. Mod. Pathol. 2018;31:101–110.
43. Marchetti A, Barberis M, Franco R, et al. Multicenter comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) assays to test PD-L1 expression for NSCLC patients to be treated with immune checkpoint inhibitors. J. Thorac. Oncol. 2017;12:1654–1663.
44. Munari E, Rossi G, Zamboni G, et al. PD-L1 assays 22C3 and SP263 are not interchangeable in non-small-cell lung cancer when considering clinically relevant cutoffs: an interclone evaluation by differently trained pathologists. Am. J. Surg. Pathol. 2018;42:1384–1389.
45. Rehman JA, Han G, Carvajal-Hausdorf DE, et al. Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small-cell lung cancer. Mod. Pathol. 2017;30:340–349.
46. Scheel AH, Baenfer G, Baretton G, et al. Interlaboratory concordance of PD-L1 immunohistochemistry for non-small-cell lung cancer. Histopathology 2018;72:449–459.
47. Soo RA, Lim JSY, Asuncion BR, et al. Determinants of variability of five programmed death ligand-1 immunohistochemistry assays in non-small-cell lung cancer samples. Oncotarget 2018;9:6841–6851.
48. Tsao MS, Kerr KM, Kockx M, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of Blueprint Phase 2 project. J. Thorac. Oncol. 2018;13:1302–1311.
49. Vennapusa B, Baker B, Kowanetz M, et al. Development of a PD-L1 complementary diagnostic immunohistochemistry assay (SP142) for atezolizumab. Appl. Immunohistochem. Mol. Morphol. 2019;27:92–100.
50. Xu H, Lin G, Huang C, et al. Assessment of concordance between 22C3 and SP142 immunohistochemistry

assays regarding PD-L1 expression in non-small-cell lung cancer. Sci Rep. 2017;7:16956.

51. Erber R, Stohr R, Herlein S, et al. Comparison of PD-L1 mRNA expression measured with the CheckPoint Typer(R) assay with PD-L1 protein expression assessed with immunohistochemistry in non-small-cell lung cancer. Anticancer Res. 2017;37:6771–6778.
52. Kim H, Kwon HJ, Park SY, Park E, Chung JH. PD-L1 immunohistochemical assays for assessment of therapeutic strategies involving immune checkpoint inhibitors in non-small-cell lung cancer: a comparative study. Oncotarget 2017;8:98524–98532.
53. Krawczyk P, Jarosz B, Kucharczyk T, et al. Immunohistochemical assays incorporating SP142 and 22C3 monoclonal antibodies for detection of PD-L1 expression in NSCLC patients with known status of EGFR and ALK genes. Oncotarget 2017;8:64283–64293.
54. McLaughlin J, Han G, Schalper KA, et al. Quantitative assessment of the heterogeneity of PD-L1 expression in non-small-cell lung cancer. JAMA Oncol. 2016;2:46–54.
55. Neuman T, London M, Kania-Almog J, et al. A Harmonization study for the use of 22C3 PD-L1 immunohistochemical staining on Ventana's platform. J. Thorac. Oncol. 2016;11:1863–1868.
56. Pang C, Yin L, Zhou X, et al. Assessment of programmed cell death ligand-1 expression with multiple immunohistochemistry antibody clones in non-small-cell lung cancer. J. Thorac. Dis. 2018;10:816–824.
57. Parra ER, Villalobos P, Mino B, Rodriguez-Canales J. Comparison of different antibody clones for immunohistochemistry detection of programmed cell death ligand 1 (PD-L1) on non-small-cell lung carcinoma. Appl. Immunohistochem. Mol. Morphol. 2018;26:83–93.
58. Paulsen EE, Kilvaer TK, Khanehkenari MR, et al. Assessing PDL-1 and PD-1 in non-small-cell lung cancer: a novel immunoscore approach. Clin. Lung Cancer 2017;18:220–233.e8.
59. Phillips T, Simmons P, Inzunza HD, et al. Development of an automated PD-L1 immunohistochemistry (IHC) assay for non-small-cell lung cancer. Appl. Immunohistochem. Mol. Morphol. 2015;23:541–549.
60. Rebelatto MC, Midha A, Mistry A, et al. Development of a programmed cell death ligand-1 immunohistochemical assay validated for analysis of non-small-cell lung cancer and head and neck squamous cell carcinoma. Diagn. Pathol. 2016;11:95.
61. Roge R, Vyberg M, Nielsen S. Accurate PD-L1 protocols for non-small-cell lung cancer can be developed for automated staining platforms with clone 22C3. Appl. Immunohistochem. Mol. Morphol. 2017;25:381–385.
62. Sheffield BS, Fulton R, Kalloger SE, et al. Investigation of PD-L1 biomarker testing methods for PD-1 axis inhibition in non-squamous non-small-cell lung cancer. J. Histochem. Cytochem. 2016;64:587–600.
63. Smith J, Robida MD, Acosta K, et al. Quantitative and qualitative characterization of two PD-L1 clones: SP263 and E1L3N. Diagn. Pathol. 2016;11:44.
64. Tseng JS, Yang TY, Wu CY, et al. Characteristics and predictive value of PD-L1 status in real-world non-small cell lung cancer patients. J. Immunother. 2018;41:292–299.
65. Mandrekar JN. Measures of interrater agreement. J. Thorac. Oncol. 2011;6:6–7.
66. McHugh ML. Interrater reliability: the kappa statistic. Biochem. Med. (Zagreb) 2012;22:276–282.
67. Torlakovic E, Lim HJ, Adam J, et al. 'Interchangeability' of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Mod. Pathol. 2019. 10.1038/s41379-019-0327-4.
68. Adam J, Hofman V, Mansuet-Lupo A, et al. P2.09-17 real-world concordance across pathologists for PD-L1 scoring in non-small-cell lung cancer: results from a large nationwide initiative. J. Thorac. Oncol. 2019;14:S775.

69. Koelzer VH, Gisler A, Hanhart JC, et al. Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. Histopathology 2018;73:397–406.

70. Khoury T, Sait S, Hwang H, et al. Delay to formalin fixation effect on breast biomarkers. Mod. Pathol. 2009;22:1457–1467.

71. Babic A, Loftin IR, Stanislaw S, et al. The impact of pre-analytical processing on staining quality for H&E, dual hapten, dual color in situ hybridization and fluorescent in situ hybridization assays. Methods 2010;52:287–300.

72. Wolff AC, Hammond ME, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J. Clin. Oncol. 2013;31:3997–4013.

SUPPORTING INFORMATION

Table S1. Search syntax in PubMed, Embase and Cochrane Library (date of search: June 27, 2018).

Database	Search syntax	Hits
PubMed	(((lung[Title/Abstract] OR lungs[Title/Abstract] OR pulmonary[Title/Abstract]) AND (cancer*[Title/Abstract] OR carcinom*[Title/Abstract] OR neoplasm*[Title/Abstract] OR tumour*[Title/Abstract] OR tumour*[Title/Abstract])) OR lung adenocarcinom*[Title/Abstract] OR pulmonary adenocarcinom*[Title/Abstract] OR NSCLC[Title/Abstract] OR carcinoma, non small cell lung[MeSH Terms]) AND (PD-L1[Title/Abstract] OR PDL1[Title/Abstract] OR "Programmed death-ligand 1"[Title/Abstract] OR "Programmed cell death-ligand 1"[Title/Abstract] OR B7-H1[Title/Abstract] OR B7H1[Title/Abstract] OR CD274[Title/Abstract] OR B7-H1 antigen[MeSH Terms])	1,449
Embase	(((lung:ti,ab OR lungs:ti,ab OR pulmonary:ti,ab) AND (cancer*:ti,ab OR carcinom*:ti,ab OR tumour*:ti,ab OR tumour*:ti,ab OR neoplasm*:ti,ab)) OR 'lung adenocarcinom*':ti,ab OR 'pulmonary adenocarcinom*':ti,ab OR nsclc:ti,ab OR 'non small cell lung cancer'/exp) AND ('pd l1':ti,ab OR pdl1:ti,ab OR 'programmed cell death-ligand 1':ti,ab OR 'programmed death-ligand 1':ti,ab OR 'b7 h1':ti,ab OR b7h1:ti,ab OR cd274:ti,ab OR 'programmed death 1 ligand 1'/exp) AND [embase]/lim	4,082
Cochrane Library	(((lung:ti,ab OR lungs:ti,ab OR pulmonary:ti,ab) AND (cancer*:ti,ab OR carcinom*:ti,ab OR tumour*:ti,ab OR tumour*:ti,ab OR neoplasm*:ti,ab)) OR "lung adenocarcinom*":ti,ab OR "pulmonary adenocarcinom*":ti,ab OR nsclc:ti,ab) AND (PD-L1:ti,ab OR PDL1:ti,ab OR "Programmed death-ligand 1":ti,ab OR "Programmed cell death-ligand 1":ti,ab OR B7-H1:ti,ab OR B7H1:ti,ab OR CD274:ti,ab)	252

Table S2. Inclusion and exclusion criteria.

Inclusion criteria	
Population	Patients with non-small cell lung cancer (any type).
Study design	• Comparison of at least two commercially available standardized assays and/or LDTs for detection of PD-L1 expression in NSCLC patients; • Studies examining only one assay may be included when interobserver concordance and/or interlaboratory concordance is assessed.
Outcome	• Interassay and/or interobserver and/or interlaboratory concordance of PD-L1 staining/scoring are assessed; • Proper statistical analysis is performed (overall percentage of agreement should at least be given).
Exclusion criteria	
	• Comparison of assays/LDTs only used for evaluating PD-L1 expression in types of cancer other than NSCLC; • Use of cytological material only; • Use of scoring method not employed in clinical practice; • Language other than English; • Duplicate articles containing all/some of original publicized data; • Reviews, conference abstracts, case reports, editorials, book chapters, presentations.

Table S3. Quality assessment of included studies.

First author	Year	Patient selection	Test(s)	Flow and timing	Statistical analysis and reporting	Risk of Bias (RoB)
Adam[31]	2018	◐	●	●	●	Low
Brunnström[32]	2017	●	●	◑	●	Low
Chan[35]	2018	●	◑	◑	●	Moderate
Cogswell[36]	2017	○	◑	●	●	Moderate
Conde[37]	2018	●	◑	◑	●	Moderate
Cooper[33]	2017	●	●	●	◑	Low
Erber[51]	2017	◐	◐	◐	◐	High
Fujimoto[38]	2017	◐	●	◐	●	Moderate
Hendry[28]	2017	◐	◑	◑	●	Moderate
Hirsch[22]	2016	◐	●	◐	●	Moderate
Ilie[39]	2016	◐	◐	◐	●	Moderate
Ilie[40]	2017	◐	●	●	◑	Moderate
Ilie[41]	2018	◐	●	●	◑	Moderate
Keller[42]	2018	◐	◑	◑	●	Moderate
Kim[52]	2017	◐	◐	◐	◐	High
Krawczyk[53]	2017	◐	○	◐	◐	High
Marchetti[43]	2017	●	◑	◑	●	Moderate
McLaughlin[54]	2016	◐	○	◑	●	High
Munari[44]	2018	●	◐	◑	●	Moderate
Neuman[55]	2016	●	◐	○	◑	High
Pang[56]	2018	◐	◐	◐	◐	High
Parra[57]	2018	○	◐	◑	◐	High
Paulsen[58]	2017	◐	○	◑	◐	High
Phillips[59]	2015	○	◑	○	●	High
Ratcliffe[23]	2017	◐	◑	●	●	Moderate
Rebelatto[60]	2016	○	◑	◑	◑	High
Rehman[45]	2017	◐	◐	◑	●	Moderate
Rimm[24]	2017	◐	◑	◑	●	Moderate
Roach[34]	2016	●	●	◑	●	Low
Roge[61]	2017	◐	◐	◐	◐	High
Russell-Goldman[29]	2018	●	●	◑	●	Low
Scheel[25]	2016	◐	●	◑	●	Moderate
Scheel[46]	2018	◐	◑	◑	●	Moderate
Sheffield[62]	2016	◐	◑	◑	◐	High
Skov[30]	2017	◐	●	◑	●	Moderate
Smith[63]	2016	◐	◐	◐	◐	High
Soo[47]	2018	◐	◑	◑	●	Moderate
Tsao[48]	2018	◐	◑	◑	●	Moderate
Tseng[64]	2018	◐	◐	◐	◐	High

Table S3. Continued.

First author	Year	Patient selection	Test(s)	Flow and timing	Statistical analysis and reporting	Risk of Bias (RoB)
Vennapusa[49]	2018	●	●	●	◐	Moderate
Xu[50]	2017	◐	◐	◐	●	Moderate

Legend: ● = low RoB (1 point), ◐ = moderate RoB (0.5 points), ● = high RoB (0 points). Scoring system: ≥3.5 points = low RoB; ≥2.5 and <3.5 points = moderate RoB; <2.5 points = high RoB.
For references see main manuscript reference list.

Table S4. Study characteristics of all studies included for data extraction and analysis.

First author	Year	Samples analyzed	Antibodies compared	Scoring method	Cutoffs	Number of observers	Statistical methods
Adam[31]	2018	41 FFPE resection samples (AC, SCC)	Standardized assays: 22C3, 28-8, SP263 LDTs: 22C3, 28-8, SP263, SP142, E1L3N	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	1	• Interassay concordance: weighted κ, OPA • Interlaboratory concordance: weighted κ
Brunnström[32]	2017	TMAs from 55 FFPE tissue blocks (AC, SCC, LCNEC, LCC)	Standardized assays: 22C3, 28-8, SP263, SP142 LDTs: 28-8	Percentage of TC staining (membranous staining of any intensity)	1%, 5%, 25%, 50%	7	• Interassay concordance: weighted κ • Interobserver concordance: weighted κ
Chan[35]	2018	TMAs from 713 FFPE resection samples (AC, SCC, ASC, LCC, SC, LELC)	Standardized assays: 22C3, 28-8, SP263, SP142	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	2	• Interassay concordance: scatter plots with Pearson's R^2 • Interobserver concordance: ICC, Cohen's κ
Cogswell[36]	2017	40 tissue samples (20 from NSCLC, 20 from melanoma)	Standardized assays: 28-8 LDTs: E1L3N	Percentage of TC staining (partial or complete membranous staining)	Not reported	3	Determining sensitivity in staining of PD-L1
Conde[37]	2018	• Discovery cohort: 40 FFPE resection samples (SCC). • Validation cohort: 29 FFPE resection samples (SCC).	Standardized assays: SP263, SP142 LDTs: E1L3N	Percentage of TC staining (membranous and/or cytoplasmic staining)	1%, 5%, 10%, 25%, 50%	2	• Interassay concordance: Pearson's correlation coefficient • Interobserver concordance: ICC, Fleiss' κ
Cooper[33]	2017	TMAs from 108 FFPE resection samples (NSCLC)	Standardized assays: 22C3	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	10	Interobserver concordance: OPA, NPA, PPA, Cohen's κ
Fujimoto[38]	2017	40 FFPE tumor samples (AC, SCC, other)	Standardized assays: 22C3, 28-8, SP263, SP142	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	4	Interassay concordance: weighted κ

Table S4. Continued.

First author	Year	Samples analyzed	Antibodies compared	Scoring method	Cutoffs	Number of observers	Statistical methods
Hendry[28]	2017	TMAs from 423 resections (lung malignancies)	Standardized assays: 22C3, 28-8, SP263, SP142 LDTs: 22C3	Percentage of TC staining (membranous staining of any intensity)	1%, 25%, 50%.	1	Interassay concordance: ICC, OPA, NPA, PPA, Cohen's κ
Hirsch[22]	2017	38 FFPE resection and biopsy samples (NSCLC)	Standardized assays: 22C3, 28-8, SP263, SP142	Percentage of TC staining (partial or complete membranous staining)	1%, 25%.	3	Interassay concordance: scatter plots, regression analysis.
Ilie[39]	2016	56 FFPE resection samples (basaloid SCC)	Standardized assays: SP263, SP142 LDTs: 28-8	Percentage of TC staining (unclear if only membranous staining or also cytoplasmic)	• SP142: 1%, 5%, 50%; • SP263: 25%; • 28-8: 1%, 5%, 10%	3	• Interassay concordance: Spearman correlation coefficient, Cohen's κ • Interobserver concordance: OPA, κ
Ilie[40]	2017	120 FFPE resection and biopsy samples (AC, SCC)	Standardized assays: 22C3 LDTs: 22C3 (2 protocols)	Percentage of TC staining (partial or complete membranous staining of any intensity)	1%, 50%	3	• Interassay concordance: NPA, PPA, ICC. • Interobserver concordance: κ
Ilie[41]	2018	37 FFPE bronchial biopsy samples (AC, SCC)	Standardized assays: 22C3 LDTs: 22C3 (2 protocols)	Percentage of TC staining (partial or complete membranous staining of any intensity)	1%, 50%	2	Interassay concordance: ICC
Keller[42]	2018	TMAs from 370 FFPE resection samples (SCC)	LDTs: SP142, E1L3N	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	1	Interassay concordance: Spearman's ρ correlation
Marchetti[43]	2017	TMAs from 100 FFPE resection samples (AC)	Standardized assays: 22C3, SP263	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	4 pathologists/ 4 centers	• Interassay concordance: Pearson's precision analysis, weighted κ/Light's κ • Interobserver concordance: weighted κ/Fleiss' κ • Interlaboratory concordance: ICC

Table S4. Continued.

First author	Year	Samples analyzed	Antibodies compared	Scoring method	Cutoffs	Number of observers	Statistical methods
Munari[44]	2018	TMAs from 198 FFPE resection samples (AC, SCC, other)	Standardized assays: 22C3, SP263 LDTs: 22C3	Percentage of TC staining (membranous staining of any intensity)	1%, 50%	2	• Interassay concordance: OPA, PPA, NPA, Cohen's κ • Interobserver concordance: Cohen's κ
Ratcliffe[23]	2017	493 FFPE tissue samples (non-squamous carcinoma, SC, ASC)	Standardized assays: 22C3, 28-8, SP263	Percentage of TC staining (membranous staining)	1%, 10%, 25%, 50%	2	• Interassay concordance: Spearman correlation coefficient, OPA, NPA, PPA • Interobserver concordance (determined in subset of 200 samples): OPA
Rehman[45]	2017	105 FFPE tissue samples (AC, SCC)	LDTs: SP142	Percentage of TC staining (predominantly membranous staining at any intensity)	None	5	Interobserver concordance: ICC
Rimm[24]	2017	90 resection samples (AC, SCC)	Standardized assays: 22C3, 28-8 LDTs: SP142, E1L3N	Percentage of TC staining (membranous and/or cytoplasmic staining)	1%, 5%, 10%, 25%, 50%	13	• Interassay concordance: paired Wilcoxon signed-rank test, ICC • Interobserver concordance: ICC, Fleiss' κ
Roach[34]	2016	• Interobserver concordance: 62 FFPE tissue samples (NSCLC) • Interlaboratory concordance: 36 FFPE tissue samples (NSCLC)	Standardized assays: 22C3	Percentage of TC staining (partial or complete membranous staining of intensity 1+ to 3+)	50%	3 pathologists/ 3 centers	• Interobserver concordance: OPA, NPA, PPA • Interlaboratory concordance: OPA, NPA, PPA
Russell-Goldman[29]	2018	46 FFPE resection and biopsy samples (NSCLC)	LDTs: E1L3N	Percentage of TC staining (partial or complete membranous staining of any intensity)	1%, 50%	2	Interobserver concordance: Bland-Altman plots, ICC

Table S4. Continued.

First author	Year	Samples analyzed	Antibodies compared	Scoring method	Cutoffs	Number of observers	Statistical methods
Scheel[25]	2016	• Training set: 15 FFPE tissue samples (AC, SCC) • Validation set: 15 FFPE tissue samples (AC, SCC)	Standardized assays: 22C3, 28-8, SP263, SP142 LDTs: SP142, E1L3N	Percentage of TC staining (partial or complete membranous staining)	1%, 5%, 10%, 25%, 50%	9	• Interassay concordance: pairwise comparison of assays (percentage of concordance) • Interobserver concordance: Cohen's κ and Light's κ
Scheel[46]	2018	TMAs from 21 FFPE resection samples (NSCLC)	Standardized assays: 22C3, 28-8, SP263 LDTs: 22C3, 28-8, SP263, SP142, E1L3N, QR1	Percentage of TC staining (unclear if only membranous staining or also cytoplasmic)	• 3-step score: 1%, 50% • 6-step score: 1%, 5%, 10%, 25%, 50%	10 centers	• Interassay concordance: Light's κ • Interlaboratory concordance: Cohen's κ of each local result compared with central reference score
Skov[30]	2017	87 FFPE resection and biopsy samples (AC, SCC, LCNEC, NSCLC NOS, mesothelioma, metastasis to lung)	Standardized assay: 22C3, 28-8	Percentage of TC staining (partial or complete membranous staining of any intensity)	50%	1	Interassay concordance: OPA, NPA, PPA, Pearson correlation coefficient
Soo[47]	2018	18 FFPE samples from resections, biopsies, pleural fluid and FNAs (NSCLC)	Standardized assays: 22C3, 28-8, SP263, SP142 LDTs: E1L3N	Percentage of TC staining (unclear if only membranous staining or also cytoplasmic)	Not reported	3	Interassay concordance: Spearman's rank test
Tsao[48]	2018	81 paraffin samples from resections, biopsies, lymph node excisions and cytological cell blocks (AC, SCC, NSCLC NOS, SCLC)	Standardized assays: 22C3, 28-8, SP263, SP142 Unclear: 73-10	Percentage of TC staining (partial or complete membranous staining)	1%, 5%, 10% 25%, 50%, 80%	24	• Interassay concordance: comparison of best-fit curves • Interobserver concordance: ICC, Fleiss' κ

Table S4. Continued.

First author	Year	Samples analyzed	Antibodies compared	Scoring method	Cutoffs	Number of observers	Statistical methods
Vennapusa[49]	2018	• Interobserver concordance: 80 FFPE resection and biopsy samples (NSCLC) • Interlaboratory concordance: 28 FFPE resection and biopsy samples (NSCLC)	Standardized assays: SP142	Percentage of TC staining (membranous staining of any intensity)	Combined scoring of TCs and ICs (TCs: 1%, 5%, 50%; ICs: 1%, 5%, 10%).	3 pathologists/ 3 centers	• Interobserver concordance: OPA, NPA, PPA. • Interlaboratory concordance: OPA, NPA, PPA
Xu[50]	2017	135 resection samples (AC, SCC, other)	Standardized assays: 22C3, SP142	Percentage of TC staining (partial or complete membranous staining)	1%, 5%, 50%	2	Interassay concordance: weighted κ, McNemar-Bowker test.

Abbreviations: AC=adenocarcinoma; ASC=adenosquamous carcinoma; FFPE=formalin-fixed paraffin-embedded; FNA=fine needle aspiration; IC=immune cell; ICC=intraclass correlation coefficient; κ=kappa; LCC=large cell carcinoma; LCNEC=large cell neuro-endocrine carcinoma; LDT=laboratory-developed test; LELC=lymphoepithelioma-like carcinoma; NPA=negative percent agreement; NSCLC NOS=non-small cell lung carcinoma not otherwise specified; OPA=overall percent agreement; PPA=positive percent agreement; SC=sarcomatoid carcinoma; SCC=squamous cell carcinoma; SCLC=small cell lung carcinoma; TC=tumor cell; TMA=tissue-microarray. For references see main manuscript reference list.

Table S5. Results from studies assessing interassay concordance of TC staining.

First author	Year	Study results (interassay concordance)
Adam[31]	2018	• High concordance between assays 22C3, 28-8, and SP263 (κ 0.71-0.89). • 14 of 27 LDTs (51.8%) concordant with one of the reference assays (κ >0.75).
Brunnström[32]	2017	• Interassay concordance range for comparison of all standardized assays and 28-8 LDT: κ 0.45-0.91. Lowest values for comparison of SP142 with other assays (0.45-0.63). • Agreement between assays higher with 50% cutoff than with 1% cutoff.
Chan[35]	2018	• High agreement between 22C3, 28-8 and SP263 (Pearson R^2 0.841-0.873). Lower correlation with SP142 ($R^2 \approx 0.70$). • Lower OPA (68.6-82%) at 1% cutoff compared to 50% cutoff (94.4-97.9%)
Cogswell[36]	2017	• Higher sensitivity using 28-8 compared to E1L3N. • 28-8 assay more frequently detected PD-L1 positive TCs (22 vs. 6) compared with E1L3N.
Conde[37]	2018	• Very good correlation between E1L3N and SP263 in both cohorts (ρ = 0.94 and ρ = 0.99). • Lower correlation between SP263 and SP142 (ρ = 0.88 and ρ = 0.87).
Fujimoto[38]	2017	• Higher concordance between 22C3, 28-8 and SP263 (κ 0.64-0.71) than between SP142 and other assays (κ 0.39-0.55). • Higher agreement at 50% cutoff than at 1% cutoff.
Hendry[28]	2017	• Good agreement between 22C3, 28-8, SP263 and SP142 on continuous scale (ICC 0.674). Higher when SP142 was excluded (ICC 0.755). • Moderate agreement between assays with use of clinical cutoffs (κ 0.43). • Excellent agreement between 22C3 assay and 22C3 LDT (ICC 0.921, κ 0.897 for cutoff 50%).
Hirsch[22]	2017	• High correlation between 22C3, 28-8 and SP263, lower correlation for all comparisons including SP142. • Replacement of validated cutoff with any other cutoff reduced overall agreement for each assay.
Ilie[39]	2016	• Poor correlation between SP142 and SP263 (κ 0.362) or 28-8 (κ 0.412). • Good correlation between SP263 and 28-8 (κ 0.883).
Ilie[40]	2017	High concordance between two 22C3 LDTs and 22C3 assay (ICC 98.7-99.9%).
Ilie[41]	2018	High correlation between two 22C3 LDTs and 22C3 assay (ICC 0.999 and 1.000).
Keller[42]	2018	Significant correlation of TPS between E1L3N and SP142 (r = 0.781; P < 0.001), although E1L3N showed higher sensitivity.
Marchetti[43]	2017	• Correlation between 22C3 and SP263 0.89-0.97 for 4 participating centers. • 50% cutoff: κ values for all centers 0.844-1. • 1% cutoff: κ values for all centers 0.62-0.83.
Munari[44]	2018	• OPA between 22C3 assay and SP263 77.3% (κ 0.518) and 68.6% (κ 0.390). 22C3 stained significantly lower proportion of cases than SP263. • OPA between 22C3 LDT and SP263 81.5% (κ 0.624) and 76.1% (κ 0.572). • OPA between 22C3 assay and 22C3 LDT 84.7% (κ 0.595) and 80.3% (κ 0.583).
Ratcliffe[23]	2017	• High associations between assays 22C3, 28-8 and SP263 (Spearman correlations >0.9). • OPA of >90% between assays at multiple cutoffs. • NPA and PPA >85% for each comparison at different cutoffs.
Rimm[24]	2017	• Only scores of 28-8 and E1L3N not statistically significantly different. • SP142 greatest magnitude of difference compared to the other 3 assays. • ICC based on average scores 0.81, which increased to 0.97 after SP142 exclusion.
Scheel[25]	2016	• OPA 41%-72% for pairwise comparison of all standardized assays/LDTs. Highest concordance between 22C3 and 28-8. • SP142 stained lower proportions of TCs than 22C3 and 28-8, SP263 stained higher proportions of TCs.

Table S5. Continued.

First author	Year	Study results (interassay concordance)
Scheel[46]	2018	• Similar proportions of TC staining for 22C3 and 28-8, fewer TCs stained with SP142, more TCs stained with SP263. • Similar staining patterns to 22C3 and 28-8 for 6 of 11 LDTs. • Substantial to near-perfect concordance between assay 22C3 and 28-8. • Moderate concordance between assay SP263 and 22C3 or 28-8.
Skov[30]	2017	• High level of agreement between 22C3 and 28-8 (R^2 0.95). • OPA, NPA and PPA high for all cutoffs (1%, 5%, 10% and 50%).
Soo[47]	2018	Considerable variation in TC staining. Lowest correlation between 28-8 and SP142 (R^2 0.25). Highest between 22C3 and E1L3N (R^2 0.71).
Tsao[48]	2018	• Close approximation between best-fit curves of 22C3, 28-8 and SP263. • Lower sensitivity of SP142, higher sensitivity of 73-10.
Xu[50]	2017	• Assay SP142 stained fewer TCs compared to 22C3. • Using the 22C3 scoring algorithm, κ was 0.481 between assays. • Using the SP142 scoring algorithm, κ was 0.324 between assays.

Abbreviations: IC=immune cell; ICC=intraclass correlation coefficient; κ=kappa; LDT=laboratory-developed test; NPA=negative percent agreement; OPA=overall percent agreement; PPA=positive percent agreement; ρ=Pearson's correlation coefficient; TC=tumor cell.
For references see main manuscript reference list.

Table S6. Results from studies assessing interobserver concordance of TC scoring.

First author	Year	Study results (interobserver concordance)
Brunnström[32]	2017	Varying κ values for antibodies: • 22C3 0.71-0.95; • 28-8 assay 0.80-0.93; • SP263 0.75-0.91; • SP142 0.81-0.96; • 28-8 (LDT) 0.80-0.95. Number of differently classified cases significantly higher for 1% cutoff.
Chan[35]	2018	• ICC highest for SP263 (0.967), then 22C3 (0.963), 28-8 (0.932) and last SP142 (0.916). • Higher agreement for 50% cutoff than for 1% cutoff.
Conde[37]	2018	• High ICCs for all antibodies (assays SP263 and SP142 and E1L3N) in both cohorts. • Highest concordance for 50% cutoff.
Cooper[33]	2017	Assay 22C3: • 1% cutoff: OPA 84.2%, κ 0.68; • 50% cutoff: OPA 81.9%, κ 0.58.
Ilie[39]	2016	High interobserver agreement for all antibodies: • SP263 OPA 98%, κ 0.976; • SP142 OPA 92%, κ 0.910; • 28-8 OPA 96%, κ 0.935.
Ilie[40]	2017	• 1% cutoff: κ 1 for all antibodies (22C3 assay and LDTs); • 50% cutoff: κ 1 for both 22C3 LDTs, κ 0.99 for 22C3 assay.
Marchetti[43]	2017	Higher κ for 22C3 and SP263 for 50% cutoff (κ 0.931 and 0.942) than for 1% cutoff (κ 0.754 and 0.798)
Munari[44]	2018	Good concordance between pathologists: • SP263 κ 0.73; • 22C3 assay κ 0.77; • 22C3 LDT κ 0.72.
Ratcliffe[23]	2017	Assays 22C3, 28-8 and SP263: • OPAs at cutoffs 10%, 25% and 50% were >85%; • OPAs lower at 1% cutoff (75.9%-77.0%).
Rehman[45]	2017	SP142 LDT: ICC 94%
Rimm[24]	2017	Assays 22C3 and 28-8 and E1L3N and SP142 LDT: • ICCs between 0.83 and 0.88; • Agreement higher at 50% cutoff (κ 0.75) than at 1% cutoff (κ 0.54).
Roach[34]	2016	Assay 22C3: OPA 92.7%, NPA 92.6%, PPA 92.8%.
Russell-Goldman[29]	2018	• Very high agreement (ICC 0.96) for E1L3N. • Agreement higher at 50% cutoff than at 1% cutoff (98% vs. 79%).
Scheel[25]	2016	• Training set: κ 0.50 for E1L3N, κ 0.49 for SP142. Higher concordance with use of dichotomous cutoff criteria (κ 0.61-0.80). • Validation set: moderate concordance levels for the assays 22C3, 28-8, SP263 and SP142 (κ 0.47-0.49). Higher concordance with use of dichotomous cutoff criteria (κ 0.59-0.80).
Tsao[48]	2018	Assays 22C3, 28-8, SP263, SP142 and 73-10: • ICC for glass slide reading 0.88-0.93. For digital reading 0.80-0.91; • High-level reliability at various cutoffs, especially 5%, 10%, 25% and 50% (κ > 0.7); • Slightly diminished reliability at 1% and 80% cutoff.
Vennapusa[49]	2018	Assay SP142: OPAs of 92.7%, 93.8% and 93.5% for TC1/IC1, TC2/IC2, and TC3/IC3.

Abbreviations: IC=immune cell; ICC=intraclass correlation coefficient; κ=kappa; LDT=laboratory-developed test; NPA=negative percent agreement; OPA=overall percent agreement; PPA=positive percent agreement; TC=tumor cell.
For references see main manuscript reference list.

Table S7. Results from studies assessing interlaboratory concordance of TC staining.

First author	Year	Study results (interlaboratory concordance)
Adam[31]	2018	Concordance for assays 22C3, 28-8 and SP263: κ 0.79-0.94.
Marchetti[43]	2017	• Assay 22C3: ICC 0.973 for 4 centers. • Assay SP263: ICC 0.986 for 4 centers.
Roach[34]	2016	Assay 22C3: OPA 88.3%, NPA 90.3%, PPA 85.2%.
Scheel[46]	2018	• 6-step scoring system: substantial concordance for assays 22C3, 28-8 and SP263 (κ 0.63-0.69), moderate concordance for LDTs (κ 0.43). • 3-step scoring system: nearly perfect concordance for assays 22C3, 28-8 and SP263 (κ 0.73-0.89), moderate concordance for LDTs (κ 0.5).
Vennapusa[49]	2018	Assay SP142: inter-site agreement of 87.6%, 87.6% and 91.0% for TC1/IC1, TC2/IC2, and TC3/IC3.

Abbreviations: IC=immune cell; ICC=intraclass correlation coefficient; κ=kappa; LDT=laboratory-developed test; NPA=negative percent agreement; OPA=overall percent agreement; PPA=positive percent agreement; TC=tumor cell. For references see main manuscript reference list.

Data S1. Supplementary results: concordance of immune cell staining and scoring.

Interassay concordance

Eleven studies assessed interassay concordance of immune cell (IC) staining (Table S8). Many of these found poor agreement of IC staining between assays and/or LDTs[28, 31, 35-37] or greater variability in staining pattern for ICs than for TCs[22, 50]. One study[39] showed much higher concordance (κ 0.721) between assays SP263 and 28-8 than between assays SP142 and SP263 (κ 0.018) and between assays SP142 and 28-8 (κ 0.134). Similarly, Rimm et al.[24] found that interassay concordance of IC staining greatly approved after exclusion of SP142 (increase of ICC from 0.27 to 0.80). The Blueprint phase 2 study[48], moreover, reported comparable distributions of IC scores among 22C3, 28-8 and SP263, while antibody SP142 showed lesser staining of ICs compared with the other antibodies.

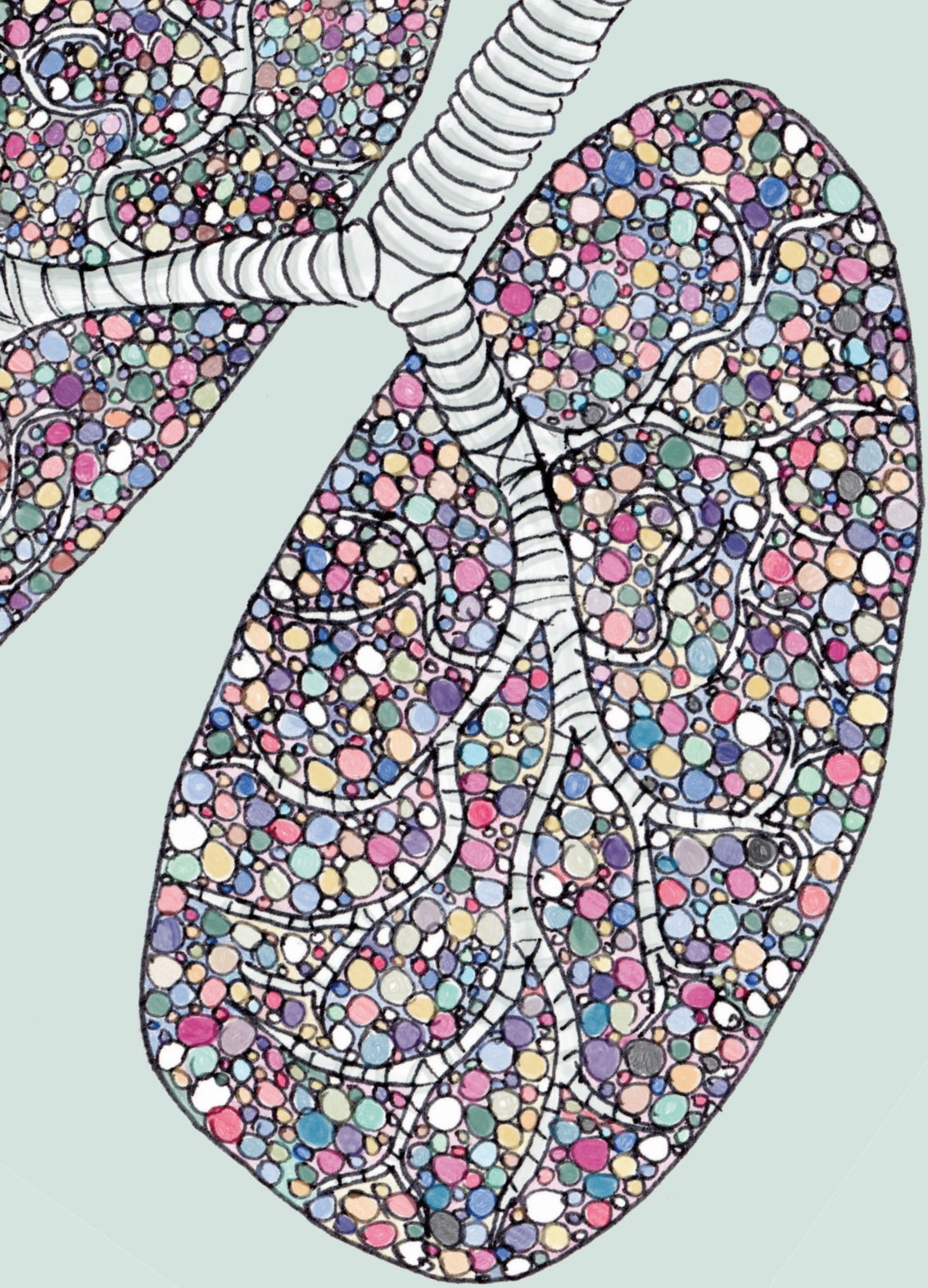
Interobserver concordance

Seven studies assessed interobserver concordance of IC scoring (Table S8). All of these reported lower concordance values of IC scoring compared to TC scoring for all antibodies[24, 25, 29, 37, 39, 45, 48]. Only Conde et al.[37] reported comparable concordance levels for both IC and TC scoring in a cohort of 40 patients, while also reporting lower concordance levels for IC scoring than for TC scoring in another (validation) cohort of 29 patients.

Table S8. Results from studies assessing interassay and/or interobserver concordance of IC staining/scoring.

First author	Year	Study results (interassay concordance of IC staining)	Study results (interobserver concordance of IC scoring)
Adam[31]	2018	Poor OPA when comparing assays 28-8, 22C3 and SP263 and when comparing LDTs to these assays.	NA
Chan[35]	2018	Low concordance for IC scoring between all assays (R^2 0.263-0.682).	NA
Cogswell[36]	2017	Assay 28-8 detected PD-L1 positive ICs more frequently than E1L3N (15 vs. 7).	NA
Conde[37]	2018	Correlation for ICs lower than for TCs, lowest correlation for comparisons involving SP142.	High ICCs for all antibodies (SP263, SP142, E1L3N) in discovery cohort (0.92-0.96), lower ICCs in validation cohort (0.76-0.81).
Hendry[28]	2017	Poor overall and pairwise agreement between IC stainings (ICC 0.212).	NA
Hirsch[22]	2017	Greater variability in IC staining than in TC staining between assays. Highest concordance between 22C3 and 28-8.	NA
Ilie[39]	2016	• Poor agreement between SP142 and SP263 (κ 0.018) or 28-8 (κ 0.134). • Good correlation between SP263 and 28-8 (κ 0.721).	Agreement for ICs lower than for TCs: • SP142 OPA 81%, κ 0.786; • SP263 OPA 87%, κ 0.832; • 28-8 OPA 86%, κ 0.817.
Rehman[45]	2017	NA	IC scoring much less concordant (ICC 27%) than TC scoring (ICC 94%).
Rimm[24]	2017	ICC was 0.27 for comparison of 22C3, 28-8, E1L3N and SP142. ICC increased to 0.80 after SP142 exclusion.	ICCs for IC scoring much lower (0.17-0.23) than for TC scoring (0.83-0.33).
Russell-Goldman[29]	2018	NA	• Moderate agreement (ICC 0.47). • Agreement higher at 10% cutoff than at 1% cutoff (77% vs. 75%).
Scheel[25]	2016	NA	In both training set and validation set low concordance of IC scoring (mostly κ <0.2).
Scheel[46]	2018	• Similar IC staining patterns for assay 22C3 and 28-8, more intense staining with SP263. • Distinct IC staining pattern with SP142.	NA
Tsao[48]	2018	• Distribution of IC scores comparable among 22C3, 28-8 and SP263. • Greater and lesser staining of ICs by 73-10 and SP142.	Overall poor agreement (κ 0.11-0.28 for glass slide reading and κ 0.08-0.27 for digital reading).
Xu[50]	2017	Variability in staining pattern greater for ICs than for TCs. Lower PD-L1 detection levels with SP142 compared to 22C3.	NA

Abbreviations: IC=immune cell; ICC=intraclass correlation coefficient; κ=kappa; LDT=laboratory-developed test; NA=not applicable; OPA=overall percent agreement; TC=tumor cell. For references see main manuscript reference list.



3

FORMALIN FIXATION FOR OPTIMAL CONCORDANCE OF PROGRAMMED DEATH-LIGAND 1 IMMUNOSTAINING BETWEEN CYTOLOGIC AND HISTOLOGIC SPECIMENS FROM PATIENTS WITH NON-SMALL CELL LUNG CANCER

Koomen BM
van der Starre-Gaal J
Vonk JM
von der Thüsen JH
van der Meij JJC
Monkhorst K
Willems SM
Timens W
't Hart NA

Cancer Cytopathology. 2021;129(4):304-317

ABSTRACT

Background

Immunohistochemical staining of programmed death-ligand 1 (PD-L1) is used to determine which patients with non–small cell lung cancer (NSCLC) may benefit most from immunotherapy. Therapeutic management of many patients with NSCLC is based on cytology instead of histology. In this study, concordance of PD-L1 immunostaining between cytology cell blocks and their histologic counterparts was analyzed. Furthermore, the effect of various fixatives and fixation times on PD-L1 immunoreactivity was studied.

Methods

Paired histologic and cytologic samples from 67 patients with NSCLC were collected by performing fine-needle aspiration on pneumonectomy/lobectomy specimens. Formalin-fixed agar-based or CytoLyt/PreservCyt-fixed Cellient cell blocks were prepared. Sections from cell blocks and tissue blocks were stained with SP263 (standardized assay) and 22C3 (laboratory-developed test) antibodies. PD-L1 scores were compared between histology and cytology. In addition, immunostaining was compared between PD-L1–expressing human cell lines fixed in various fixatives at increasing increments in fixation duration.

Results

Agar cell blocks and tissue blocks showed substantial agreement ($\kappa = 0.70$ and $\kappa = 0.67$, respectively), whereas fair-to-moderate agreement was found between Cellient cell blocks and histology ($\kappa = 0.28$ and $\kappa = 0.49$, respectively). Cell lines fixed in various alcohol-based fixatives showed less PD-L1 immunoreactivity compared with those fixed in formalin. In contrast to SP263, additional formalin fixation after alcohol fixation resulted in preserved staining intensity using the 22C3 laboratory-developed test and the 22C3 pharmDx assay.

Conclusions

Performing PD-L1 staining on cytologic specimens fixed in alcohol-based fixatives could result in false-negative immunostaining results, whereas fixation in formalin leads to higher and more histology-concordant PD-L1 immunostaining. The deleterious effect of alcohol fixation could be reversed to some degree by post-fixation in formalin.

INTRODUCTION

In non–small cell lung cancer (NSCLC), immunohistochemical expression of programmed death-ligand 1 (PD-L1) on tumor cells is shown to predict the likelihood of response to anti-PD1/anti–PD-L1 immunotherapy[1]. The detection of PD-L1 in tumor tissue by immunohistochemistry (IHC) is required as a companion diagnostic for the immune checkpoint inhibitor pembrolizumab, which may only be prescribed as first-line monotherapy to patients with advanced NSCLC when tumor cells have ≥50% PD-L1 expression[2, 3]. For patients with locally advanced (stage III) NSCLC whose disease has not progressed after treatment with radiation and platinum-based chemotherapy, the anti–PD-L1 checkpoint inhibitor durvalumab may be prescribed. This drug, however, only received approval by the European Medicines Agency for patients whose tumors have PD-L1 expression in ≥1% tumor cells[4]. For the immune checkpoint inhibitors nivolumab and atezolizumab, PD-L1 IHC is used as a complementary diagnostic, rendering it less important to gather adequate tissue for PD-L1 immunostaining. Nevertheless, it may identify patients who could respond better to treatment with these checkpoint inhibitors than others and aid in the assessment of risks and benefits for individual patients[5, 6], especially when the presence of tumor-infiltrating lymphocytes is also taken into account[7].

Because clinical trials addressing immune checkpoint inhibitors included patients with tissue-based diagnoses, the use of IHC assays for evaluating PD-L1 expression in tumor cells is validated only in histologic specimens[8, 9]. In clinical practice, however, the management of many patients with advanced NSCLC is based on cytology[10]. Diagnostic cytology by fine-needle aspiration (FNA) is less invasive than tissue-based diagnostics through histologic biopsies and thus is preferable[8, 11]. The sensitivity, specificity, and positive predictive values of PD-L1 immunocytochemistry (ICC), however, are unclear, with only a limited number of studies focusing on histologic and cytologic correlation of PD-L1 immunostaining[8, 9, 12-16].

In addition, routinely used fixatives in cytology are often based on methanol or ethanol compared with formalin-based fixatives in histology[17]. This might negatively affect the staining intensity of IHC assays[18-21], resulting in false-negative analyses. The effect of different pre-analytical variables on PD-L1 immunoreactivity in cytologic specimens is largely unknown, leaving it unclear whether fixatives other than formalin can reliably be used for determining PD-L1 expression through immunostaining.

The objective of the current multicenter study was to compare PD-L1 immunostaining in matched histologic specimens and cytologic cell blocks by using FNA material and histologic samples from the same resected lung tumor. Second, the effects of various fixation solutions as well as different fixation times on PD-L1 immunostaining were studied

using PD-L1–expressing human cell lines. Preliminary and limited results from the study were previously reported in an item of correspondence[22]. Complete, extended results are presented here.

MATERIALS AND METHODS

Collection of Histologic and Cytologic Specimens

Five Dutch pathology laboratories collaborated to collect material for the comparison of PD-L1 immunostaining in histologic and cytologic specimens from patients with NSCLC (University Medical Center Groningen, Isala Zwolle, Erasmus Medical Center Rotterdam, Netherlands Cancer Institute Amsterdam, and Pathology Friesland). In each center, paired histologic and cytologic samples were collected. To do so, FNAs were performed on pneumonectomy or lobectomy specimens with a palpable or visible tumor to obtain cytologic samples that were as close as possible to routine FNA-derived specimens. The collection of cytologic material was performed before further preparation and fixation of the resection specimen. It has been demonstrated previously that collecting FNA material this way can be done without compromising routine histologic evaluation of the tumor; thus it is a safe method that can be used within the outlines of the code of conduct for responsible use of residual human tissue for research established by the Federation of Dutch Medical Scientific Societies.[23] Histology samples were taken after 18 to 72 hours following routine protocols used in clinical practice. The mean estimated fixation time was between 18 hours (overnight fixation) and 72 hours (resection specimens that remained in formalin over the weekend). All patient material was used anonymously and was collected and used in accordance with the Federation of Dutch Medical Scientific Societies code of conduct and with the General Data Protection Regulation.

After obtaining the FNA specimen, each center was allowed to use their routine method for fixation and generation of a cell block (see Supporting Table 1). In this way, concordance of PD-L1 immunostaining could be analyzed between histologic tissue and cell blocks that were processed in different ways, reflecting normal, everyday practice. Depending on the locally developed protocols, either a formalin-fixed, agar-based cell block or a CytoLyt/PreservCyt-fixed Cellient cell block was prepared. The Cellient Automated Cell Block System (Hologic Inc) was used for the latter. From each cell block, a slide was cut and stained with hematoxylin and eosin to check for the presence of tumor cells. If enough viable tumor cells were present (≥100 tumor cells), sequential, 3-µm-thick slices were cut for staining with 2 separate PD-L1 antibodies, i.e. the Ventana SP263 standardized assay (Ventana Medical Systems Inc) and a Dako 22C3 laboratory-developed test (LDT). Formalin-fixed paraffin-embedded (FFPE) tissue sections were cut from a routinely made histologic tissue block from the same tumor. These sections were also stained for PD-L1

with both antibodies (SP263 and 22C3) and were used for histologic comparison (for a schematic representation of the study design, see Fig. 1).

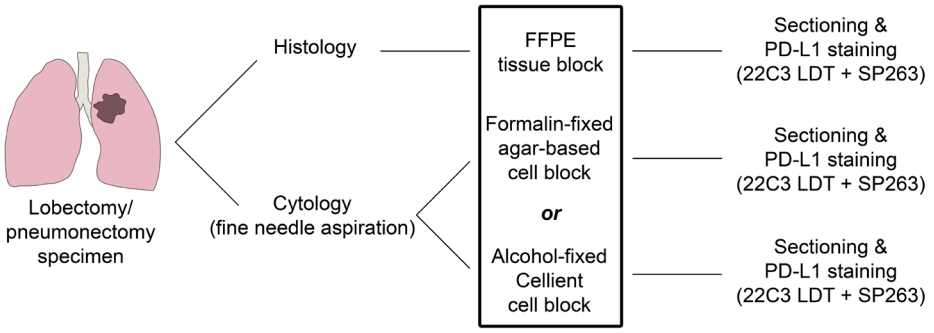


Figure 1. This is a schematic representation of the study design. FFPE indicates formalin-fixed, paraffin-embedded; LDT, laboratory-developed test (using the 22C3 antibody); PD-L1, programmed death-ligand 1; SP263, antibody used in the standardized assay.

Preparation of PD-L1–Expressing Cell Lines

To further evaluate the effects of different fixatives on PD-L1 immunostaining in cytologic samples, commercially available cell lines with high PD-L1 expression were used (T-cell non-Hodgkin lymphoma cell lines from HistoCyte Laboratories Ltd[24]). The cell lines were fixed in either 5 mL 10% neutral-buffered formalin (NBF) or 25 mL CytoLyt, PreservCyt, CytoRich Red, or Carbowax. CytoLyt and PreservCyt (Hologic Inc) are methanol-based fixatives, while Carbowax (Dow Chemical Company) contains ethanol and polyethylene glycol. CytoRich Red (Thermo Scientific) is alcohol-based as well (methanol, isopropyl alcohol, and ethylene glycol) but also contains a small amount of formaldehyde. Each is used as a (pre-)fixation solution in clinical practice. In addition, cell lines were fixed in CytoLyt, PreservCyt, CytoRich Red, or Carbowax followed by 30-minute fixation in 10% NBF. Various fixation times were used, which allowed us to evaluate the effect of different fixation periods on PD-L1 immunostaining. The fixation periods used were 2 hours and 24 hours for all fixatives and 48 hours for the cell lines fixed in NBF and CytoLyt only. This design allowed for 20 different fixation schemes (see Supporting Table 2). Subsequently, agarose pellets were created for each cell line and were then processed into paraffin blocks. Cores from each block were assembled in paraffin-embedded cell-microarray (CMA) blocks. These were sent to the University Medical Center in Groningen, where slides were cut (3-µm thickness) for PD-L1 immunostaining.

PD-L1 IHC Staining and Scoring

Sections cut from each cell block and tissue block were stained with the Ventana SP263

antibody and the Dako 22C3 antibody. All staining of patient material was performed within 1 center (University Medical Center Groningen). Staining of slides with the SP263 standardized assay was carried out on a Ventana Benchmark Ultra platform according to the manufacturer's instructions. The Dako 22C3 was used as an LDT, also using the Ventana Benchmark Ultra platform. Previously, this LDT was compared with the Dako 22C3 pharmDx assay on Dako Link 48, for optimization and validation of its use in routine clinical practice. The addition of an amplification step led to the best protocol, which was comparable to a previously published protocol by Adam et al.[25]. The same SP263 standardized assay and the 22C3 LDT were used to stain sections from the CMA blocks containing the cell line cores. In addition, sections from the CMA blocks were stained with the Dako 22C3 pharmDx assay, used on a Dako Autostainer Link 48 platform according to the manufacturer's instructions (performed in Martini Hospital, Groningen, the Netherlands).

Staining patterns were analyzed in cores from each cell line and compared visually between the 20 different fixation schemes. Differences in PD-L1 expression between cell lines were quantified by determining PD-L1 H-scores for each cell core, using an application in Visiopharm software (Visiopharm A/S)[26]. The H-score was calculated by determining staining intensity in each cell (divided into levels 0, 1+, 2+, and 3+), followed by application of the following formula: 1 * (% of cells with staining intensity level 1+) + 2 * (% of cells with staining intensity level 2+) + 3 * (% of cells with staining intensity level 3+).[27, 28] All stained slides from the paired cell and tissue blocks of included patients were reviewed independently by 2 trained pathologists. Cases of disagreement were resolved through discussion. PD-L1 expression was scored in tumor cells according to the guidelines provided by Roche/Ventana or Dako as part of the PD-L1 IHC pharmDx test. For each slide, the pathologists determined the tumor proportion score (TPS), which is constructed by determining the percentage of viable tumor cells that show membranous PD-L1 immunostaining relative to the total amount of tumor cells. This score was used to categorize the samples into 3 groups: TPS <1% (negative), TPS 1% to 49% (weakly positive), and TPS ≥50% (strongly positive).

Statistical Analysis

To assess agreement of the PD-L1 TPS between histologic and cytologic samples, weighted κ values (linear weights) were calculated. Furthermore, the Cohen κ was calculated using data dichotomized according to the 1% and 50% cutoffs. Overall percent agreement (OPA), positive percent agreement, and negative percent agreement were determined for both cutoffs using histology as the reference standard. In addition, the McNemar-Bowker test of symmetry was applied to assess whether the categorization of PD-L1 expression differed significantly between histologic and cytologic samples. P values <.05 were considered statistically significant. Both interobserver and interassay agreement were assessed

using weighted κ and Cohen κ values for the 1% and 50% cutoffs. Statistical analysis was performed using RStudio version 1.1.456 (R Foundation for Statistical Computing) and IBM SPSS Statistics version 25 (IBM Corporation).

RESULTS

Sample Selection

Paired histologic and cytologic samples from 85 patients were collected. Fifteen were excluded, because these patients had a diagnosis other than NSCLC. Three patients were excluded because 1 or both of their samples contained an insufficient number of viable tumor cells (<100). The remaining 67 patients all had a diagnosis of NSCLC, with various histologic subtypes (adenocarcinoma, n = 38; squamous cell carcinoma, n = 25; pleomorphic carcinoma, n = 2; adenosquamous carcinoma, n = 1; and NSCLC not otherwise specified, n = 1). Cytologic material from 33 patients was processed into agar-based cell blocks. The cytologic material from the other 34 patients was processed into Cellient cell blocks.

Overall Concordance Between Histology and Cytology

When the 22C3 LDT was used for determining PD-L1 expression in histologic samples, 25 samples (37%) had a TPS <1%, 20 samples (30%) had a TPS from 1% to 49%, and 22 samples (33%) had a TPS ≥50%. When we used the 22C3 antibody on cytologic samples, 34 samples (51%) had a TPS <1%, 22 samples (33%) had a TPS from 1% to 49%, and 11 samples (16%) had a TPS ≥50%. Of all 67 samples, 39 (58%) showed concordant results between histology and cytology. A weighted κ value of 0.49 was identified, which can be described as moderate agreement[29]. Dichotomizing data according to the 1% and 50% cutoffs also resulted in moderate agreement levels (κ = 0.49 and κ = 0.50, respectively) (see Supporting Table 3).

When the SP263 antibody was used on histologic samples, 25 samples (37%) had a TPS <1%, 24 samples (36%) had a TPS from 1% to 49%, and 18 samples (27%) had a TPS ≥50%. Performing SP263 immunostaining on cytologic samples resulted in a TPS <1% in 37 samples (55%), a TPS from 1% to 49% in 18 samples (27%), and a TPS ≥50% in 12 samples (18%). With this antibody, 45 of 67 samples (67%) showed concordance between histology and cytology. The agreement observed can be described as moderate (κ = 0.59). Dichotomization according to the 1% and 50% cutoffs resulted in moderate-to-substantial agreement levels, suggesting a somewhat higher level of agreement between cytologic and histologic material at the 50% cutoff (κ = 0.66) compared with the 1% cutoff (κ = 0.53) with use of the SP263 antibody (see Supporting Table 3).

Concordance Between Histology and Agar or Cellient Cell Blocks

Separate analyses of agar-based and Cellient-processed cell blocks revealed a higher degree of agreement between the formalin-fixed agar cell blocks and histology than between the alcohol-fixed Cellient cell blocks and histology. When using the 22C3 LDT, substantial agreement levels were found when analyzing concordance between agar cell blocks and histology (OPA, 73%; κ = 0.70) (Table 1). Agreement between Cellient cell blocks and histology can be described as fair (OPA, 44%; κ = 0.28). In addition, a comparison of the categorization of PD-L1 expression between histologic samples and alcohol-fixed Cellient samples revealed a statistically significant difference (P < .01), whereas no statistically significant difference was observed between histologic samples and formalin-fixed, agar-based samples (P = .407) (see Supporting Table 4). An overview of concordance and discordance between matched samples using the 22C3 LDT is displayed by Figure 2.

Table 1. Concordance of PD-L1 expression (TPS) between histological and cytological specimens when 22C3 LDT was used to stain for PD-L1.

	Agreement of TPS in 3 categories (<1%; 1-49%; ≥50%)		**Agreement when dichotomizing TPS at 1% cutoff**				**Agreement when dichotomizing TPS at 50% cutoff**			
	OPA (%)	Weighted κ (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's κ (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's κ (95% CI)
Agar (formalin-fixed)	73%	0.70 (0.51-0.88)	88%	90%	83%	0.74 (0.49-0.99)	85%	67%	95%	0.65 (0.37-0.94)
Cellient (alcohol-fixed)	44%	0.28 (0.06-0.49)	62%	48%	85%	0.28 (0.00-0.57)	76%	20%	100%	0.26 (-0.05-0.57)

Analyses were performed for formalin-fixed agar-based cell blocks and alcohol-fixed Cellient cell blocks separately. Abbreviations: CI = confidence interval; κ = kappa; LDT = laboratory-developed test; NPA = negative percent agreement; OPA = overall percent agreement; PD-L1 = programmed death-ligand 1; PPA = positive percent agreement; TPS = tumor proportion score.

Similar results were observed for the SP263 antibody (Fig. 2), with substantial agreement (OPA, 73%; κ = 0.67) observed for formalin-fixed agar cell blocks, whereas moderate agreement (OPA, 62%; κ = 0.49) was observed when analyzing agreement between alcohol-fixed Cellient cell blocks and their histologic counterparts (Table 2). A comparison of the categorization of PD-L1 expression between histology and both groups of cytologic samples showed a statistically significant difference for the Cellient-processed samples (P < .05). As with the 22C3 LDT, no statistically significant difference was found between histologic samples and agar cell blocks (P = .247) (see Supporting Table 5).

Table 2. Concordance of PD-L1 expression (TPS) between histological and cytological specimens when SP263 standardized assay was used to stain for PD-L1.

	Agreement of TPS in 3 categories (<1%;1-49%; ≥50%)		Agreement when dichotomizing TPS at 1% cutoff				Agreement when dichotomizing TPS at 50% cutoff			
	OPA (%)	Weighted κ (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's κ (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's κ (95% CI)
Agar (formalin-fixed)	73%	0.67 (0.47-0.87)	85%	83%	90%	0.67 (0.40-0.94)	88%	67%	96%	0.67 (0.37-0.98)
Cellient (alcohol-fixed)	62%	0.49 (0.24-0.73)	68%	47%	93%	0.38 (0.11-0.65)	88%	56%	100%	0.65 (0.33-0.97)

Analyses were performed for formalin-fixed agar-based cell blocks and alcohol-fixed Cellient cell blocks separately. Abbreviations: CI = confidence interval; κ = kappa; LDT = laboratory-developed test; NPA = negative percent agreement; OPA = overall percent agreement; PD-L1 = programmed death-ligand 1; PPA = positive percent agreement; TPS = tumor proportion score.

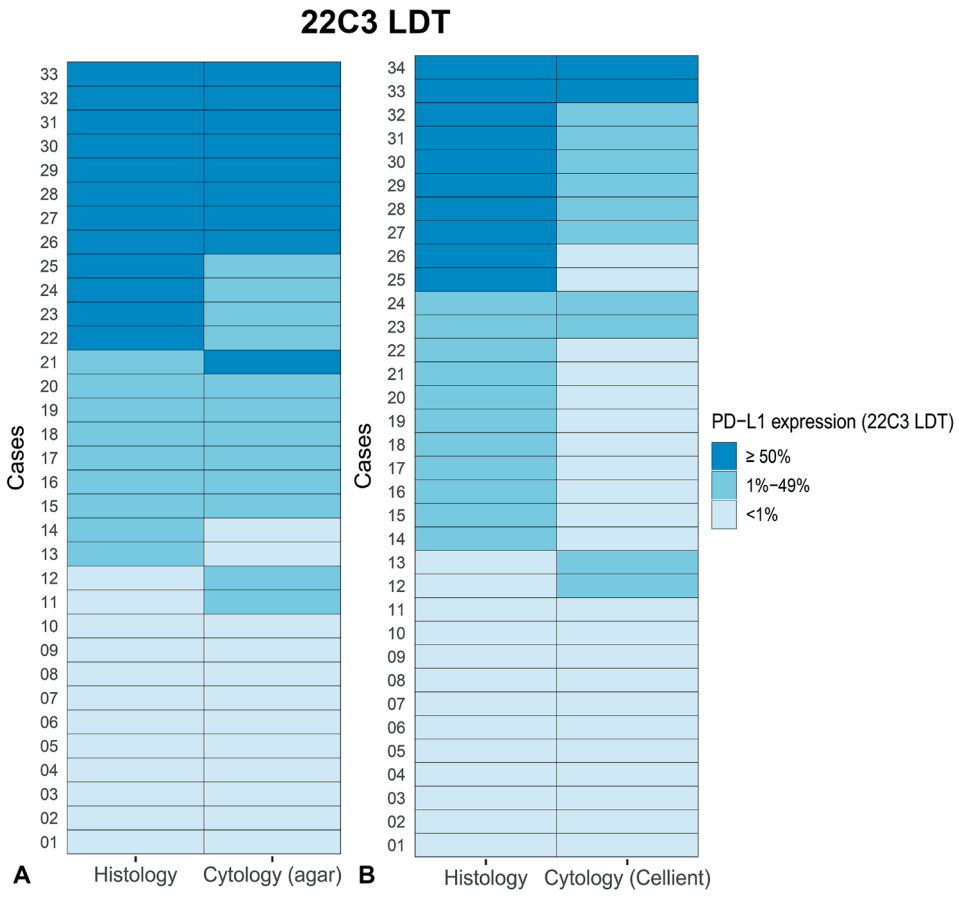


Figure 2. [Legend and figure continued on the next page]

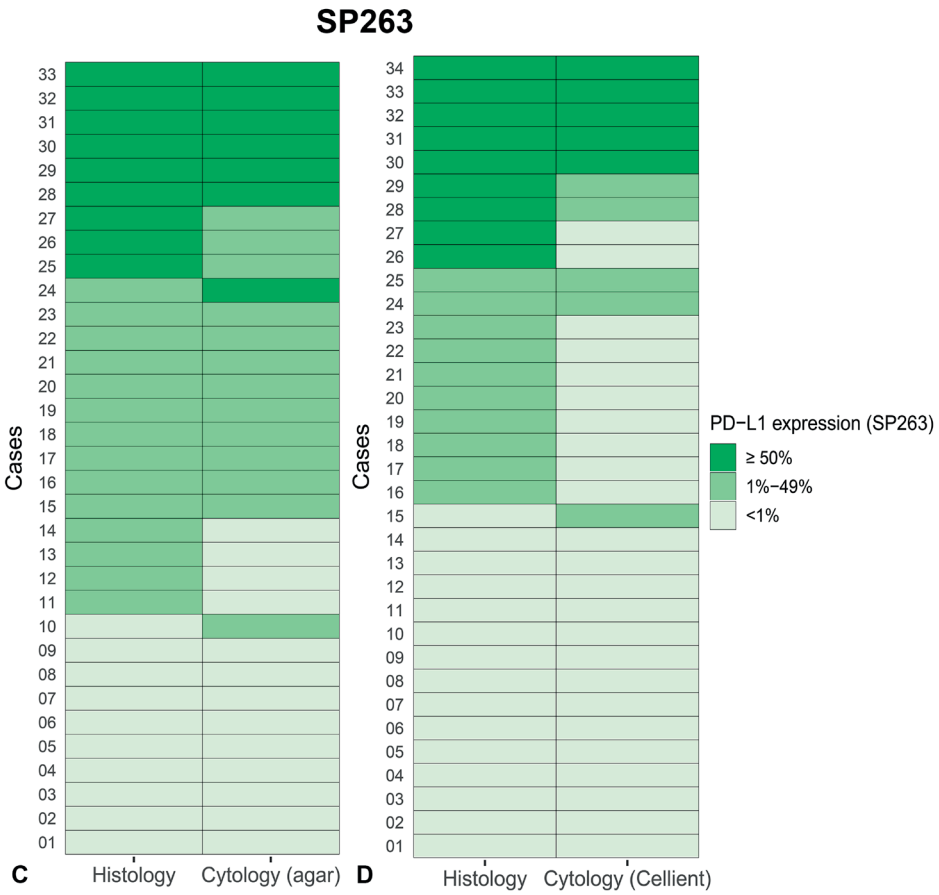


Figure 2. The concordance of programmed death-ligand 1 (PD-L1) expression (tumor proportion score) is illustrated between matched histologic and cytologic samples from individual patients. (**A**) Histology and formalin-fixed, agar-based cell blocks stained with 22C3 (the laboratory-developed test [LDT]) are compared. Nine of 33 cases (27%) show discordance. (**B**) Histology and alcohol-fixed Cellient cell blocks stained using the 22C3 LDT are compared. Nineteen of 34 cases (56%) show discordance. (**C**) Histology and formalin-fixed, agar-based cell blocks stained with SP263 (the standardized assay) are compared. Nine of 33 cases (27%) show discordance. (**D**) Histology and alcohol-fixed Cellient cell blocks stained with SP263 are compared. Thirteen of 34 cases (38%) show discordance.

Analyzing the data after dichotomization at the 1% and 50% cutoffs again resulted in lower concordance values for alcohol-fixed Cellient samples compared with formalin-fixed, agar-based samples (Tables 1 and 2). This applied to both antibodies, although concordance levels for the Cellient samples were lower for the 22C3 LDT ($\kappa = 0.28$ and $\kappa = 0.26$ for the 1% and 50% cutoffs, respectively) than for the SP263 antibody ($\kappa = 0.38$ and $\kappa = 0.65$, respectively). The lowest agreement levels were observed for Cellient cell blocks at the 50% cutoff and using the 22C3 antibody. Figure 3 displays an exemplary case in which the

surgical resection specimen had a TPS ≥50% with both antibodies, whereas the Cellient samples had a TPS between 1% and 49%.

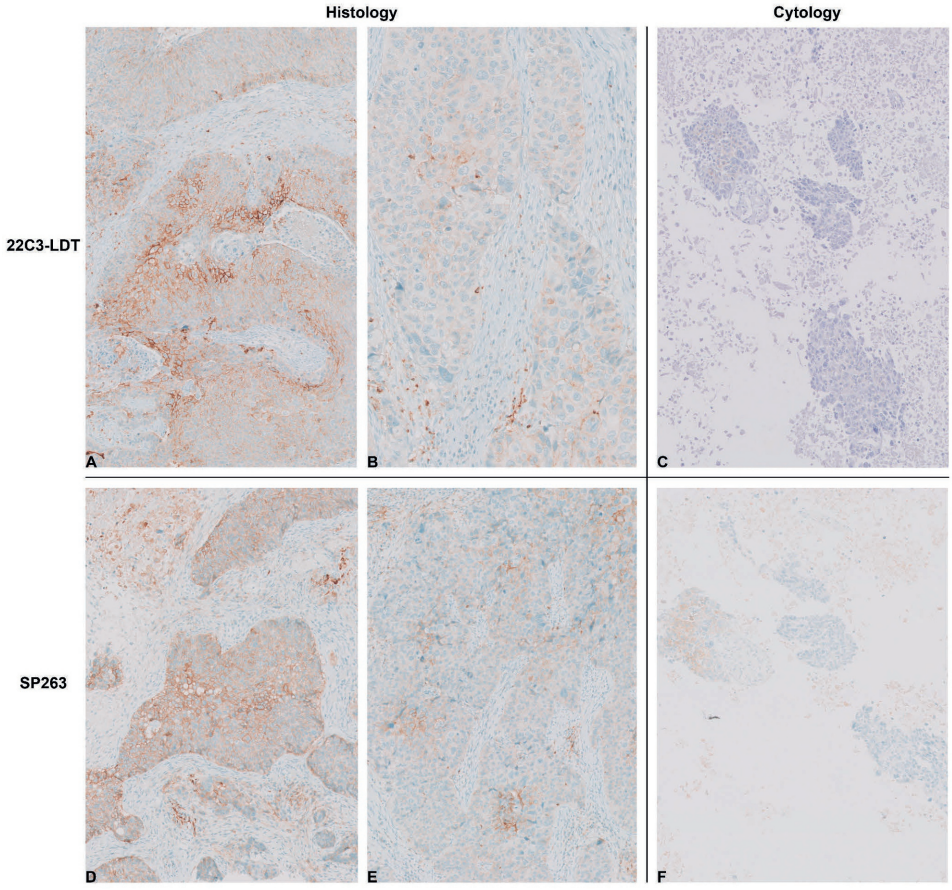


Figure 3. This is an example of a case in which the programmed death-ligand 1 (PD-L1) tumor proportion score (TPS) was ≥50% in the resection specimen and ranged from 1% to 49% in the cytologic material (alcohol-fixed Cellient cell block). (**A,B**) The PD-L1 TPS was ≥50% in a resection specimen stained with 22C3 (the laboratory-developed test [LDT]), with (**A**) part of the tumor showing strong staining intensity and (**B**) another part showing somewhat weaker staining intensity. (**C**) The PD-L1 TPS was between 1% and 49% in a Cellient cell block stained with 22C3. (**D,E**) The PD-L1 TPS was ≥50% in a resection specimen stained with SP263 (the standardized assay), with (**D**) part of the tumor showing strong staining intensity and (**E**) another part showing somewhat weaker staining intensity. (**F**) The PD-L1 TPS was between 1% and 49% in a Cellient cell block stained with SP263 (original magnification ×10 in **A-F**).

Interpathologist and Interassay Agreement

Agreement of PD-L1 scoring between the 2 pathologists was high for histologic samples with the use of both 22C3 LDT and SP263. An analysis of concordance of PD-L1 staining between both antibodies also showed high agreement, especially in histologic material (for more detailed results of interpathologist agreement analysis, see Supporting Results A and Supporting Table 6; for more detailed results of interassay agreement analysis, see Supporting Results B and Supporting Table 7).

PD-L1–Expressing Cell Lines

Twenty PD-L1–expressing cell line specimens were created using different fixatives and various fixation times. Formalin-fixed cell lines, which were used as controls, showed clear membranous staining on all cells with all staining protocols (SP263, 22C3 LDT, and 22C3 pharmDx). Fixation in CytoLyt, PreservCyt, or Carbowax resulted in lower staining intensity compared with fixation in formalin with all antibodies. Longer fixation times (24 and 48 hours, the latter for CytoLyt only) resulted in even less immunoreactivity compared with 2-hour fixation, which was most clearly visible for SP263; whereas, for the 22C3 LDT, this phenomenon was observed in cells fixed in Carbowax only. When the 22C3 pharmDx assay was used, no apparent difference between the different fixation times was discernible. Fixation in CytoRich Red did not result in lower staining intensity compared with formalin, regardless of the antibody or the fixation time used (Fig. 4).

When fixation in CytoLyt, PreservCyt, or Carbowax was followed by fixation in formalin, staining intensity was preserved using the 22C3 LDT (Fig. 5). This effect was apparent irrespective of fixation time. Similarly, using the 22C3 pharmDx assay, most of the cell lines that were fixed in formalin after alcohol fixation showed a stronger staining intensity than was observed in the cell lines without formalin fixation (see Supporting Fig. 1). This effect was most prominent in cell lines that were fixed for a total duration of 2 hours, although an improvement in staining intensity was also discernible in cell lines fixed in CytoLyt or PreservCyt for 23.5 hours followed by formalin fixation. IHC staining with SP263 showed a similar effect in the cell line fixed in Carbowax for 1.5 hours followed by 0.5 hours of formalin fixation only. SP263 showed no or only negligible preserved staining intensity after formalin fixation in the other cell lines (Fig. 6). Quantification of the difference in PD-L1 staining intensity between the cell lines with and without additional formalin fixation confirmed the beneficial effect of formalin post-fixation on the 22C3 LDT (see Supporting Fig. 2).

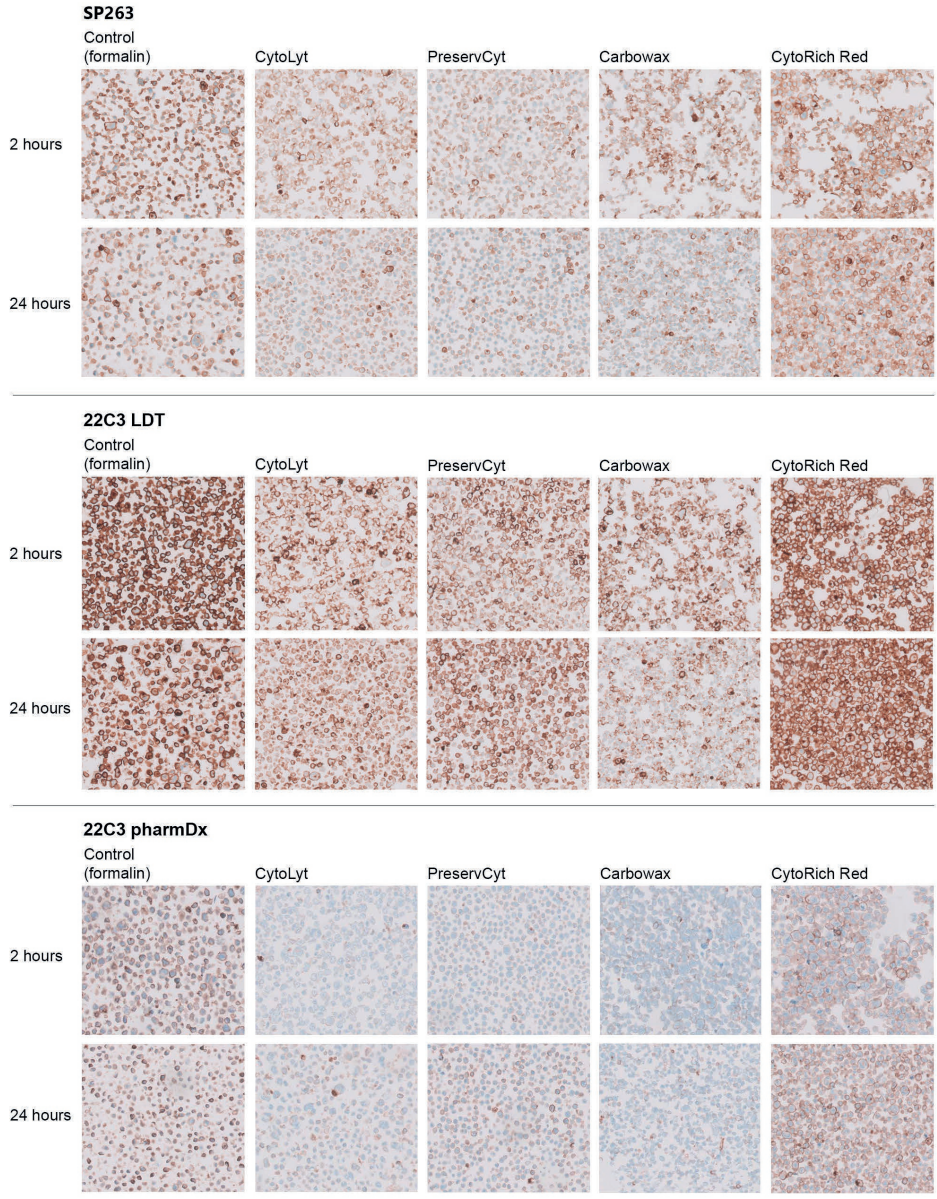


Figure 4. Immunostaining patterns are illustrated of programmed death-ligand 1 (PD-L1)–expressing cell lines fixed with different fixation schemes. Results are shown for cell lines fixed in formalin (control) and in CytoLyt, PreservCyt, Carbowax, and CytoRich Red for 2 and 24 hours. PD-L1 immunostaining was performed with (*Top*) the standardized assay SP263, (*Middle*) the 22C3 laboratory-developed test (LDT), and (*Bottom*) the 22C3 pharmDx assay.

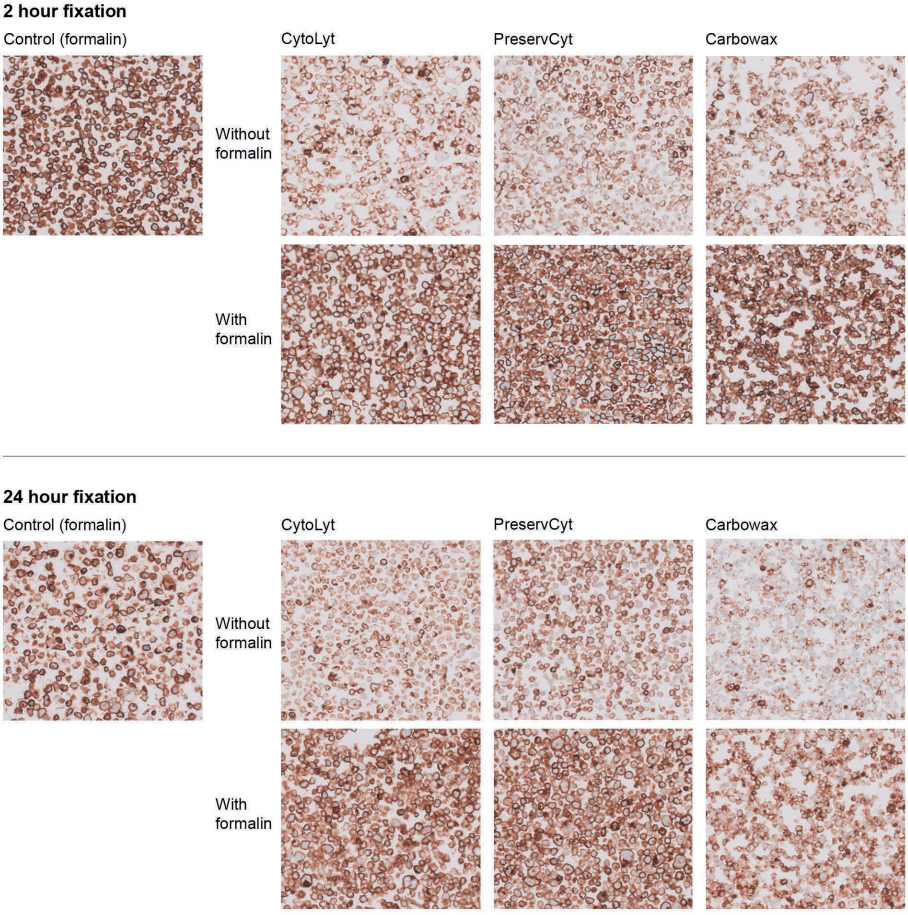


Figure 5. Immunostaining patterns of programmed death-ligand 1 (PD-L1)–expressing cell lines fixed in CytoLyt, PreservCyt, or Carbowax, either with or without additional formalin fixation for 0.5 hours, are illustrated using the 22C3 antibody (the laboratory-developed test [LDT]). Results are shown for a total fixation duration of (*Top*) 2 hours and (*Bottom*) 24 hours.

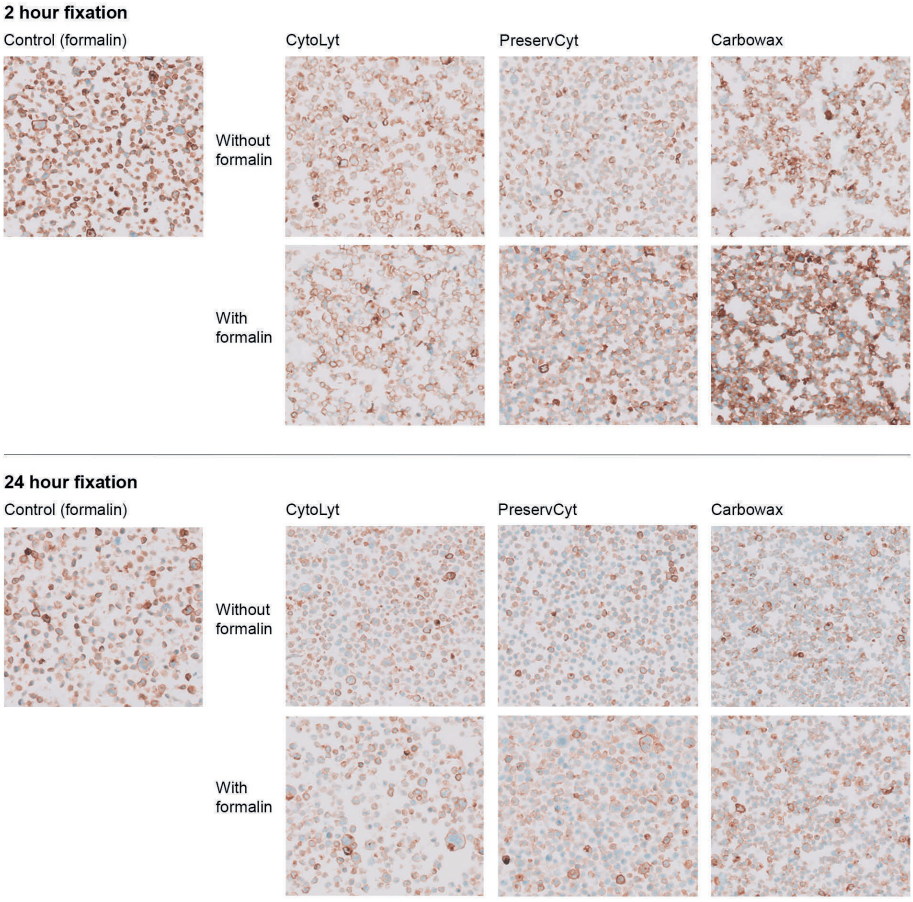


Figure 6. Immunostaining patterns of programmed death-ligand 1 (PD-L1)–expressing cell lines fixed in CytoLyt, PreservCyt, or Carbowax, either with or without additional formalin fixation for 0.5 hours, are illustrated using the SP263 antibody (the standardized assay). Results are shown for a total fixation duration of (*Top*) 2 hours and (*Bottom*) 24 hours.

DISCUSSION

In this multicenter study, concordance of PD-L1 immunostaining between matched cell blocks and histologic FFPE tissue was investigated using 2 different PD-L1 IHC antibodies (SP263 and 22C3 LDT). First, differences were observed between alcohol-fixed Cellient cell blocks and formalin-fixed agar-based cell blocks, with the Cellient material showing a clear decrease in membranous PD-L1 staining. Second, the effect of different fixatives on PD-L1 immunostaining was studied by analyzing PD-L1–expressing cell lines fixed in formalin, CytoLyt, PreservCyt, CytoRich Red, and Carbowax.

The overall concordance was moderate to substantial between histologic and cytologic specimens from the same tumor. When agar and Cellient cell blocks were analyzed separately, however, concordance levels were much higher for agar cell blocks than for Cellient-processed material. Therefore, it seems likely that the use of different modes of processing cytology results in variations in PD-L1 immunostaining and thus different levels of concordance with validated histologic PD-L1 protocols. We hypothesized that the observed differences between the 2 types of cell blocks could be explained by the use of alternative fixatives, resulting in various levels of PD-L1 immunoreactivity. Notably, the agar-based cytologic samples were fixed in formalin, whereas the Cellient samples were fixed in CytoLyt and PreservCyt, both of which are methanol-based fixatives. Other studies have shown a detrimental effect of CytoLyt fixation on IHC detection of some antigens, such as thyroid transcription factor-1 (TTF-1)[18] and Ki-67[21]. A study by Lloyd et al.[30] demonstrated a similar effect on PD-L1 staining, with CytoLyt fixation showing poor PD-L1 immunostaining results. Gosney at al.[31] observed no effect of alcohol-based fixatives, including CytoLyt, on PD-L1 immunostaining with the 22C3 pharmDx assay in cytology specimens. However, the number of cases that were fixed in CytoLyt was low. Most cases were fixed in CytoRich Red, the use of which, as discussed below, did not result in diminished PD-L1 immunostaining in our PD-L1–expressing cell lines either. Furthermore, all specimens had a post-fixation step with 10% NBF for a minimum of 45 minutes, likely resulting in the preservation of PD-L1 immunoreactivity. As indicated by our results, this preserving effect is especially prominent for the extracellular 22C3 antibody.

Diminished PD-L1 immunoreactivity caused by alcohol fixation holds significant implications for clinical practice. Although it is desirable to optimize all individual antibody staining protocols for alcohol-fixed material, perfect concordance with staining intensity in histology cannot always be achieved. Because pathologists frequently use panels of markers to make a diagnosis, a slightly lower immunocytochemical staining intensity of a single marker or a few diagnostic markers can be accepted. In predictive ICC, however, pathologists mostly rely on a single marker to allow clinicians to make a treatment decision. A low membranous staining intensity in histology might result in a false-negative result in cytology. Therefore, using PD-L1 ICC could lead to wrongful denial of treatment with durvalumab to patients with unresectable, stage III NSCLC in Europe (prescribed only to patients with a TPS ≥1%)[4] and of treatment with pembrolizumab as first-line monotherapy to patients with metastasized NSCLC (prescribed only to patients with a TPS ≥50%)[2]. Because many PD-L1 tests for patients with advanced-stage NSCLC are performed on cytologic samples, it is of the utmost importance that the methods used to process cytologic material do not negatively affect PD-L1 staining.

Our analysis of PD-L1–expressing cell lines fixed in different fixatives supports the hypothesis that the use of fixatives other than formalin could result in less PD-L1 immunoreactivity. Cell

lines fixed in methanol-based CytoLyt or PreservCyt and cell lines fixed in ethanol-based Carbowax exhibited lower staining intensity with the SP263 and 22C3 standardized assays and the 22C3 LDT. It is known that formalin fixation and alcohol fixation have different effects on proteins, leading to different alterations of the 3-dimensional protein structure[32, 33]. This might explain differences in staining results. Many IHC antibodies have primarily been developed for use on FFPE samples, thus targeting epitopes after formalin fixation. These epitopes, however, may not remain good targets after alcohol fixation. Notably, fixation in CytoRich Red did not result in lower immunostaining, although this fixative contains alcohol elements as well. In contrast to CytoLyt, PreservCyt, and Carbowax, CytoRich Red contains a small amount of formaldehyde (<1%), possibly explaining the preserved PD-L1 immunostaining[34]. A side effect of CytoRich Red, however, is DNA degradation[34], rendering the solution less appropriate for routine FNA practice in the selected population, in which treatment decisions rely on adequate predictive molecular analysis as well.

Interestingly, the addition of formalin to cell lines fixed in CytoLyt, PreservCyt, or Carbowax revealed a positive effect on PD-L1 immunostaining using both the 22C3 LDT and the 22C3 pharmDx assay. Similar to these results, a study by Torous et al.[35] showed no significant difference in PD-L1 categorization (TPS <1%, TPS 1%-49%, and TPS ≥50%) between cell blocks fixed in CytoLyt followed by formalin fixation and FFPE surgical resection specimens using the 22C3 pharmDx assay to stain for PD-L1. Although that study did not use paired cytology and resection specimens, the results suggest that additional formalin fixation could be helpful in preserving PD-L1 staining intensity in cytology specimens that were fixed in an alcohol-based fixative. In our study, however, we did not observe improved PD-L1 staining results in most of the cell lines post-fixed in formalin when stained with the SP263 antibody. It is known that fixation effects may differ between different antibodies that target the same protein, which also was demonstrated in a study by Buonocore et al.[21]. These differences might be explained by the finding that different antibodies targeting the same protein normally target different epitopes[33]. In the case of PD-L1 IHC, the SP263 antibody binds to an epitope in the cytoplasmic domain of PD-L1, whereas the 22C3 antibody binds to the extracellular domain of PD-L1[36]. It has been suggested before that this may lead to different tumor cell staining results[37], rendering it plausible that variation in epitopes might also result in different effects of formalin post-fixation between antibodies. Perhaps other differences between epitopes, such as the degree of glycosylation, which has been shown to affect the accuracy of IHC staining[38], also could lead to variation in fixation effects between antibodies. It is noteworthy, however, that we only applied 30 minutes of additional formalin fixation in our cell lines, so we cannot exclude the possibility that longer formalin fixation times also might result in a positive effect on PD-L1 immunostaining using the SP263 antibody. In any case, the positive effects of formalin post-fixation observed with the 22C3 antibody seem promising. It would be worthwhile to investigate whether additional fixation in formalin after alcohol fixation also

results in preserved PD-L1 staining intensity in patient samples using various PD-L1 IHC antibodies and varying fixation times.

Various studies have proposed intratumoral heterogeneity of PD-L1 expression as the cause of disagreement in PD-L1 immunostaining between histology and cytology[9, 13, 16]. In our study, it is possible that PD-L1 intratumoral heterogeneity also played a part in creating discordance between surgical resection specimens and cytology cell blocks, especially in the comparison of histology and agar-based cell blocks. It seems very unlikely, however, that discordance between histology and Cellient cell blocks can be attributed completely to PD-L1 intratumoral heterogeneity because the discordance in these cell blocks was much more pronounced compared with that in the agar cell blocks, and most of the discordant cases (89% and 92% for 22C3LDT and SP263, respectively) showed lower PD-L1 immunostaining on cytology compared with histology. Combined with results from the analysis of our PD-L1–expressing cell lines fixed in different fixatives, a decrease in PD-L1 immunostaining caused by alcohol fixation seems the most likely explanation for the observed discordance.

It should be mentioned that the negative effects of alcohol-based fixatives on PD-L1 immunoreactivity were observed using PD-L1 IHC protocols that were validated for use on FFPE tissue. Altering the immunostaining protocol to optimize its use on cytology specimens fixed in alcohol might result in better staining results. This has been demonstrated before with other IHC antibodies[20], although changing IHC protocols for use on cytologic specimens does not always result in improved immunostaining[20, 21]. Notably, it has also been described that pathologists often use techniques designed for histology on cytologic specimens without considering the differences in specimen preparation and how these differences could affect the interpretation of immunostaining[39]. The use of PD-L1 IHC protocols, validated for FFPE tissue, on cytology specimens fixed in alcohol-based fixatives could result in less PD-L1 immunoreactivity and might result in false-negative results.

In conclusion, our study demonstrates the negative effects of methanol-based and ethanol-based fixatives on PD-L1 immunostaining using the SP263 standardized assay, the 22C3 pharmDx assay, and a 22C3 LDT, each validated for use on FFPE tissue specimens. Therefore, if cytologic specimens need to be used for ICC testing of PD-L1 expression, the fixative of choice is formalin. Methanol-based or ethanol-based fixatives should be avoided unless they are used with a meticulously validated ICC protocol that has been designed specifically for use on specimens fixed in these fixatives. Post-fixation in formalin potentially may reverse the negative effect of alcohol fixation to some degree. This preserving effect should be studied in patient samples using various PD-L1 antibodies to assess the applicability of additional formalin fixation after alcohol fixation in clinical practice.

ACKNOWLEDGEMENTS

This study was supported by an unrestricted research grant from Roche Diagnostics Nederland BV, received by Nils A. ’t Hart. The funding source was not involved in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

REFERENCES

1. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med. 2016;375:1823-1833.
2. Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(suppl 4):iv192-iv237.
3. European Society for Medical Oncology (ESMO). Metastatic Non-Small-Cell Lung Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. ESMO; 2019. Available from: https://www.esmo.org/content/download/227453/3874538/1/ESMO-CPG-mNSCLC-18SEPT2019.pdf (Accessed 10 January 2020).
4. European Medicines Agency (EMA). Imfinzi (durvalumab). EMA; 2018. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/imfinzi (Accessed 5 April 2019).
5. Jorgensen JT. Companion and complementary diagnostics: clinical and regulatory perspectives. Trends Cancer. 2016;2:706-712.
6. Scheerens H, Malong A, Bassett K, et al. Current status of companion and complementary diagnostics: strategic considerations for development and launch. Clin Transl Sci. 2017;10:84-92.
7. Teng MWL, Ngiow SF, Ribas A, Smyth MJ. Classifying cancers based on T cell infiltration and PD-L1. Cancer Res. 2015;75:2139-2145.
8. Russell-Goldman E, Kravets S, Dahlberg SE, Sholl LM, Vivero M. Cytologic-histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. Cancer Cytopathol. 2018;126:253-263.
9. Skov BG, Skov T. Paired comparison of PD-L1 expression on cytologic and histologic specimens from malignancies in the lung assessed with PD-L1 IHC 28-8pharmDx and PD-L1 IHC 22C3pharmDx. Appl Immunohistochem Mol Morphol. 2017;25:453-459.
10. Skov BG, Hogdall E, Clementsen P, et al. The prevalence of EGFR mutations in non-small cell lung cancer in an unselected Caucasian population. APMIS. 2015;123:108-115.
11. Roh MH. The utilization of cytologic fine-needle aspirates of lung cancer for molecular diagnostic testing. J Pathol Transl Med. 2015;49:300-309.
12. Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death-ligand 1 antibody to determine programmed death-ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. Cancer Cytopathol. 2018;126:264-274.
13. Noll B, Wang WL, Gong Y, et al. Programmed death ligand 1 testing in non-small cell lung carcinoma cytology cell block and aspirate smear preparations. Cancer Cytopathol. 2018;126:342-352.
14. Sakakibara R, Inamura K, Tambo Y, et al. EBUS-TBNA as a promising method for the evaluation of tumor PD-L1 expression in lung cancer. Clin Lung Cancer. 2017;18:527-534.e1.
15. Wang H, Agulnik J, Kasymjanova G, et al. Cytology cell blocks are suitable for immunohistochemical testing for PD-L1 in lung cancer. Ann Oncol. 2018;29:1417-1422.
16. Hernandez A, Brandler TC, Zhou F, Moreira AL, Schatz-Siemers N, Simsir A. Assessment of programmed death-ligand 1 (PD-L1) immunohistochemical expression on cytology specimens in non-small cell lung carcinoma. Am J Clin Pathol. 2019;151:403-415.
17. Nambirajan A, Jain D. Cell blocks in cytopathology: an update. Cytopathology. 2018;29:505-524.

18. Gruchy JR, Barnes PJ, Dakin Hache KA. CytoLyt® fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. Appl Immunohistochem Mol Morphol. 2015;23:297-302.
19. Nietner T, Jarutat T, Mertens A. Systematic comparison of tissue fixation with alternative fixatives to conventional tissue fixation with buffered formalin in a xenograft-based model. Virchows Arch. 2012;461:259-269.
20. Sauter JL, Grogg KL, Vrana JA, Law ME, Halvorson JL, Henry MR. Young Investigator Challenge: validation and optimization of immunohistochemistry protocols for use on Cellient cell block specimens. Cancer Cytopathol. 2016;124:89-100.
21. Buonocore DJ, Konno F, Jungbluth AA, et al. CytoLyt fixation significantly inhibits MIB1 immunoreactivity whereas alternative Ki-67 clone 30-9 is not susceptible to the inhibition: critical diagnostic implications. Cancer Cytopathol. 2019;127:643-649.
22. T Hart NA, van der Starre-Gaal J, Vonk JM, Timens W. Essential preanalytics in PD-L1 immunocytochemistry. Histopathology. 2019;74:362-364.
23. Federation of Dutch Medical Scientific Societies (FEDERA). Human Tissue and Medical Research: Code of Conduct for Responsible Use. FEDERA; 2011. Available from: https://www.federa.org/sites/default/files/digital_version_first_part_code_of_conduct_in_uk_2011_12092012.pdf (Accessed 15 January 2019).
24. HistoCyte Laboratories. PD-L1 Analyte Control Datasheet. Available from: http://www.histocyte.com/images/pdf/HCL0019_HCL0020_HCL0021_PI_V003.pdf (Accessed 16 September 2019).
25. Adam J, Le Stang N, Rouquette I, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. Ann Oncol. 2018;29:953-958.
26. Visiopharm. PD-L1, Cell Lines, Cancer. Available at: https://visiopharm.com/app-center/app/pd-l1-cell-lines-cancer/ (Accessed 25 September 2020).
27. John T, Liu G, Tsao MS. Overview of molecular testing in non-small-cell lung cancer: mutational analysis, gene copy number, protein expression and other biomarkers of EGFR for the prediction of response to tyrosine kinase inhibitors. Oncogene. 2009;28(suppl 1):S14-S23.
28. Smith J, Robida MD, Acosta K, et al. Quantitative and qualitative characterization of two PD-L1 clones: SP263 and E1L3N. Diagn Pathol. 2016;11:44.
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-174.
30. Lloyd IE, Zhou W, Witt BL, Chadwick BE. Characterization of PD-L1 immunohistochemical expression in cell blocks with different specimen fixation and processing methods. Appl Immunohistochem Mol Morphol. 2019;27:107-113.
31. Gosney JR, Haragan A, Chadwick C, et al. Programmed death ligand 1 expression in EBUS aspirates of non-small cell lung cancer: is interpretation affected by type of fixation? Cancer Cytopathol. 2020;128:100-106.
32. Miller RT, Swanson PE, Wick MR. Fixation and epitope retrieval in diagnostic immunohistochemistry: a concise review with practical considerations. Appl Immunohistochem Mol Morphol. 2000;8:228-235.
33. Ramos-Vara JA. Technical aspects of immunohistochemistry. Vet Pathol. 2005;42:405-426.
34. Dejmek A, Zendehrokh N, Tomaszewska M, Edsjo A. Preparation of DNA from cytological material: effects of fixation, staining, and mounting medium on DNA yield and quality. Cancer Cytopathol. 2013;121:344-353.
35. Torous VF, Rangachari D, Gallant BP, Shea M, Costa DB, VanderLaan PA. PD-L1 testing using the clone 22C3 pharmDx kit for selection of patients with non-small cell lung cancer to receive immune checkpoint inhibitor

therapy: are cytology cell blocks a viable option? J Am Soc Cytopathol. 2018;7:133-141.

36. Lawson NL, Dix CI, Scorer PW, et al. Mapping the binding sites of antibodies utilized in programmed cell death ligand-1 predictive immunohistochemical assays for use with immuno-oncology therapies. Mod Pathol. 2020;33:518-530.
37. Jotatsu T, Oda K, Yatera K. PD-L1 immunohistochemistry in patients with non-small cell lung cancer. J Thorac Dis. 2018;10(Suppl 18):S2127-S2129.
38. Lee HH, Wang YN, Xia W, et al. Removal of N-linked glycosylation enhances PD-L1 detection and predicts anti-PD-1/PD-L1 therapeutic efficacy. Cancer Cell. 2019;36:168-178.e4.
39. Fowler LJ, Lachar WA. Application of immunohistochemistry to cytology. Arch Pathol Lab Med. 2008;132:373-383.

SUPPORTING INFORMATION

Supporting Table 1. Overview of methods for fixation of FNA material and generation of cell blocks used by each participating pathology laboratory.

Laboratory number	Number of cases in tudy	Collection fluid; fixation time	Second fixative; fixation time	Cell block method
1	8	CytoLyt*; 20 minutes	PreservCyt*; ≥10 minutes	Cellient Automated Cell Block System
2	26	CytoLyt; 30 minutes	PreservCyt; ≥10 minutes	Cellient Automated Cell Block System
3	14	Carbowax**; 30 minutes	Formalin; ≥120 minutes	Agar; processed into FFPE cell block
4	11	Unifix***; 30-120 minutes	No second fixative	Agar; processed into FFPE cell block
5	8	Formalin; ≥30 minutes	No second fixative	Agar; processed into FFPE cell block

* CytoLyt and PreservCyt contain methanol.
** Carbowax contains ethanol.
*** Unifix contains 4% formaldehyde, no alcohol.
Abbreviations: FFPE = formalin-fixed paraffin-embedded; FNA = fine needle aspiration.

Supporting Table 2. Overview of 20 cell lines with high PD-L1 expression created with different fixatives and fixation times.

Fixative 1	Fixation time fixative 1 (hours)	Fixative 2	Fixation time fixative 2 (hours)	Total fixation time (hours)
Formalin	2	-	-	2
CytoLyt	2	-	-	2
CytoLyt	1.5	Formalin	0.5	2
PreservCyt	2	-	-	2
PreservCyt	1.5	Formalin	0.5	2
CytoRich Red	2	-	-	2
CytoRich Red	1.5	Formalin	0.5	2
Carbowax	2	-	-	2
Carbowax	1.5	Formalin	0.5	2
Formalin	24	-	-	24
CytoLyt	24	-	-	24
CytoLyt	23.5	Formalin	0.5	24
PreservCyt	24	-	-	24
PreservCyt	23.5	Formalin	0.5	24
CytoRich Red	24	-	-	24
CytoRich Red	23.5	Formalin	0.5	24
Carbowax	24	-	-	24
Carbowax	23.5	Formalin	0.5	24
CytoLyt	48	-	-	48
Formalin	48	-	-	48

Supporting Table 3. Concordance of PD-L1 expression (TPS) between histological and all cytological material (both formalin-fixed agar cell blocks and alcohol-fixed Cellient cell blocks).

	Agreement of TPS in 3 categories (<1%; 1-49%; ≥50%)		**Agreement when dichotomizing TPS at 1% cutoff**				**Agreement when dichotomizing TPS at 50% cutoff**			
	OPA (%)	Weighted κ (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's κ (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's κ (95% CI)
22C3 LDT	58%	0.49 (0.34-0.65)	75%	69%	84%	0.49 (0.29-0.70)	81%	45%	98%	0.50 (0.27-0.72)
SP263	67%	0.59 (0.43-0.74)	76%	67%	92%	0.53 (0.34-0.73)	88%	61%	98%	0.66 (0.44-0.88)

Abbreviations: CI = confidence interval; κ = kappa; LDT = laboratory-developed test; NPA = negative percent agreement; OPA = overall percent agreement; PD-L1 = programmed death-ligand 1; PPA = positive percent agreement; TPS = tumor proportion score.

Supporting Table 4. Cross tabulation of PD-L1 expression (TPS in categories) of paired histological and cytological samples (formalin-fixed/agar-based and alcohol-fixed/Cellient processed), when 22C3 LDT was used to stain for PD-L1.

		a. Cytology (agar)				**b. Cytology (Cellient)**			
		<1%	1-49%	≥50%	Total	<1%	1-49%	≥50%	Total
Histology	<1%	10	2	0	12	11	2	0	13
	1-49%	2	6	1	9	9	2	0	11
	≥50%	0	4	8	12	2	6	2	10
Total		12	12	9	33	22	10	2	34
P-value		0.407				*<0.01*			

Numbers of concordant and discordant samples are displayed for formalin-fixed agar-based cell blocks (**a**) and alcohol-fixed Cellient cell blocks (**b**), both compared to histology. PD-L1 IHC was performed using 22C3 LDT. Differences in the categorization of PD-L1 expression between cytological and histological samples were analyzed by performing the McNemar-Bowker test of symmetry.

Supporting Table 5. Cross tabulation of PD-L1 expression (TPS in categories) of paired histological and cytological samples (formalin-fixed/agar-based and alcohol-fixed/Cellient processed), when SP263 standardized assay was used to stain for PD-L1.

		a. Cytology (agar)				**b. Cytology (Cellient)**			
		<1%	1-49%	≥50%	Total	<1%	1-49%	≥50%	Total
Histology	<1%	9	1	0	10	14	1	0	15
	1-49%	4	9	1	14	8	2	0	10
	≥50%	0	3	6	9	2	2	5	9
Total		13	13	7	33	24	5	5	34
P-value		0.247				*<0.05*			

Numbers of concordant and discordant samples are displayed for formalin-fixed agar-based cell blocks (**a**) and alcohol-fixed Cellient cell blocks (**b**), both compared to histology. PD-L1 IHC was performed using SP263 standardized assay. Differences in the categorization of PD-L1 expression between cytological and histological samples were analyzed by performing the McNemar-Bowker test of symmetry.

Supporting Results A

Interpathologist agreement of PD-L1 scoring

Agreement between the two pathologists was high for histological samples with use of both 22C3 LDT and SP263 (κ 0.87 and 0.83). Compared to histological samples, interpathologist agreement was lower in cytological specimens for both antibodies (Supporting Table 6).

Supporting table 6. Interpathologist agreement of PD-L1 scoring (two observers).

		Agreement of TPS in 3 categories (<1%; 1-49%; ≥50%)		Agreement when dichotomizing TPS at 1% cutoff		Agreement when dichotomizing TPS at 50% cutoff	
		OPA (%)	Weighted κ (95% CI)	OPA (%)	Cohen's κ (95% CI)	OPA (%)	Cohen's κ (95% CI)
22C3 LDT	**Histology**	89%	0.87 (0.77-0.97)	96%	0.90 (0.79-1.00)	93%	0.83 (0.68-0.98)
	Cytology	79%	0.74 (0.61-0.86)	87%	0.74 (0.58-0.89)	93%	0.74 (0.52-0.96)
SP263	**Histology**	85%	0.83 (0.73-0.93)	93%	0.84 (0.71-0.98)	93%	0.81 (0.65-0.97)
	Cytology	79%	0.74 (0.62-0.87)	88%	0.76 (0.61-0.92)	91%	0.72 (0.50-0.93)

Abbreviations: CI = confidence interval; κ = kappa; NPA = negative percent agreement; OPA = overall percent agreement; PD-L1 = programmed death-ligand 1; PPA = positive percent agreement; TPS = tumor proportion score.

Supporting Results B

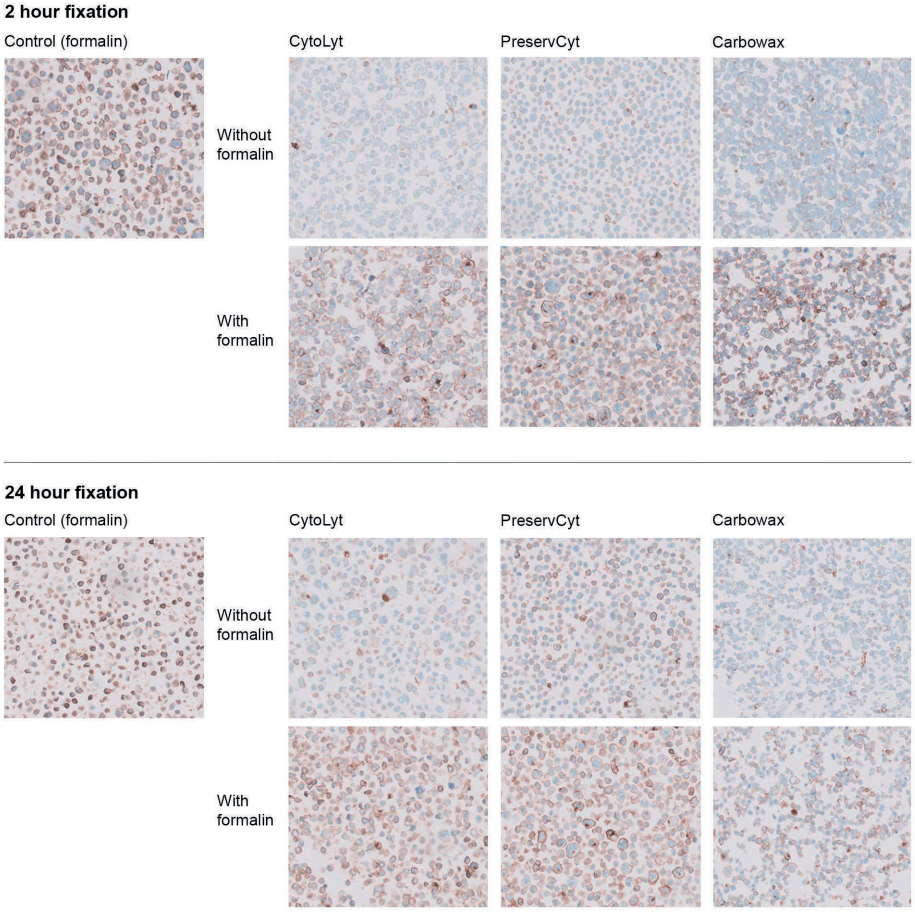
Interassay agreement of PD-L1 staining

Interassay concordance was highest in histological material (κ 0.87). Concordance levels were lower for Cellient cell blocks than for agar cell blocks (Supporting Table 7).

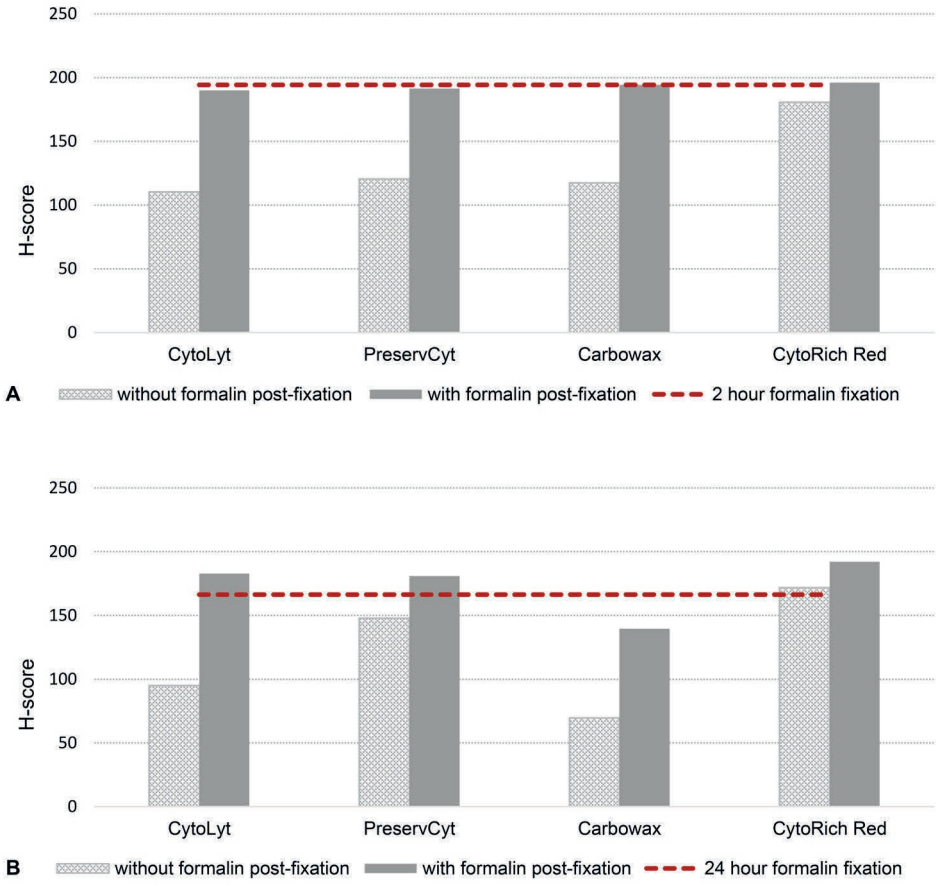
Supporting Table 7. Agreement of PD-L1 staining between 22C3 LDT and SP263 standardized assay.

	Agreement of TPS in 3 categories (<1%; 1-49%; ≥50%)		Agreement when dichotomizing TPS at 1% cutoff		Agreement when dichotomizing TPS at 50% cutoff	
	OPA (%)	Weighted κ (95% CI)	OPA (%)	Cohen's κ (95% CI)	OPA (%)	Cohen's κ (95% CI)
Histology	90%	0.87 (0.77-0.96)	94%	0.87 (0.75-1.00)	96%	0.86 (0.72-0.99)
Cytology (overall)	84%	0.77 (0.64-0.90)	90%	0.79 (0.64-0.94)	93%	0.74 (0.52-0.96)
Cytology (agar)	85%	0.82 (0.66-0.98)	91%	0.81 (0.59-1.00)	94%	0.84 (0.61-1.00)
Cytology (Cellient)	82%	0.67 (0.43-0.92)	88%	0.73 (0.48-0.98)	91%	0.53 (0.07-1.00)

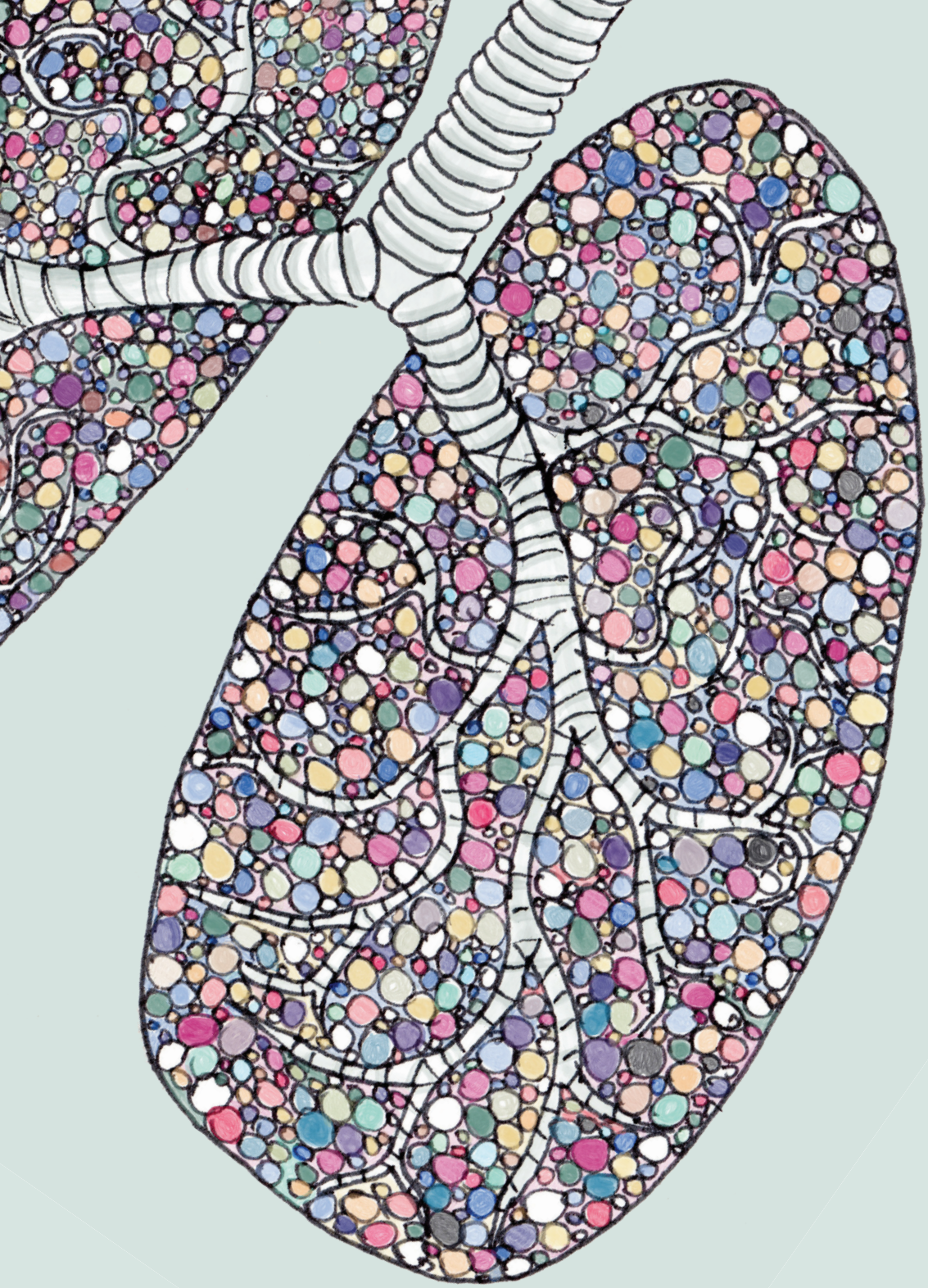
Abbreviations: CI = confidence interval; κ = kappa; NPA = negative percent agreement; OPA = overall percent agreement; PD-L1 = programmed death-ligand 1; PPA = positive percent agreement; TPS = tumor proportion score.



Supporting Figure 1. Immunostaining patterns of programmed death-ligand 1 (PD-L1)–expressing cell lines fixed in CytoLyt, PreservCyt or Carbowax, either with or without additional formalin fixation of 0.5 hours, are illustrated using the 22C3 pharmDx assay. Results are shown for a total fixation duration of (*Top*) 2 hours and (*Bottom*) 24 hours.



Supporting Figure 2. Programmed death-ligand 1 (PD-L1) expression (expressed in H-scores) in cell lines fixed in alcohol-based fixatives with and without formalin post-fixation, stained with 22C3 LDT. **A**: PD-L1 H-scores for a total fixation duration of 2 hours. The dotted line represents the H-score of a cell line fixed in formalin only. **B**: PD-L1 H-scores for a total fixation duration of 24 hours. The dotted line represents the H-score of a cell line fixed in formalin only.



4

FALSE NEGATIVE PROGRAMMED DEATH-LIGAND 1 IMMUNOSTAINING IN ETHANOL-FIXED ENDOBRONCHIAL ULTRASOUND-GUIDED TRANSBRONCHIAL NEEDLE ASPIRATION SPECIMENS OF NON-SMALL CELL LUNG CANCER PATIENTS

Koomen BM*
Vreuls W*
de Boer M
de Ruiter EJ
Hoelters J
Vink A
Willems SM

Histopathology. 2021;79(4):480-490.

* *These authors contributed equally to this work.*

ABSTRACT

Aims

Programmed death-ligand 1 (PD-L1) immunostaining is used to predict which non-small cell lung cancer (NSCLC) patients will respond best to treatment with programmed death protein-1/PD-L1 inhibitors. PD-L1 immunostaining is sometimes performed on alcohol-fixed cytological specimens instead of on formalin-fixed paraffin-embedded (FFPE) biopsies or resections. We studied whether ethanol pre-fixation of clots from endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) results in diminished PD-L1 immunostaining as compared with formalin fixation.

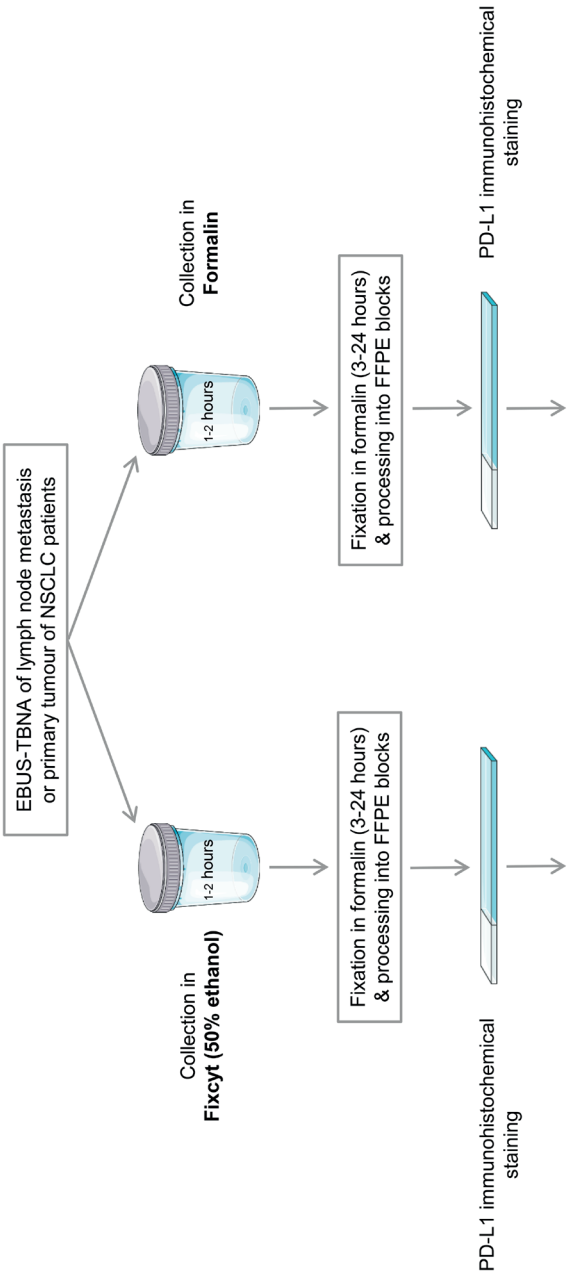
Methods and results

FFPE cell blocks from EBUS-TBNA specimens of 54 NSCLC patients were identified. For each case, paired samples were available, consisting of clots directly immersed in formalin and clots prefixed in Fixcyt (50% ethanol). Serial sections were immunostained for PD-L1 by use of the standardized SP263 assay and the 22C3 antibody as a laboratory-developed test (LDT). PD-L1 positivity was determined with two cutoffs (1% and 50%). Concordance of PD-L1 positivity between the formalin-fixed (gold standard) and ethanol-prefixed material was assessed. When the 22C3 LDT was used, 30% and 36% of the ethanol-prefixed specimens showed false-negative results at the 1% and 50% cutoffs, respectively (kappa 0.64 and 0.68). When SP263 was used, 22% of the ethanol-prefixed specimens showed false-negative results at the 1% cutoff (kappa 0.67). At the 50% cutoff, concordance was higher (kappa 0.91), with 12% of the ethanol-prefixed specimens showing false-negative results.

Conclusion

Ethanol fixation of EBUS-TBNA specimens prior to formalin fixation can result in a considerable number of false-negative PD-L1 immunostaining results when a 1% cutoff is used and immunostaining is performed with SP263 or the 22C3 LDT. The same applies to use of the 50% cutoff when immunostaining is performed with the 22C3 LDT.

Graphical abstract



False-negative PD-L1 immunostaining in ethanol pre-fixed specimens compared to formalin-fixed specimens

INTRODUCTION

Immunotherapy with checkpoint blockade has become an integral part of cancer treatment, with several immune checkpoint inhibitors having been registered for various forms of cancer[1]. In patients with advanced non-small cell lung cancer (NSCLC), inhibition of the programmed death-1 (PD-1) protein or its ligand programmed death-ligand 1 (PD-L1) may lead to clinical benefit[2-7]. Registered drugs are the PD-1 inhibitors nivolumab and pembrolizumab and the PD-L1 inhibitors atezolizumab and durvalumab. Various clinical trials have shown better clinical results or a trend for better efficacy of PD-1 and PD-L1 inhibitors in patients with higher expression of PD-L1 on tumor cells, as measured with immunohistochemistry (IHC)[2, 3, 5, 6, 8, 9]. In clinical practice, therefore, pathologists determine the percentage of tumor cells that show PD-L1 expression, also known as the tumor proportion score (TPS), which is used to predict which patients might respond best to treatment with PD-1 or PD-L1 inhibitors.

In predicting these chances of response to anti-PD-1 and anti-PD-L1 therapeutics, two cutoffs for the PD-L1 TPS are clinically relevant. A cutoff of 50% is relevant for prescription of pembrolizumab. This drug may be prescribed as first-line therapy combined with chemotherapy to patients with stage IV NSCLC (without *EGFR* or *ALK* mutations), regardless of PD-L1 expression[9, 10]. However, patients whose tumors show a TPS of ≥50% may receive first-line pembrolizumab monotherapy[9, 10], exposing them to far less treatment toxicity. Second, the 1% cutoff is used for prescription of durvalumab in various European countries, as the European Medicines Agency has advised prescription of this drug as consolidation treatment to stage III NSCLC patients whose tumors show a TPS of ≥1%[11]. The PD-1/PD-L1 inhibitors nivolumab and atezolizumab may be prescribed as second-line treatment to patients with metastatic NSCLC regardless of PD-L1 expression[9]. However, measurement of PD-L1 expression could aid in predicting the chances of response to these drugs as well, and may in that way guide clinicians in their treatment decisions for the individual patient[12].

The efficacy of PD-1/PD-L1 inhibitors in relation to PD-L1 expression was determined in clinical trials that performed IHC on formalin-fixed paraffin-embedded (FFPE) surgical biopsies or resections[13]. However, because a large proportion of NSCLC patients are diagnosed at an advanced stage of disease, diagnosis and treatment decisions are often based on cytological specimens from readily accessible sites[14]. Because of their minimally invasive character, techniques such as endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) and transthoracic fine needle aspiration are preferable in this setting[14, 15]. The material collected during these procedures can be fixed and processed into cell blocks in a great variety of ways, which may quite often not result in FFPE material[16]. As the use of PD-L1 antibodies for immunostaining on cytological

specimens fixed and processed in various ways has not been validated in clinical trials, it is unclear whether the results of PD-L1 immunostaining on these specimens can adequately predict response to PD-1/PD-L1 inhibitors.

So far, there have been several studies assessing the concordance of PD-L1 immunostaining between surgical biopsies or resections and cytological cell blocks. Most of these have concluded that PD-L1 testing can safely be performed on the latter, on the basis of high levels of concordance between histology and cell blocks[13, 17-20]. However, these studies used formalin-fixed cytological material, whereas, in clinical practice, laboratories may use alcohol (methanol or ethanol)-based fixatives or transport media[21]. A potential advantage of using alcohols such as ethanol for fixation is their beneficial effect on nucleic acid preservation as compared with formalin fixation, resulting in higher-quality material for molecular analysis[22]. However, it is known from other studies that alcohol fixation can result in decreased intensity of immunostaining for various antibodies[23-27]. Potentially, alcohol fixation could have a similar negative effect on PD-L1 immunostaining. This could have an important impact on the selection of NSCLC patients who are eligible for immunotherapy, potentially leading to denial of treatment options that these patients could benefit from. In this study, the effect of ethanol pre-fixation followed by delayed formalin fixation on PD-L1 immunostaining was determined, by the use of paired formalin-fixed and ethanol-fixed EBUS-TBNA samples from NSCLC patients.

MATERIALS AND METHODS

Patients and specimens

For this study, NSCLC patients who had undergone EBUS-TBNA on lymph nodes or primary tumors at the Canisius-Wilhelmina Hospital (CWZ) (Nijmegen, The Netherlands) between November 2015 and August 2018 were identified retrospectively. Aspirated material from all patients was split into two separate fixatives, with approximately two-thirds of the material (the yield from sampling the lymph node or primary tumor twice) being collected in 20 ml of 10% neutral buffered formalin, and one-third of the material (the yield from sampling once) being collected in 20 ml of Fixcyt (50% ethanol and 2% polyethylene glycol solution). The average pre-fixation duration was 1 h, with a maximum of 2 h. Upon arrival of the EBUS-TBNA material at the pathology laboratory, visible clots from the material fixed in Fixcyt were directly placed into formalin for further fixation, with post-fixation times ranging from 3 h to 24 h. The total fixation duration (formalin fixation only or Fixcyt fixation followed by formalin fixation) was equal between the formalin material and Fixcyt material collected from the same patient. Clots from both the formalin-fixed and Fixcyt-fixed aspirated material were then processed into paraffin-embedded blocks, with the same routine biopsy processing program being used for both specimens. Both the

formalin-fixed and Fixcyt-fixed blocks from each patient were collected from the pathology archive. The study was approved by the institutional review board at CWZ. Data and tissue samples were all handled according to the General Data Protection Regulation. All patient material was used anonymously, and collection and use of material was in accordance with the code of conduct for the responsible use of residual human tissue for research, established by the Federation of Dutch Medical Scientific Societies[28].

PD-L1 immunostaining

Consecutive 3-µm sections were cut from both formalin-fixed and Fixcyt-fixed blocks of all included patients. Sections were stained with hematoxylin and eosin (H&E) and with two PD-L1 antibodies, i.e. Ventana (Ventana Medical Systems, Inc., Tucson, Arizona, U.S.A.) SP263 and Dako (Agilent Technologies, Santa Clara, California, U.S.A.) 22C3. The H&E-stained sections were used to determine the number of tumor cells present in each FFPE block, after which patients with one or more blocks containing <100 viable tumor cells were excluded. For immunostaining with SP263, a standardized assay was used on the Ventana Benchmark Ultra platform at Utrecht University Medical Centre (UMCU) (Utrecht, The Netherlands), according to the manufacturer's instructions. On the basis of retrospective analysis of pathology reports generated in 2017 and 2018, the average PD-L1 positivity rates in histological material of NSCLC patients at UMCU were 57% and 28% at the 1% cutoff and at the 50% cutoff, respectively. Similar PD-L1 expression rates have been reported in a large series of NSCLC specimens[29]. The 22C3 antibody was used at CWZ in a laboratory-developed test (LDT) on the Dako Omnis platform [dilution 1:25, 30 min of incubation; heat-induced epitope retrieval (97°C) with low-pH buffer; EnVision FLEX+ (Agilent Technologies, Santa Clara, California, U.S.A.) mouse LINKER detection kit; EnVision FLEX DAB Enhancer (5 min)], validated for use in clinical practice through comparison of sections from 100 NSCLC patients with sections from the same patients immunostained for PD-L1 in an academic referral hospital (all FFPE tissue sections). The 22C3 LDT has shown good to excellent results in several external quality assessment schemes performed by NordiQC and the European Society of Pathology. The average PD-L1 positivity rates were 69% at the 1% cutoff and 39% at the 50% cutoff, on the basis of data from 2017 and 2018. The rate at the 50% cutoff is higher than the positivity rates reported in the Keynote trials that assessed the efficacy of pembrolizumab[4, 5], but positivity rates up to 42.7% have been reported by others[30].

Assessment of PD-L1 immunostaining

All PD-L1-immunostained slides containing ≥100 viable tumor cells were scored by an experienced pathologist, certified for PD-L1 scoring, together with a pathology resident, under a double-headed microscope. The TPS was established by determining the percentage of PD-L1-positive tumor cells relative to the total number of tumor cells. Membranous immunostaining of any intensity was considered to be valid, whereas

cytoplasmic immunostaining and immunostaining of necrotic tumor cells were disregarded. The TPS was determined with scores of 0%, 1%, 5% or 10% for the TPS ranging from 0% to 10%. For scores of >10%, a 10% increment was used. Scoring of all slides immunostained with one antibody was performed within one session. Scoring of all slides immunostained with the other antibody was performed in a second session, with a period of 3 weeks between the scoring sessions and rearrangement of the order of slides before the second session. The scorers were blinded to the fixative that was used, and did not know which slides belonged to the same patient.

Statistical analysis

The concordance of PD-L1 immunostaining between formalin-fixed and Fixcyt-fixed material was assessed by calculating intraclass correlation coefficients (ICCs) for continuous PD-L1 scores. These scores were then dichotomized according to two clinically relevant cutoffs for the TPS, i.e. ≥1% and ≥50%. The overall percentage agreement (OPA), the positive percentage agreement (PPA) and the negative percentage agreement were calculated for both cutoffs, with the material fixed in formalin being used as a reference standard. On the basis of the guideline on principles of analytical validation of immunohistochemical assays from the College of American Pathologists (CAP) Pathology and Laboratory Quality Center, an overall agreement of at least 90% between the formalin-fixed and Fixcyt-fixed material was regarded as acceptable[31]. Cohen's kappa (κ) values were also determined. According to McHugh's suggested interpretation of the kappa statistic for clinical laboratories, values ≥0.80 were considered to indicate strong agreement and were deemed to be acceptable for clinical practice[32]. Statistical analysis was performed with rstudio version 1.1.456 and IBM spss statistics version 25.

RESULTS

Patient and specimen characteristics

Formalin-fixed and Fixcyt-fixed blocks from 67 NSCLC patients were collected. Of these, 54 cases were eligible for inclusion. The remaining 13 cases had insufficient numbers of viable tumor cells (<100) in one or both blocks, and were therefore excluded. Patient and specimen characteristics of the included cases are shown in Table 1.

When a cutoff of ≥1% was used to determine PD-L1 positivity, 33 (61%) and 32 (59%) formalin-fixed specimens, respectively, were PD-L1-positive when immunostaining was performed with the 22C3 LDT and with SP263. With the same cutoff, 23 (43%) and 27 (50%) Fixcyt-fixed specimens were PD-L1-positive. With a cutoff of ≥50%, 14 (26%) formalin-fixed specimens were PD-L1-positive when immunostaining was performed with the 22C3 LDT, and 16 (30%) were PD-L1-positive when immunostaining was performed with SP263.

Of the Fixcyt-fixed specimens, 10 (19%) and 14 (26%), respectively, showed a TPS of ≥50% when immunostaining was performed with the 22C3 LDT and with SP263.

Table 1. Patient and specimen characteristics.

Characteristic	*n* (%)
Diagnosis	
Adenocarcinoma	32 (59)
Squamous cell carcinoma	15 (28)
NSCLC NOS	6 (11)
Adenosquamous carcinoma	1 (2)
TBNA source	
Lymph node	47 (87)
Primary tumor	7 (13)
PD-L1 expression (TPS) formalin ≥1%	
22C3 LDT	33 (61)
SP263	32 (59)
PD-L1 expression (TPS) formalin ≥50%	
22C3 LDT	14 (26)
SP263	16 (30)
PD-L1 expression (TPS) Fixcyt ≥1%	
22C3 LDT	23 (43)
SP263	27 (50)
PD-L1 expression (TPS) Fixcyt ≥50%	
22C3 LDT	10 (19)
SP263	14 (26)

Abbreviations: LDT = laboratory-developed test; NOS = not otherwise specified; NSCLC = non-small cell lung cancer; PD-L1 = programmed death-ligand 1; TBNA = transbronchial needle aspiration; TPS = tumor proportion score.

Comparison of PD-L1 immunostaining between formalin-fixed and Fixcyt-fixed specimens when immunostaining was performed with the 22C3 LDT

When the concordance of TPS was assessed on a continuous scale for material immunostained by use of the 22C3 LDT, the correlation between the formalin-fixed and Fixcyt-fixed specimens was on the boundary between moderate and good [ICC 0.76; 95% confidence interval (CI) 0.60–0.86][33]. However, dichotomization of the TPS at the 1% cutoff and the 50% cutoff resulted in lower concordance levels (Cohen's κ of 0.64 and 0.68, respectively, and OPAs of <90%) (Table 2). Differences in categorization of the TPS between the two types of material, with both cutoffs, are shown in Figure 1A. When the 1% cutoff was used to determine PD-L1 positivity, 10 (30%) of 33 cases that were PD-L1-positive in the formalin-fixed specimen showed false-negative results in the Fixcyt-fixed specimen, resulting in a PPA of only 70%. Figure 2A,B shows a representative example of a case with a TPS of ≥1% in the formalin-fixed specimen and a TPS of <1% in the Fixcyt-fixed specimen. When the 50% cutoff was used, the PPA was only 64%, owing to five (36%) of 14 cases showing false-negative results in the Fixcyt-fixed specimens. Figure 3A,B shows a

representative example of a case with a TPS of ≥50% in the formalin-fixed specimen and a TPS of <50% in the Fixcyt-fixed specimen when immunostaining for PD-L1 was performed with the 22C3 LDT.

Table 2. Concordance of programmed death-ligand 1 (PD-L1) positivity between specimens fixed in formalin and specimens fixed in Fixcyt for SP263 and the 22C3 laboratory-developed test (LDT), with two different cutoffs to determine PD-L1 positivity (≥1% and ≥50%).

	Concordance when a 1% cutoff was used				**Concordance when a 50% cutoff was used**			
	OPA (%)	PPA (%)	NPA (%)	Cohen's kappa (95% CI)	OPA (%)	PPA (%)	NPA (%)	Cohen's kappa (95% CI)
22C3 LDT	81	70	100	0.64 (0.45–0.83)	89	64	98	0.68 (0.44–0.92)
SP263	83	78	91	0.67 (0.47–0.87)	96	88	100	0.91 (0.78–1.00)

Abbreviations: CI = confidence interval; NPA = negative percentage agreement; OPA = overall percentage agreement; PPA = positive percentage agreement.

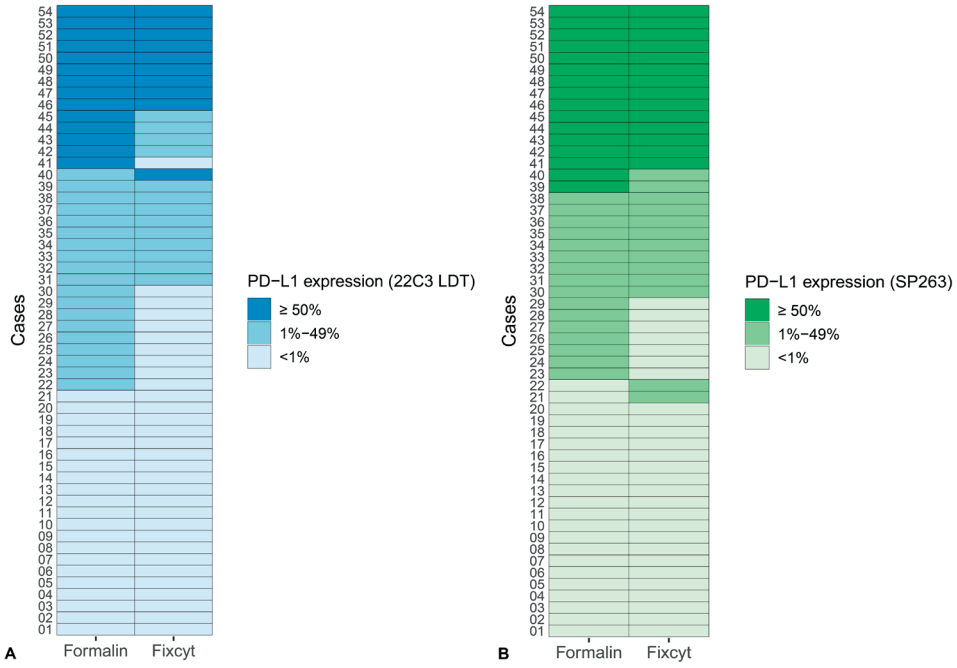


Figure 1. The programmed death-ligand 1 (PD-L1) tumor proportion score (TPS) in three categories for formalin-fixed and Fixcyt-fixed material per case, determined with the 22C3 laboratory-developed test (LDT) (**A**) and the SP263 standardized assay (**B**). Cases for which colors do not correspond showed discordant TPS values between formalin-fixed and Fixcyt-fixed material.

Comparison of PD-L1 immunostaining between formalin-fixed and Fixcyt-fixed specimens when immunostaining was performed with SP263

When immunostaining was performed with SP263, analysis of the concordance of the TPS on a continuous scale showed high agreement (ICC 0.92; 95% CI 0.90–0.95). Again, however, dichotomization of the TPS at the 1% cutoff resulted in much lower concordance levels (OPA of <90% and Cohen's κ of 0.67) (Table 2). Figure 1B shows that, when this cutoff was used, seven (22%) of 32 cases that were PD-L1-positive in the formalin-fixed specimen showed false-negative results in the Fixcyt-fixed specimen (PPA of 78%). A representative example of this phenomenon (lower intensity of PD-L1 immunostaining in Fixcyt-fixed material than in formalin-fixed material when SP263 was used) is shown in Figure 2C,D. Dichotomization of the TPS at the 50% cutoff resulted in much higher concordance levels than those found when the 1% cutoff was used (Cohen's κ of 0.91 versus 0.67 and PPA of 88% versus 78%) (Table 2). Figure 3C,D shows a representative example of a case in which the PD-L1 immunostaining was scored as ≥50% in both the formalin-fixed and the Fixcyt-fixed material when SP263 was used.

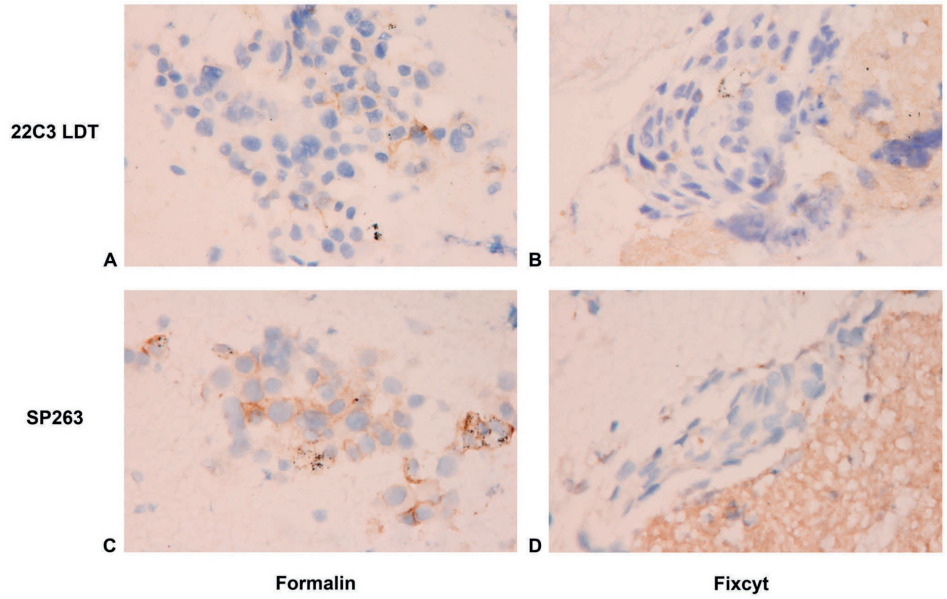


Figure 2. Images of the programmed death-ligand 1 (PD-L1) immunostaining pattern of an exemplary case showing a tumor proportion score (TPS) of ≥1% in formalin-fixed material when immunostaining was performed with the 22C3 laboratory-developed test (LDT) (**A**) and with SP263 (**C**). The Fixcyt-fixed material showed a TPS of <1% when immunostaining was performed with the 22C3 LDT (**B**) and with SP263 (**D**).

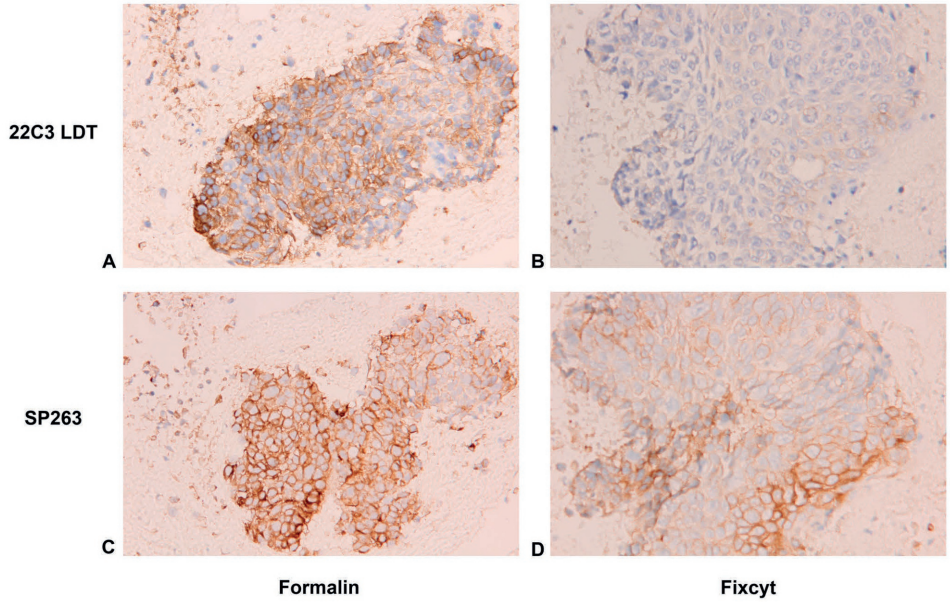


Figure 3. Images of the programmed death-ligand 1 (PD-L1) immunostaining pattern of an exemplary case showing a tumor proportion score (TPS) of ≥50% in formalin-fixed material when immunostaining was performed with the 22C3 laboratory-developed test (LDT) (**A**) and with SP263 (**C**). The Fixcyt-fixed material showed a TPS of <50% when immunostaining was performed with the 22C3 LDT (**B**) and a TPS of ≥50% when immunostaining was performed with SP263 (**D**).

Comparison of PD-L1 immunostaining between the 22C3 LDT and SP263

See Supporting information for the results of a comparison of PD-L1 immunostaining between the 22C3 LDT and SP263 for both formalin-fixed and Fixcyt-fixed material.

DISCUSSION

In this retrospective study, the effect of pre-fixation in an ethanol-based fixative on PD-L1 immunostaining was studied with two PD-L1 immunohistochemical assays (the 22C3 LDT and the SP263 standardized assay) validated for use on FFPE tissue. It was shown that fixation of EBUS-TBNA material in Fixcyt (ethanol-based) prior to formalin fixation resulted in a considerable number of false-negative PD-L1 immunostaining results when the 1% cutoff was used to determine PD-L1 positivity. When PD-L1 positivity was determined with the 50% cutoff, immunostaining by use of the 22C3 LDT again resulted in a substantial number of false-negative PD-L1 immunostaining results in the Fixcyt-fixed material, whereas this was not the case for SP263.

Few other studies have assessed the influence of ethanol (pre-)fixation on PD-L1 immunostaining results in cytology cell blocks. Wang et al.[34] concluded that alcohol fixation (methanol or ethanol) does not affect PD-L1 immunostaining. However, no comparisons between paired samples of the same patients were made, and the authors state that it is possible that most of the cases fixed in both alcohol and formalin had short alcohol pre-fixation times. A study by Gosney et al.[35], which did use paired samples of aspirates fixed in alcohol-based fixatives and formalin, also reported no effect of the use of alcohol-based fixatives on the expression of PD-L1 or its interpretation. Most of the alcohol-fixed samples in the study, however, were fixed in CytoRich Red, a solution that also contains formaldehyde. These samples were therefore exposed to more formalin during the processing procedure than if they had been fixed in an ethanol-based fixative such as Fixcyt, which could potentially explain the differences in results from those of our own study. A study by Jain et al.[36] found an overall concordance of PD-L1 immunostaining of 88.4% between small biopsies and matched liquid-based cytology smears, but, similarly to the procedure of Gosney et al.[35], these smears were fixed in CytoRich Red. Similarly to our findings, Lloyd et al. showed a negative effect of CytoLyt (a methanol-based fixative) on PD-L1 immunostaining in PD-L1-expressing cell lines[37].

The occurrence of false-negative PD-L1 immunostaining results is problematic, as it could lead to denial of potentially beneficial treatment options to patients with NSCLC. The use of EBUS-TBNA specimens for PD-L1 testing is most often seen in patients who present with locally advanced or metastatic disease at diagnosis. These patients are precisely the ones who could benefit from treatment with PD-1/PD-L1 inhibitors. It is thus of the utmost importance that pathologists can accurately determine the PD-L1 TPS for these patients. Hence, pathologists should be aware of the risks of using an ethanol-based (pre-)fixative. The concordance levels of the PD-L1 TPS were lowest for both the 22C3 LDT and SP263 with use of the 1% cutoff, indicating that the use of ethanol-based fixatives could result in the wrongful denial of durvalumab as consolidation treatment to stage III NSCLC patients in clinical practice in various European countries[11]. The combination of Fixcyt-fixed material and immunostaining by use of the 22C3 LDT also resulted in disappointing concordance levels when a 50% cutoff was used. This could result in stage IV NSCLC patients being prescribed a more toxic first-line treatment regimen of pembrolizumab combined with chemotherapy, whereas they could have been treated with a less toxic treatment option consisting of pembrolizumab alone[9, 10].

Interestingly, the use of SP263 resulted in high concordance of the PD-L1 TPS between formalin-fixed and Fixcyt-fixed material when the 50% cutoff was used for dichotomization. Overall, slightly stronger immunostaining intensity was seen in sections immunostained with SP263 than in those immunostained by use of the 22C3 LDT, with discordant cases more often showing a lower TPS when the 22C3 LDT was used than when SP263 was used. A

decrease in immunostaining intensity due to ethanol (pre-)fixation might therefore be less problematic when SP263 is used than when the 22C3 LDT is used. It could also be that the negative effect of ethanol (pre-)fixation on PD-L1 immunoreactivity is smaller with the use of SP263 than with use of the 22C3 LDT. Although these two antibodies target the same protein, they do target different epitopes of this protein[38]. Perhaps the alteration in tertiary structure of the PD-L1 protein caused by alcohol fixation[21] results in one epitope being more capable of binding with the PD-L1 antibody than the other. A study by Munari et al.[39] also showed a high concordance rate between ethanol-fixed material (cytological smears) and FFPE surgical resection material at the 50% cutoff, when the SP263 standardized assay was used. Scoring of PD-L1 positivity with the 50% cutoff thus seems to be feasible on ethanol-prefixed material when SP263 is used. However, we would suggest assessing the concordance between formalin-fixed and ethanol-prefixed material in more 'critical samples', i.e. samples that show a PD-L1 TPS closer to the threshold of PD-L1 positivity[40], before drawing a more definitive conclusion on the actual feasibility of using SP263 on ethanol-prefixed material to determine PD-L1 positivity at a 50% cutoff in clinical practice.

It has been described previously that alcohol fixation leads to false-negative immunostaining results when conventional IHC protocols, validated for FFPE tissue, are used[23-27, 41]. Nevertheless, in clinical practice, cell blocks are often seen as ideal for immunostaining[42], even though different fixatives, including alcohol-based fixatives, may be used in preparing them. Fowler and Lachar state that it is a common mistake in cytology laboratories to not carefully examine the cell block methodology and its potential negative impact on IHC interpretation[43]. They advise comparison of immunostaining results between cytological samples and surgical pathology samples, prior to the introduction of any new cell block method[43]. This is in line with the guideline from the CAP, which states that laboratories should test a sufficient number of cases to ensure that IHC assays achieve the expected results, if they are used on cytological specimens that are not processed in the same manner as the tissues used for assay validation[31]. Similarly, Rekhtman et al.[44], who recently developed a modified HistoGel-based cell block preparation method that includes the addition of 95% ethanol, state that laboratories who adopt their method should consider revalidation of immunostains. In the literature, there are some examples demonstrating that changing an IHC protocol can lead to good immunostaining results when alcohol-fixed specimens are used, whereas the standard IHC protocol, validated for FFPE tissue, showed reduced immunostaining intensity[23]. However, in clinical practice, many laboratories use standardized PD-L1 assays, such as the standardized SP263 assay and the 22C3 pharmDx assay, which received Food and Drug Administration approval and/or CE-IVD marking for use only with standardized protocols designed by the manufacturers[45-47]. Moreover, changing an IHC protocol does not always result in improved immunostaining[23, 25], leaving it unsure whether changing PD-L1 IHC protocols would actually result in fewer false-negative immunostaining results when they are used on ethanol-(pre)fixed specimens.

This study has some limitations. First, the sample size is rather small. Because of the retrospective nature of this study, the availability of patient material was dependent on the number of NSCLC patients who had actually undergone EBUS-TBNA in clinical practice and had their aspirated material fixed in both Fixcyt and formalin. However, we do believe that the design of this study, which used paired samples to compare PD-L1 immunostaining between ethanol-prefixed and formalin-fixed material, provides a valuable contribution to the current literature. Second, we used the cell blocks fixed in formalin as a reference standard in our analyses. Ideally, we would have liked to use true histological specimens as the gold standard for our comparisons. Unfortunately, matched histological specimens were not available for our study cases. Third, we only used two PD-L1 antibodies in our study, whereas, in clinical practice, laboratories may use other PD-L1 antibodies, such as 28-8 in the pharmDx assay or E1L3N. We cannot draw any conclusions regarding the effect of ethanol-based fixation on PD-L1 immunostaining with antibodies and protocols other than the ones that we used. Fourth, the samples in our study were fixed in ethanol for only a short period of time, with a maximum of 2 h. In clinical practice, aspirated material might sometimes rest in a fixative for a longer amount of time, especially when it is collected just before or during the weekend. Potentially, a longer duration of ethanol (pre-)fixation could result in even more detrimental effects on PD-L1 immunostaining. Also, we cannot draw definitive conclusions on the effect of formalin post-fixation after alcohol fixation on PD-L1 immunostaining, as we did not know the individual fixation times for the samples in our study, and were therefore unable to determine whether there was any difference in PD-L1 results between specimens with shorter and longer formalin post-fixation times. Fifth, the uneven distribution of aspirated material between formalin and Fixcyt could have played a role in causing discrepancies between formalin-fixed and Fixcyt-fixed samples in our study, as the Fixcyt-fixed material might have contained a smaller number of tumor cells. We did not, however, observe prominent differences in the numbers of tumor cells between the two types of specimens when scoring PD-L1. Moreover, nearly one-third of the excluded cases were excluded because the formalin-fixed cell blocks did not contain enough tumor cells, whereas the Fixcyt-fixed cell blocks did. We therefore do not believe that the number of tumor cells was systematically lower in the Fixcyt-fixed specimens, diminishing the influence that the uneven distribution of the aspirated material would have had on creating discordance in PD-L1 immunostaining between the different specimen types. Finally, both intratumoral heterogeneity and intraobserver variability could potentially explain part of the discordance seen between formalin-fixed and Fixcyt-fixed material, which is also indicated by the three cases that showed higher PD-L1 expression in the Fixcyt-fixed specimen than in the formalin-fixed specimen (two cases when immunostaining was performed with SP263; one case when immunostaining was performed with the 22C3 LTD). However, as the overall discordance that we observed mainly consisted of false-negative results in Fixcyt-fixed material as compared with formalin-fixed material, rather than a mix of both false-negative and false-positive results, it seems highly unlikely that

the discordance could be explained solely by the presence of intratumoral heterogeneity or intraobserver variability.

To conclude, when SP263 and 22C3 IHC protocols, validated for use on FFPE material, are used on cytological specimens pre-fixed in an ethanol-based fixative, this results in a considerable number of false-negative PD-L1 immunostaining results. This occurs when both the 1% and the 50% cutoffs are used to determine PD-L1 positivity, although the risk of false-negative results seems to be smallest when the SP263 standardized assay is used and scoring of PD-L1 positivity is performed with the 50% cutoff. Pathologists should be aware that scoring of PD-L1 expression as negative on the basis of standard PD-L1 IHC protocols used on specimens (pre-)fixed in ethanol could lead to patients wrongfully being denied treatment options that they could actually benefit from.

ACKNOWLEDGEMENTS

The graphical abstract was created by use of the following artwork from Servier Medical Art (Les Laboratoires Servier, https://smart.servier.com): analysis flask, https://smart.servier.com/smart_image/culture-flask-4/; slide, https://smart.servier.com/smart_image/slide/. Servier Medical Art is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

REFERENCES

1. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science 2018;359:1350–1355.
2. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N. Engl. J. Med. 2015;373:1627–1639.
3. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N. Engl. J. Med. 2015;373:123–135.
4. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N. Engl. J. Med. 2016;375:1823–1833.
5. Herbst RS, Baas P, Kim D-W, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387:1540–1550.
6. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 2017;389:255–265.
7. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. N. Engl. J. Med. 2017;377:1919–1929.
8. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. Lancet. 2019;393:1819–1830.
9. Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. 2018;29(Suppl. 4):iv192–iv237.
10. European Society for Medical Oncology (ESMO). Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. 2019. Available from: https://www.esmo.org/Guidelines/Lung-and-Chest-Tumours/Metastatic-Non-Small-Cell-Lung-Cancer (Accessed 11 November 2019).
11. EMA. Imfinzi (Durvalumab). 2018. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/imfinzi (Accessed 25 September 2019).
12. Scheerens H, Malong A, Bassett K, et al. Current status of companion and complementary diagnostics: strategic considerations for development and launch. Clin. Transl. Sci. 2017;10:84–92.
13. Russell-Goldman E, Kravets S, Dahlberg SE, Sholl LM, Vivero M. Cytologic-histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. Cancer Cytopathol. 2018;126:253–263.
14. kov BG, Høgdall E, Clementsen P, et al. The prevalence of EGFR mutations in non-small cell lung cancer in an unselected Caucasian population. APMIS. 2015;123:108–115.
15. Roh MH. The utilization of cytologic fine-needle aspirates of lung cancer for molecular diagnostic testing. J. Pathol. Transl. Med. 2015;49:300–309.
16. Nambirajan A, Jain D. Cell blocks in cytopathology: an update. Cytopathology. 2018;29:505–524.
17. Skov BG, Skov T. Paired comparison of PD-L1 expression on cytologic and histologic specimens from malignancies in the lung assessed with PD-L1 IHC 28–8pharmDx and PD-L1 IHC 22C3pharmDx. Appl. Immunohistochem. Mol. Morphol. 2017;25:453–459.
18. Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death-

ligand 1 antibody to determine programmed death-ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. Cancer Cytopathol. 2018;126:264–274.

19. Sakakibara R, Inamura K, Tambo Y, et al. EBUS-TBNA as a promising method for the evaluation of tumor PD-L1 expression in lung cancer. Clin. Lung Cancer. 2017;18:527–534.
20. Hernandez A, Brandler TC, Zhou F, Moreira AL, Schatz-Siemers N, Simsir A. Assessment of programmed death-ligand 1 (PD-L1) immunohistochemical expression on cytology specimens in non-small cell lung carcinoma. Am. J. Clin. Pathol. 2019;151:403–415.
21. Eltoum I, Fredenburgh J, Myers RB, Grizzle WE. Introduction to the theory and practice of fixation of tissues. J. Histotechnol. 2001;24:173–190.
22. Chung J-Y, Song JS, Ylaya K, et al. Histomorphological and molecular assessments of the fixation times comparing formalin and ethanol-based fixatives. J. Histochem. Cytochem. 2018;66:121–135.
23. Sauter JL, Grogg KL, Vrana JA, Law ME, Halvorson JL, Henry MR. Young investigator challenge: validation and optimization of immunohistochemistry protocols for use on cellient cell block specimens. Cancer Cytopathol. 2016;124:89–100.
24. Gruchy JR, Barnes PJ, Dakin Hache KA. CytoLyt(R) fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. Appl. Immunohistochem. Mol. Morphol. 2015;23:297–302.
25. Buonocore DJ, Konno F, Jungbluth AA, et al. CytoLyt fixation significantly inhibits MIB1 immunoreactivity whereas alternative Ki-67 clone 30–9 is not susceptible to the inhibition: critical diagnostic implications. Cancer Cytopathol. 2019;127:643–649.
26. Gorman BK, Kosarac O, Chakraborty S, Schwartz MR, Mody DR. Comparison of breast carcinoma prognostic/predictive biomarkers on cell blocks obtained by various methods: cellient, formalin and thrombin. Acta Cytol. 2012;56:289–296.
27. Hanley KZ, Birdsong GG, Cohen C, Siddiqui MT. Immunohistochemical detection of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 expression in breast carcinomas: comparison on cell block, needle-core, and tissue block preparations. Cancer. 2009;117:279–288.
28. FEDERA. Human tissue and medical research: Code of Conduct for responsible use. 2011. Available from: https://www.federa.org/codes-conduct (Accessed 30 September 2019).
29. Evans M, O'Sullivan B, Hughes F, et al. The clinicopathological and molecular associations of PD-L1 expression in non-small cell lung cancer: analysis of a series of 10,005 cases tested with the 22C3 assay. Pathol. Oncol. Res. 2020;26:79–89.
30. Jin Y, Shen X, Pan Y, et al. Correlation between PD-L1 expression and clinicopathological characteristics of non-small cell lung cancer: a real-world study of a large Chinese cohort. J. Thorac. Dis. 2019;11:4591–4601.
31. Fitzgibbons PL, Bradley LA, Fatheree LA, et al. Principles of analytic validation of immunohistochemical assays: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch. Pathol. Lab. Med. 2014;138:1432–1443.
32. McHugh ML. Interrater reliability: the kappa statistic. Biochem. Med. (Zagreb) 2012; 22; 276–282.
33. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 2016;15:155–163.
34. Wang H, Agulnik J, Kasymjanova G, et al. Cytology cell blocks are suitable for immunohistochemical testing for

PD-L1 in lung cancer. Ann. Oncol. 2018;29:1417–1422.

35. Gosney JR, Haragan A, Chadwick C, et al. Programmed death ligand 1 expression in EBUS aspirates of non-small cell lung cancer: is interpretation affected by type of fixation? Cancer Cytopathol. 2020;128:100–106.
36. Jain D, Sukumar S, Mohan A, Iyer VK. Programmed death-ligand 1 immunoexpression in matched biopsy and liquid-based cytology samples of advanced stage non-small cell lung carcinomas. Cytopathology 2018;29:550–557.
37. Lloyd IE, Zhou W, Witt BL, Chadwick BE. Characterization of PD-L1 immunohistochemical expression in cell blocks with different specimen fixation and processing methods. Appl. Immunohistochem. Mol. Morphol. 2019;27:107–113.
38. Giraldo N, Taube J. PD-L1 and other immunological diagnosis tools. In Zitvogel L, Kroemer G eds. Oncoimmunology: a practical guide for cancer immunotherapy. Cham, Switzerland: Springer International Publishing. 2017:375–377.
39. Munari E, Zamboni G, Sighele G, et al. Expression of programmed cell death ligand 1 in non-small cell lung cancer: comparison between cytologic smears, core biopsies, and whole sections using the SP263 assay. Cancer Cytopathol. 2019;127:52–61.
40. Thunnissen E, de Langen AJ, Smit EF. PD-L1 IHC in NSCLC with a global and methodological perspective. Lung Cancer. 2017;113:102–105.
41. Kruger AM, Stevens MW, Kerley KJ, Carter CD. Comparison of the CellientTM automated cell block system and agar cell block method. Cytopathology. 2014;25:381–388.
42. Schmitt F, Cochand-Priollet B, Toetsch M, Davidson B, Bondi A, Vielh P. Immunocytochemistry in Europe: results of the European Federation of Cytology Societies (EFCS) inquiry. Cytopathology. 2011;22:238–242.
43. Fowler LJ, Lachar WA. Application of immunohistochemistry to cytology. Arch. Pathol. Lab. Med. 2008;132:373–383.
44. Rekhtman N, Buonocore DJ, Rudomina D, et al. Novel modification of HistoGel-based cell block preparation method: improved sufficiency for molecular studies. Arch. Pathol. Lab. Med. 2018;142:529–535.
45. Dako Agilent Pathology Solutions. PD-L1 IHC 22C3 pharmDx is CE-IVD-marked for in vitro diagnostic use. Available from: https://www.agilent.com/cs/library/usermanuals/public/29171_22C3-ihc-pharmdx-interpretation-manual-eu.pdf (Accessed 20 November 2019).
46. US Food and Drug Administration: Dako PD-L1 IHC 22C3 pharmDx. 2015. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150013c.pdf (Accessed 20 November 2019).
47. Ventana Medical Systems, Inc: VENTANA PD-L1 (SP263) Assay. 2019. Available from: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp263-assay2.html (Accessed 20 November 2019).

SUPPORTING INFORMATION

Supplementary results

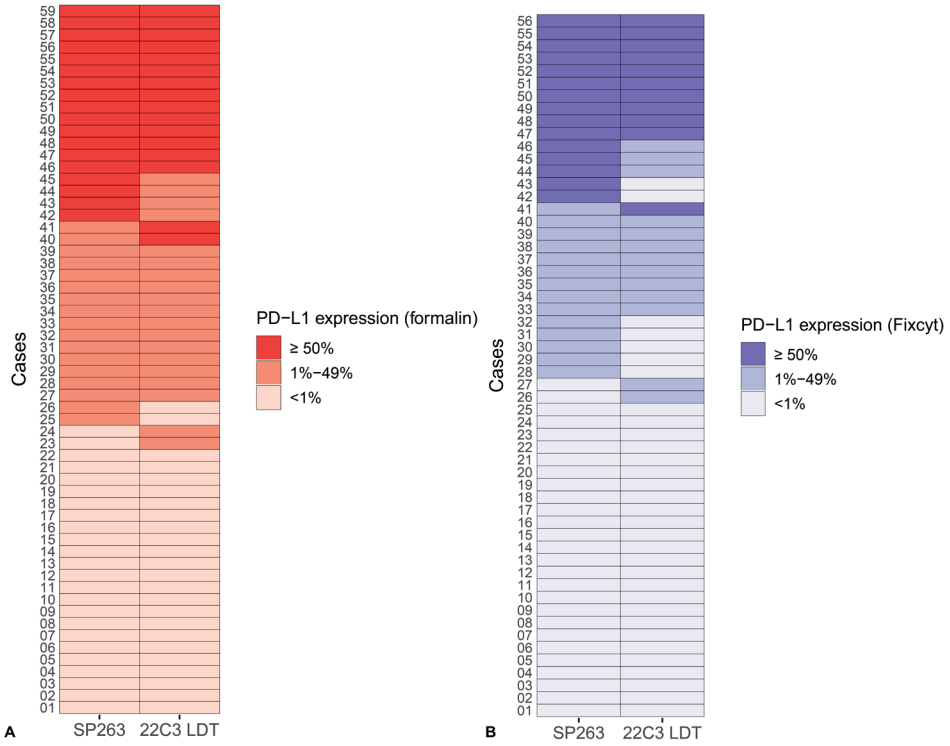
Analysis of concordance between SP263 and the 22C3 laboratory-developed test (LDT) of programmed death-ligand 1 (PD-L1) tumor proportion score (TPS) on a continuous scale showed high agreement when performed on formalin-fixed material (intraclass correlation coefficient (ICC) 0.91; 95% confidence interval (CI) 0.85-0.95). Dichotomization of TPS at the 1% and 50% cutoff resulted in a high concordance level at the 1% cutoff (Cohen's kappa (κ) 0.86) and a somewhat lower concordance level at the 50% cutoff (Cohen's κ 0.75), with overall percentage agreement (OPA) being ≥90% for both cutoffs (Table S1). Differences in categorization of TPS between the two antibodies, with both cutoffs, are displayed by Figure S1A. Of 10 discordant cases, six (60%) showed a lower TPS with the 22C3 LDT as compared with SP263.

When analyzing the Fixcyt-fixed material, concordance between SP263 and the 22C3 LDT was lower in both the analysis of TPS on a continuous scale (ICC 0.82; CI 0.67-0.90) and the analysis of TPS using dichotomized data (OPA 84% and Cohen's κ 0.68 at 1% cutoff; OPA 89% and Cohen's κ 0.70 at 50% cutoff) (Table S1). Differences in categorization of TPS between the two antibodies, with both cutoffs, are displayed by Figure S1B. There were 13 discordant cases, of which 10 (77%) showed a lower TPS in the 22C3 LDT as compared with SP263.

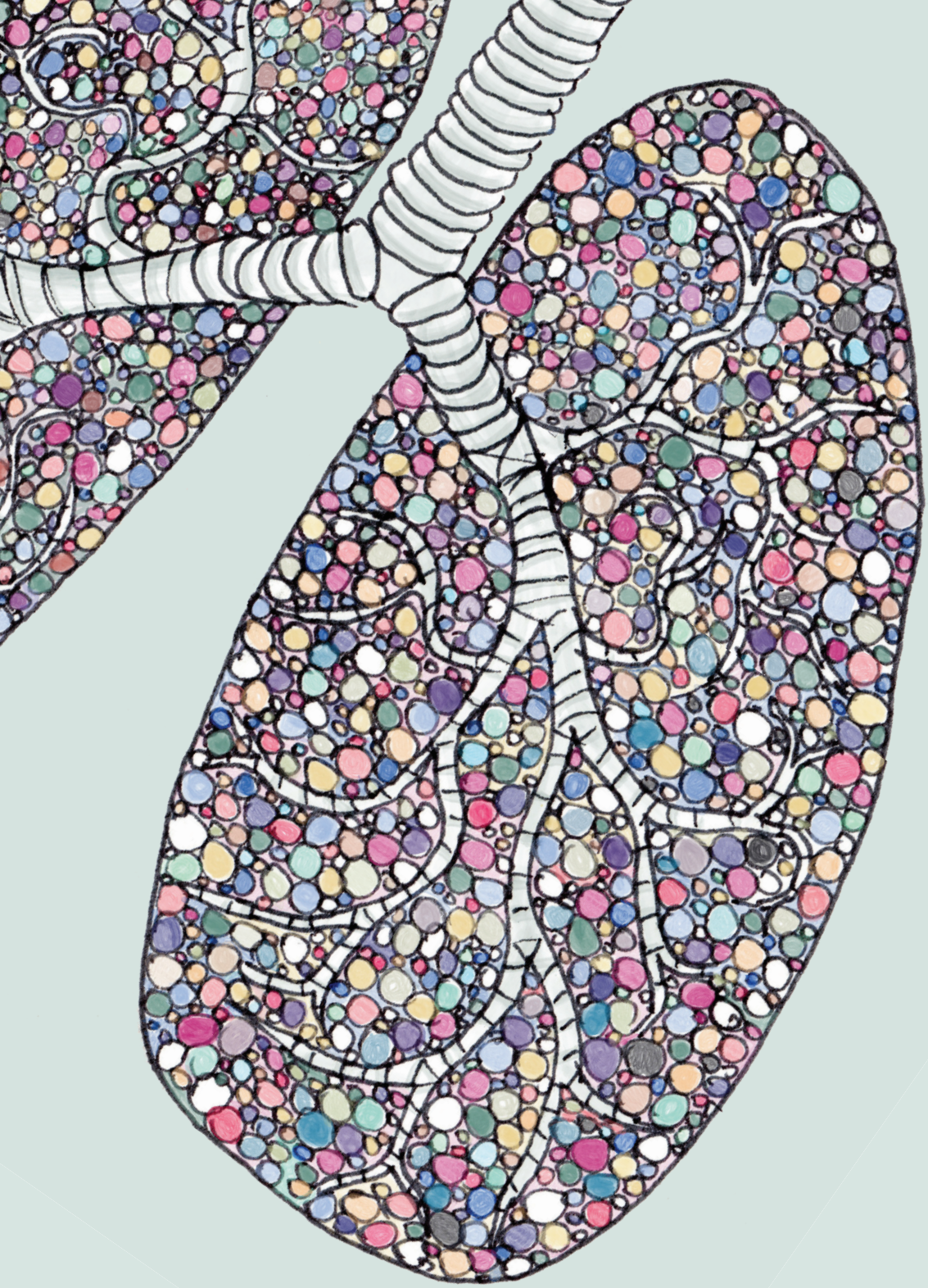
Table S1. Concordance of programmed death-ligand 1 (PD-L1) positivity between SP263 and the 22C3 laboratory-developed test for specimens fixed in formalin and in Fixcyt, with two different cutoffs to determine PD-L1 positivity (≥1% and ≥50%).

	Concordance when a 1% cutoff was used		**Concordance when a 50% cutoff was used**	
	OPA (%)	Cohen's kappa (95% CI)	OPA (%)	Cohen's kappa (95% CI)
Formalin	93	0.86 (0.72-1.00)	90	0.75 (0.56-0.94)
Fixcyt	84	0.68 (0.49-0.87)	89	0.70 (0.48-0.93)

Abbreviations: CI = confidence interval; OPA = overall percentage agreement.



Supporting figure S1. The programmed death-ligand 1 (PD-L1) tumor proportion score (TPS) in three categories for material immunostained by use of the SP263 standardized assay and the 22C3 laboratory-developed test (LDT) per case, determined for formalin-fixed material (**A**) and for Fixcyt-fixed material (**B**). Cases for which colors do not correspond, showed discordant TPS values between SP263 and the 22C3 LDT.



5

CONSIDERABLE INTERLABORATORY VARIATION IN PD-L1 POSITIVITY IN A NATIONWIDE COHORT OF NON-SMALL CELL LUNG CANCER PATIENTS

Koomen BM
Voorham QJM
Epskamp-Kuijpers CCHJ
van Dooijeweert C
van Lindert ASR
Deckers IAG
Willems SM

Lung Cancer. 2021;159:117-126

ABSTRACT

Objectives

Immunohistochemical expression of programmed death-ligand 1 (PD-L1) is used as a predictive biomarker for prescription of immunotherapy to non-small cell lung cancer (NSCLC) patients. Accurate assessment of PD-L1 expression is therefore crucial. In this study, the extent of interlaboratory variation in PD-L1 positivity in the Netherlands was assessed, using real-world clinical pathology data.

Materials and Methods

Data on all NSCLC patients in the Netherlands with a mention of PD-L1 testing in their pathology report from July 2017 to December 2018 were extracted from PALGA, the nationwide network and registry of histo- and cytopathology in the Netherlands. PD-L1 positivity rates were determined for each laboratory that performed PD-L1 testing, with separate analyses for histological and cytological material. Two cutoffs (1% and 50%) were used to determine PD-L1 positivity. Differences between laboratories were assessed using funnel plots with 95% confidence limits around the overall mean.

Results

6,354 patients from 30 laboratories were included in the analysis of histology data. At the 1% cutoff, maximum interlaboratory variation was 39.1% (32.7%–71.8%) and ten laboratories (33.3%) differed significantly from the mean. Using the 50% cutoff, four laboratories (13.3%) differed significantly from the mean and maximum variation was 23.1% (17.2%–40.3%). In the analysis of cytology data, 1,868 patients from 23 laboratories were included. Eight laboratories (34.8%) differed significantly from the mean in the analyses of both cutoffs. Maximum variation was 41.2% (32.2%–73.4%) and 29.2% (14.7%–43.9%) using the 1% and 50% cutoffs, respectively.

Conclusion

Considerable interlaboratory variation in PD-L1 positivity was observed. Variation was largest using the 1% cutoff. At the 50% cutoff, analysis of cytology data demonstrated a higher degree of variation than the analysis of histology data.

INTRODUCTION

Therapies targeting programmed death-1 (PD-1) receptor or its ligand programmed death-ligand 1 (PD-L1) have become a relevant component of standard treatment regimens in patients with advanced non-small cell lung cancer (NSCLC)[1, 2]. In clinical practice, oncologists rely on results from PD-L1 immunohistochemistry (IHC) to make treatment decisions for NSCLC patients. PD-L1 IHC is performed in pathology laboratories and assessed by pathologists, who determine the percentage of tumor cells that show PD-L1 expression relative to the total amount of tumor cells present (tumor proportion score (TPS)). Based on the TPS, patients are offered different treatment options. For instance, in many European countries, the PD-L1 inhibitor durvalumab, prescribed as consolidation treatment to stage III NSCLC patients, is only reimbursed for patients whose tumors show a PD-L1 TPS of ≥ 1%. This is based on advice from the European Medicines Agency (EMA)[3]. Similarly, only patients with advanced NSCLC whose tumor biopsy expresses a PD-L1 TPS of ≥ 50% are offered the option of monotherapy with the PD-1 inhibitor pembrolizumab, which is significantly better tolerated than the standard combination strategy which includes chemotherapy[1, 2]. Therefore, it is crucial that TPS is determined reliably by all pathology laboratories that perform PD-L1 IHC.

An important factor in accurate assessment of PD-L1 TPS is the scoring performed by pathologists. Studies have shown that in a real-world clinical setting, substantial variation exists in histologic grading of various tumor types between laboratories[4, 5, 6, 7]. Regarding PD-L1, it has been reported that interpathologist agreement of PD-L1 scoring on tumor cells may be high, but is often found to be decreased when using the 1% cutoff[8]. Discordance in PD-L1 scoring between pathologists might result in a high degree of variation in PD-L1 positivity rates between laboratories. Whether this is in fact true for clinical practice, is of yet unknown.

Other analytical and pre-analytical factors of IHC could also contribute to interlaboratory variation in positivity rates of IHC biomarkers[9, 10]. With PD-L1 IHC, laboratories may for instance differ in their choice of antibody, as multiple antibodies are available and it is unfeasible for laboratories to use more than one antibody for the same test[10]. Some studies have questioned whether different PD-L1 assays and laboratory-developed tests (LDTs) can be used interchangeably[8, 11]. Furthermore, whereas processing of histological material is often performed in a similar way by different laboratories, many differences can be found in processing of cytological material[12, 13, 14]. It is possible that these differences might increase interlaboratory variation in PD-L1 positivity.

A high degree of variation in PD-L1 positivity rates between pathology laboratories could result in patients being scored as PD-L1 positive in one laboratory and as PD-L1 negative

in another. This could lead to patients being denied effective treatment options or being exposed to unnecessary toxicity. In order to determine whether this may be the case in the real-world setting, we conducted a retrospective cohort study to assess the existence of interlaboratory variation in PD-L1 positivity in clinical practice in the Netherlands, using real-world clinical pathology data of NSCLC patients.

MATERIALS AND METHODS

Data extraction

Data were extracted from a national database governed by the PALGA foundation, the nationwide network and registry of histo- and cytopathology in the Netherlands. This database contains excerpts from pathology reports dating back to 1971 and manages all pathology records from all Dutch pathology laboratories since 1991[15]. Patients at each individual institution connected to PALGA can opt-out of consenting to the use of their data for research, which is estimated to occur in 3% of all patients[16]. All personal data in the database are pseudonymized by a Trusted Third Party (ZorgTTP, Houten, the Netherlands), ensuring that individual patients are not identifiable. According to the Central Committee on Research involving Human Subjects (CCMO), this type of study does not require approval from an ethics committee in the Netherlands. This study was approved by PALGA's Scientific Council and Privacy Committee and all data was handled according to the General Data Protection Regulation (GDPR).

Data were retrieved on all NSCLC patients in the Netherlands with a mention of PD-L1 testing in their pathology report between 1 July 2017 and 31 December 2018. Patients with two primary lung tumors were excluded, since treatment of the one tumor might have influenced PD-L1 expression on the other tumor[17, 18]. For each patient, the following data were extracted: age, sex, histologic subtype, number of PD-L1 tests performed, source of material for PD-L1 test(s), type of material for PD-L1 test(s), antibody and protocol (commercial assay or LDT) used for PD-L1 test(s), reported TPS and number of tumor cells (TCs) present. Additionally, information on academic or non-academic status was subtracted for each laboratory.

Analysis of interlaboratory variation in PD-L1 positivity

To study variation in PD-L1 positivity, we compared the percentages of reported PD-L1 positive patients between all laboratories that performed PD-L1 testing. We determined PD-L1 positivity based on two cutoffs for TPS relevant in clinical practice, i.e. ≥ 1% and ≥ 50%. Interlaboratory variation was studied with separate analyses for each of these cutoffs: one analysis with data dichotomized according to the 1% cutoff, and one with data dichotomized according to the 50% cutoff. Moreover, since processing of cytological

material may vary greatly between laboratories, we analyzed histological and cytological material separately.

For each patient, only one PD-L1 test was included for analysis. If patients had had PD-L1 testing performed on both histological and cytological material, they were included in both the analysis of histological material and the analysis of cytological material. Only laboratories that performed PD-L1 testing in ≥ 30 patients during the study period were included in the final analyses. If test results of patients with multiple PD-L1 tests performed on the same type of material (either all on histology or all on cytology) were discordant, these patients were excluded. Patients with inconclusive test results only and patients with tests without a reported TPS were excluded as well.

Analysis based on IHC antibody and protocol

In order to get a preliminary idea of the role that different antibodies might play in causing variation in PD-L1 positivity, separate analyses were performed that incorporated information on use of IHC antibody and protocol (commercial assay or LDT) by individual laboratories in plots displaying interlaboratory variation in PD-L1 positivity rates. These plots were created using histology data only.

Statistical analysis

Patient and PD-L1 test characteristics were summarized using counts and proportions for histological and cytological material separately. Differences between PD-L1 positive and negative subgroups were tested by using a chi-square (χ^2)-test for categorical variables and a t-test for continuous variables.

The overall proportions of PD-L1 positivity were determined for histological and cytological material separately, using both the 1% cutoff and the 50% cutoff. These overall proportions were considered the mean national proportion. Differences in PD-L1 positivity rates between laboratories were assessed by creating funnel plots, which display the mean national proportion of PD-L1 positivity with its 95% confidence limits and the percentage of PD-L1 positive patients plotted against the total number of patients tested for each laboratory. This allows for comparison of laboratories to each other and to the mean[19]. All laboratories falling outside the 95% confidence limits were considered to differ significantly from the mean. Academic and non-academic laboratories were indicated separately within the funnel plots.

The positivity rates displayed in the funnel plots for each individual laboratory were adjusted for case mix (i.e. differences in patient and test characteristics) by performing multivariate logistic regression analysis using predetermined variables. These variables included age, sex, histologic subtype and source of material for PD-L1 testing. For the analysis of histology

data, type of material (i.e. biopsy or resection) was also added to the logistic regression model. As information on the IHC antibody and/or protocol used for PD-L1 testing was lacking in a considerable percentage of cases (in 32.3% and 40.3% of cases using histology and cytology data, respectively), this variable was not included in the main multivariate logistic regression model. In order to get a general idea of the role that use of different IHC antibodies and protocols might play in causing interlaboratory variation in PD-L1 positivity, additional case mix adjusted PD-L1 positivity rates were calculated using multivariate logistic regression analysis that did include the variable IHC antibody/protocol. These positivity rates were then compared with the case mix adjusted positivity rates without the variable IHC antibody/protocol, by displaying both in one funnel plot. Case mix adjusted positivity rates were determined by dividing the observed percentage of PD-L1 positive patients per laboratory by the expected percentage, based on the multivariate logistic regression model, followed by multiplying with the national mean percentage of PD-L1 positivity.

All statistical analysis was performed using IBM SPSS Statistics version 25.

RESULTS

Patient selection process

Data of 9,153 NSCLC patients with a mention of PD-L1 in their pathology report were retrieved from the PALGA database. PD-L1 testing was performed on histological and/or cytological material of 8,725 of these patients, with 10,625 PD-L1 tests performed in total. Data from 42 pathology laboratories were included in the data set, of which 32 performed PD-L1 testing themselves. PD-L1 testing was performed on histological material in 6,755 cases and on cytological material in 2,300 cases. For the analysis of histology data, two laboratories that performed PD-L1 testing in < 30 patients were excluded, resulting in exclusion of 43 patients. After exclusion of patients with discordant results of multiple PD-L1 tests (n = 178), patients with inconclusive test results only (n = 166) and patients with tests with unknown TPS (n = 14), 6,354 patients from 30 laboratories remained for analysis of interlaboratory variation using data on histological material (Figure 1). The 166 patients with inconclusive test results only had a total of 177 tests performed. Of these tests, 143 (80.8%) were inconclusive because the amount of viable tumor cells was insufficient (< 100). In the remaining cases, various reasons why the test was considered inconclusive were reported, such as too much background staining, hard to distinguish tumor cells within inflammatory infiltrate, and mechanical damage to tissue. In some cases, the reason why the test was considered inconclusive was not reported. For the analysis of cytology data, patients from nine laboratories that performed PD-L1 testing in < 30 patients were excluded (n = 111). 1,868 patients from 23 laboratories remained for analysis of interlaboratory variation after exclusion of patients with discordant results of multiple

PD-L1 tests (n = 23), patients with inconclusive test results only (n = 290) and patients with tests with unknown TPS (n = 8) (Figure 1). A total of 309 tests was performed in the patients with inconclusive test results only. In 236 (76.4%) of these tests, there was an insufficient amount of viable tumor cells (< 100).

Characteristics of all patients included in the analysis of histology data are displayed in Table 1. Proportions significantly differed between PD-L1 positivity and negativity across histologic subtype, source of material, and type of material for both cutoffs (i.e. 1% and 50%) and across IHC antibody/protocol at the 1% cutoff, although the observed differences in percentages were sometimes small and not always clinically relevant. Table 2 shows the characteristics of all patients included in the analysis of cytology data. Proportions across sex, histologic subtype and IHC antibody/protocol differed significantly between PD-L1 positivity and PD-L1 negativity for both cutoffs. All differences in patient/specimen characteristics between PD-L1 positivity and PD-L1 negativity displayed in Table 1 and Table 2 were corrected for through multivariate logistic regression in our final data analyses, except for the differences in IHC antibody/protocol, since it was unknown which PD-L1 antibody and/or staining protocol was used in a considerable percentage of cases.

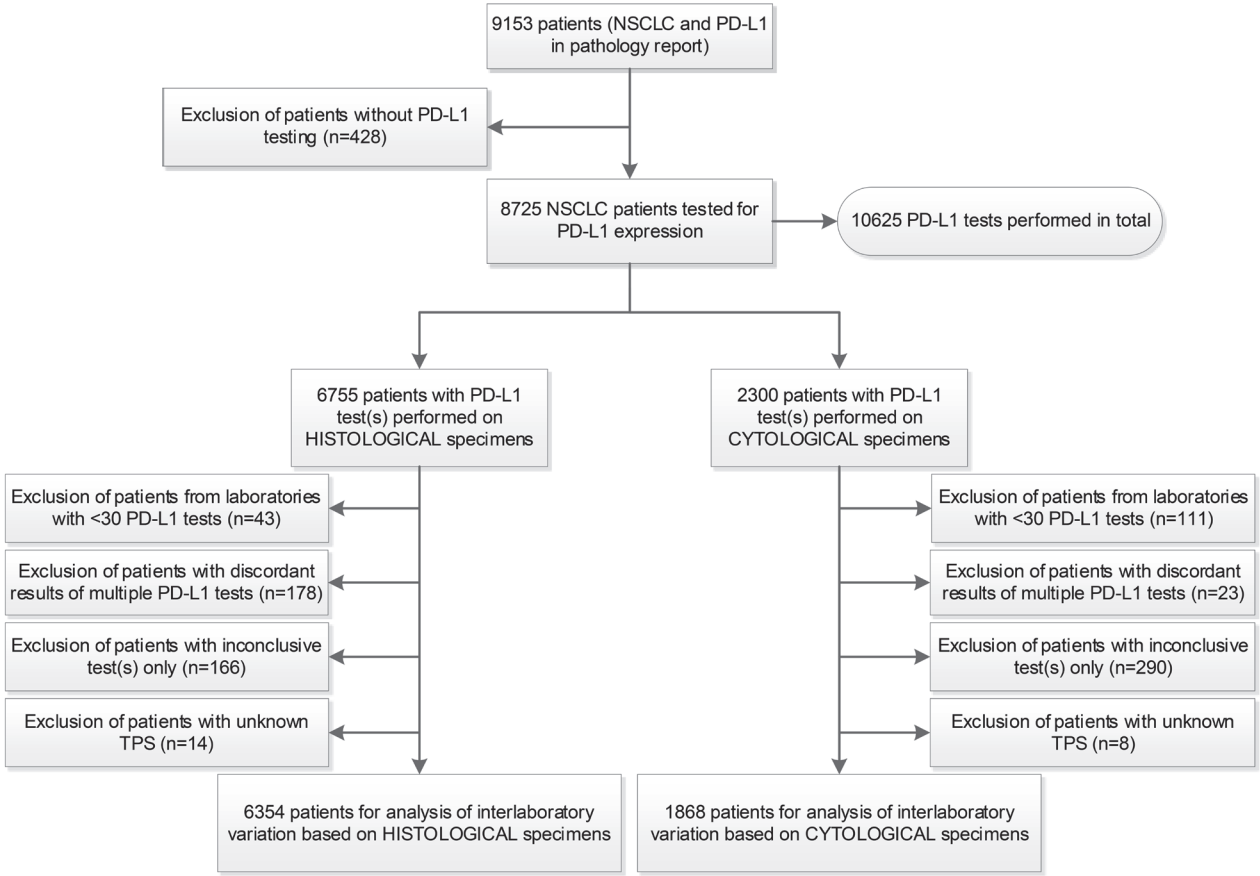


Figure 1. Flowchart of patient selection process. Abbreviations: NSCLC = non-small cell lung cancer; PD-L1 = programmed death-ligand 1; TPS = tumor proportion score.

Table 1. Characteristics for two sets of groups (PD-L1 <1% vs. PD-L1 ≥1% and PD-L1 <50% vs. PD-L1 ≥50%) using data on histological material only.

	Total (n = 6354)	**PD-L1 <1% (n= 2763)**	**PD-L1 ≥1% (n = 3591)**	***p*-value**	**PD-L1 <50% (n= 4387)**	**PD-L1 ≥50% (n = 1967)**	***p*-value**
Age in years (mean (SD))	67.1 (9.6)	67.1 (9.4)	67.1 (9.7)	0.90	67.2 (9.5)	66.9 (9.7)	0.37
Sex							
Male	3498 (55.1%)	1518 (54.9%)	1980 (55.1%)	0.88	2420 (55.2%)	1078 (54.8%)	0.79
Female	2856 (44.9%)	1245 (45.1%)	1611 (44.9%)		1967 (44.8%)	889 (45.2%)	
Histologic subtype							
AC	4139 (65.1%)	1769 (64.0%)	2370 (66.0%)	<0.001	2774 (63.2%)	1365 (69.4%)	<0.001
SCC	1484 (23.4%)	664 (24.0%)	820 (22.8%)		1116 (25.4%)	368 (18.7%)	
NSCLC NOS	504 (7.9%)	196 (7.1%)	308 (8.6%)		322 (7.3%)	182 (9.3%)	
Other	227 (3.6%)	134 (4.8%)	93 (2.6%)		175 (4.0%)	52 (2.6%)	
Source of material							
Primary tumor	3935 (61.9%)	1714 (62.0%)	2221 (61.8%)	<0.001	2769 (63.1%)	1166 (59.3%)	<0.001
Metastasis	1493 (23.5%)	700 (25.3%)	793 (22.1%)		1049 (23.9%)	444 (22.6%)	
Lymph node metastasis	863 (13.6%)	316 (11.4%)	547 (15.2%)		514 (11.7%)	349 (17.7%)	
Other	63 (1.0%)	33 (1.2%)	30 (0.8%)		55 (1.3%)	8 (0.4%)	
Type of material							
Biopsy	5467 (86.0%)	2342 (84.8%)	3125 (87.0%)	0.01	8736 (85.2%)	1731 (88.0%)	<0.01
Surgical resection	887 (14.0%)	421 (15.2%)	466 (13.0%)		651 (14.8%)	236 (12.0%)	
IHC antibody/protocol							
SP263	2084 (32.8%)	882 (31.9%)	1202 (33.5%)	<0.001	1448 (33.0%)	636 (32.3%)	0.06
22C3 LDT	1437 (22.6%)	647 (23.4%)	790 (22.0%)		1010 (23.0%)	427 (21.7%)	
22C3 pharmDx	630 (9.9%)	244 (8.8%)	386 (10.7%)		405 (9.2%)	225 (11.4%)	
22C3 ?	1262 (19.9%)	521 (18.9%)	741 (20.1%)		855 (19.5%)	407 (20.7%)	
Other	155 (2.4%)	76 (2.8%)	79 (2.2%)		110 (2.5%)	45 (2.3%)	
Unknown	786 (12.4%)	393 (14.2%)	393 (10.9%)		559 (12.7%)	227 (11.5%)	

Abbreviations: AC = adenocarcinoma; IHC= immunohistochemistry; LDT = laboratory-developed test; NOS = not otherwise specified; NSCLC = non-small cell lung cancer; PD-L1 = programmed death-ligand 1; SD = standard deviation; SCC = squamous cell carcinoma; ? = IHC protocol unknown.

Table 2. Characteristics for two sets of groups (PD-L1 <1% vs. PD-L1 ≥1% and PD-L1 <50% vs. PD-L1 ≥50%) using data on cytological material only.

	Total (n = 1868)	**PD-L1 <1% (n= 907)**	**PD-L1 ≥1% (n = 961)**	***p*-value**	**PD-L1 <50% (n= 1330)**	**PD-L1 ≥50% (n = 538)**	***p*-value**
Age in years (mean (SD))	66.9 (10.1)	66.9 (10.2)	66.8 (10.1)	0.79	66.9 (10.2)	66.7 (10.1)	0.72
Sex							
Male	998 (53.4%)	523 (57.7%)	475 (49.4%)	<0.001	743 (55.9%)	255 (47.4%)	0.001
Female	870 (46.6%)	384 (42.3%)	486 (50.6%)		587 (44.1%)	283 (52.6%)	
Histologic subtype							
AC	1445 (77.4%)	663 (73.1%)	782 (81.4%)	<0.001	996 (74.9%)	449 (83.3%)	0.001
SCC	243 (13.0%)	141 (15.5%)	102 (10.6%)		193 (14.5%)	50 (9.3%)	
NSCLC NOS	160 (8.6%)	89 (9.8%)	71 (7.4%)		125 (9.4%)	35 (6.5%)	
Other	20 (1.1%)	14 (1.5%)	6 (0.6%)		16 (1.2%)	4 (0.7%)	
Source of material							
Primary tumor	100 (5.4%)	50 (5.5%)	50 (5.2%)	0.23	73 (5.5%)	27 (5.0%)	0.19
Metastasis	128 (6.9%)	69 (7.6%)	59 (6.1%)		90 (6.8%)	38 (7.1%)	
Lymph node metastasis	1047 (56.0%)	520 (57.3%)	527 (54.8%)		735 (55.3%)	312 (58.0%)	
Pleural effusion	494 (26.4%)	219 (24.1%)	275 (28.6%)		351 (26.4%)	143 (26.6%)	
Bronchial brush/fluid	99 (5.3%)	49 (5.4%)	50 (5.2%)		81 (6.1%)	18 (3.3%)	
IHC antibody/protocol							
SP263	483 (25.9%)	274 (30.2%)	209 (21.7%)	<0.001	380 (28.6%)	103 (19.1%)	<0.001
22C3 LDT	464 (24.8%)	228 (25.1%)	236 (24.6%)		326 (24.5%)	138 (25.7%)	
22C3 pharmDx	104 (5.6%)	47 (5.2%)	57 (5.9%)		75 (5.6%)	29 (5.4%)	
22C3 ?	372 (19.9%)	146 (16.1%)	226 (23.5%)		243 (18.3%)	129 (24.0%)	
Other	63 (3.4%)	21 (2.3%)	42 (4.4%)		38 (2.9%)	25 (4.6%)	
Unknown	382 (20.4%)	191 (21.1%)	191 (19.9%)		268 (20.2%)	114 (21.2%)	

Abbreviations: AC = adenocarcinoma; IHC= immunohistochemistry; LDT = laboratory-developed test; NOS = not otherwise specified; NSCLC = non-small cell lung cancer; PD-L1 = programmed death-ligand 1; SD = standard deviation; SCC = squamous cell carcinoma; ? = IHC protocol unknown.

Interlaboratory variation in PD-L1 positivity: Histology

Thirty laboratories performed PD-L1 testing on histological material of ≥ 30 patients. Using the 1% cutoff to determine PD-L1 positivity, the mean national proportion of PD-L1 positive patients was 56.5%. The case mix adjusted positivity rates of the individual laboratories ranged from 32.7% to 71.8%, resulting in a maximum variation of 39.1% between laboratories. Ten (33.3%) laboratories differed significantly from the mean (Figure 2a).

When the 50% cutoff was used to determine PD-L1 positivity, the mean national proportion of PD-L1 positive patients was 31.0%. With this cutoff, the case mix adjusted positivity rates of the individual laboratories ranged from 17.2% to 40.3%, resulting in a maximum variation of 23.1%. Four (13.3%) laboratories differed significantly from the mean (Figure 2b).

When comparing academic and non-academic laboratories in the funnel plots of both cutoffs, there did not appear to be any obvious clustering of academic or non-academic laboratories on one side of the national mean. Of the ten laboratories that differed significantly from the mean at the 1% cutoff, four were academic. In the analysis of the 50% cutoff, two of four laboratories that differed significantly from the mean were academic laboratories.

Interlaboratory variation in PD-L1 positivity: Cytology

Twenty-three laboratories performed PD-L1 testing on cytological material of ≥ 30 patients. When the 1% cutoff was used to determine PD-L1 positivity, the mean national proportion of PD-L1 positive patients was 51.4%. Individual laboratory case mix adjusted positivity rates ranged from 32.2% to 73.4%, resulting in a maximum variation of 41.2%. Eight (34.8%) laboratories differed significantly from the mean (Figure 3a).

Using the 50% cutoff to determine PD-L1 positivity, the mean national proportion of PD-L1 positive patients was 28.8%. This time, case mix adjusted positivity rates of individual laboratories ranged from 14.7% to 43.9%, resulting in a maximum variation of 29.2%. Again, eight (34.8%) laboratories differed significantly from the mean (Figure 3b).

A comparison of academic and non-academic laboratories in the funnel plots of both cutoffs showed no obvious clustering of academic or non-academic laboratories on one side of the national mean. Of the eight laboratories that differed significantly from the mean at the 1% cutoff, three were academic. In the analysis of the 50% cutoff, four of eight laboratories that differed significantly from the mean were academic laboratories.

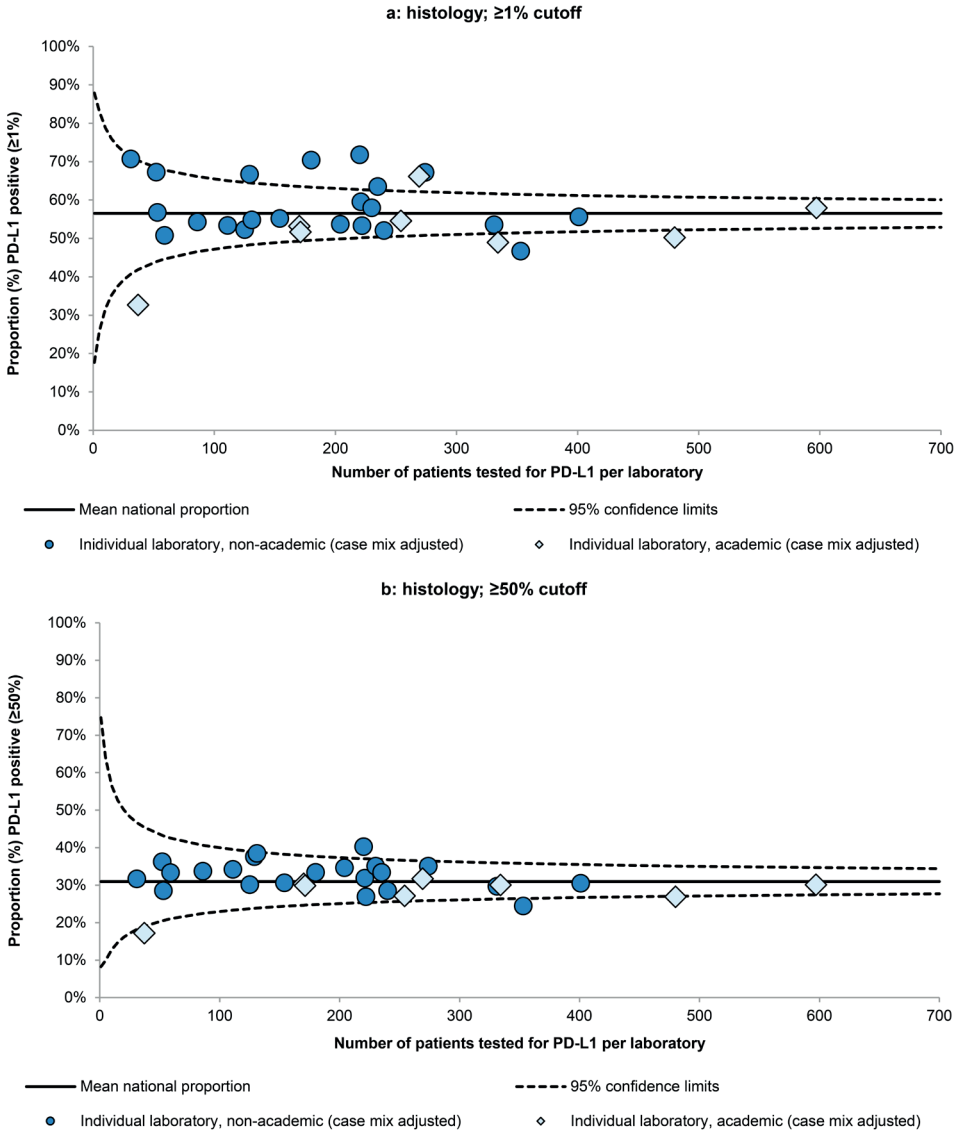


Figure 2. Funnel plots showing interlaboratory variation in PD-L1 positivity, based on histology data and using either a 1% cutoff (**a**) or a 50% cutoff (**b**) to determine PD-L1 positivity. Case mix adjusted positivity rates are displayed for each laboratory, plotted against the total number of patients tested for PD-L1. The black line shows the mean national proportion of PD-L1 positive patients, surrounded by its 95% confidence limits (dotted lines).

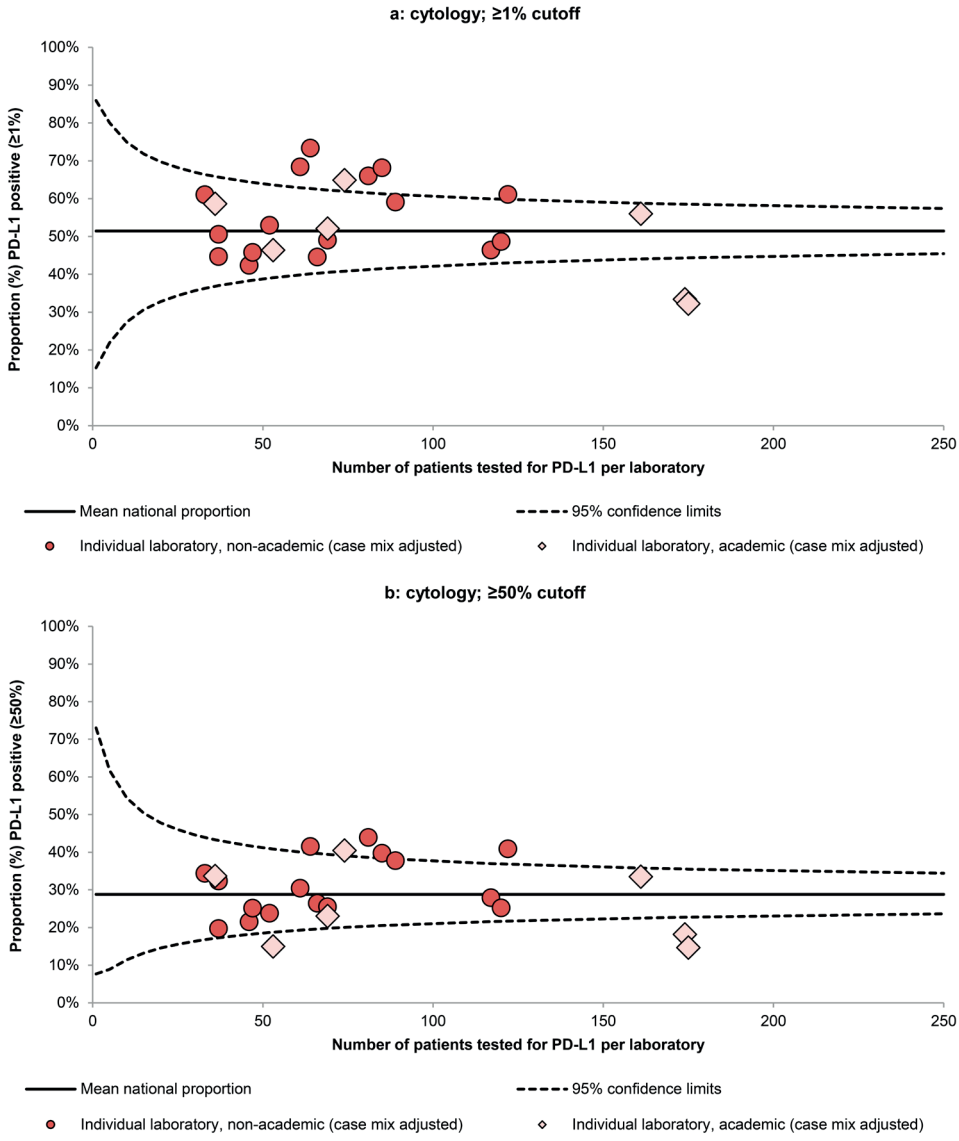


Figure 3. Funnel plots showing interlaboratory variation in PD-L1 positivity, based on cytology data and using either a 1% cutoff (**a**) or a 50% cutoff (**b**) to determine PD-L1 positivity. Case mix adjusted positivity rates are displayed for each laboratory, plotted against the total number of patients tested for PD-L1. The black line shows the mean national proportion of PD-L1 positive patients, surrounded by its 95% confidence limits (dotted lines).

Comparing laboratories that differ significantly from the mean

All 23 laboratories that were included in the analysis of cytology data, were also included in the analysis of histology data. When using the 1% cutoff to determine PD-L1 positivity, only two of these 23 laboratories differed significantly from the mean in both the analysis of histology data and the analysis of cytology data. Using the 50% cutoff, only one laboratory differed significantly from the mean in both analyses. Seven laboratories were included in the analysis of histology data only. Three of these seven laboratories differed significantly from the mean in the analysis of the 1% cutoff, and one differed significantly from the mean in the analysis of the 50% cutoff.

PD-L1 IHC antibodies and protocols

In order to analyze the role that use of different PD-L1 IHC antibodies and protocols (commercial assay or LDT) might play in causing interlaboratory variation in PD-L1 positivity, information on PD-L1 IHC antibodies and protocols used by individual laboratories was incorporated in the funnel plots for histological material (see Supplementary Figure 1a and b). There did not appear to be any obvious clustering of specific IHC antibodies on one side of the national mean and the laboratories that differed significantly from the mean used various antibodies. Additionally, no obvious differences were observed between commercial assays and LDTs.

PD-L1 positivity rates that were adjusted for case mix with inclusion of the variable IHC antibody/protocol ranged from 33.2% to 68.4% at the 1% cutoff, resulting in a maximum variation of 35.2% between laboratories. This range was smaller than the range in PD-L1 positivity rates adjusted for case mix without IHC antibody/protocol (maximum variation 35.2% vs. 39.1%). With the inclusion of IHC antibody/protocol in case mix adjustment, nine (30.0%) laboratories differed significantly from the mean, which is one less than in the analysis excluding IHC antibody/protocol from case mix adjustment. From the laboratories with known PD-L1 antibody and protocol, inclusion of IHC antibody/protocol in case mix adjustment showed the largest effect on laboratories that used the commercial 22C3 pharmDx assay (Supplementary Figure 2a).

At the 50% cutoff, PD-L1 positivity rates adjusted for case mix with inclusion of the variable IHC antibody/protocol ranged from 17.3% to 38.4%, resulting in a maximum variation of 21.1%. There was slightly less variation compared to the analysis of PD-L1 positivity rates adjusted for case mix not including IHC antibody/protocol (maximum variation 21.1% vs. 23.1%). With the inclusion of IHC antibody/protocol in case mix adjustment, only two (6.7%) laboratories differed significantly from the mean, which was two less than in the analysis excluding IHC antibody/protocol from case mix adjustment. Similar to the analysis at the 1% cutoff, inclusion of IHC antibody/protocol in case mix adjustment showed the largest effect on laboratories that used the commercial 22C3 pharmDx assay (Supplementary Figure 2b).

DISCUSSION

In this nationwide cohort of NSCLC patients, using real-world clinical pathology data, a considerable amount of variation in PD-L1 positivity was found between laboratories. The amount of variation was largest when a 1% cutoff was used to determine PD-L1 positivity. When using a 50% cutoff, use of cytological material for PD-L1 testing also resulted in a substantial amount of variation. Use of the 50% cutoff to determine PD-L1 positivity on histological material resulted in the smallest amount of interlaboratory variation in PD-L1 positivity.

With regard to the analysis of histology data, the difference in amount of variation between the 1% and the 50% cutoff is prominent. The degree of variation seen at the 50% cutoff is a lot smaller compared to the 1% cutoff. Most likely, the higher amount of variation seen when using the 1% cutoff is caused to a large extent by a higher degree of interpathologist variation at this cutoff compared to the 50% cutoff. Several studies have, in fact, demonstrated lower concordance levels of PD-L1 scoring on TCs between pathologists at the 1% cutoff in comparison to the 50% cutoff[20, 21, 22, 23, 24, 25, 26]. Apparently, determining PD-L1 positivity using a 1% cutoff is harder than determining PD-L1 positivity at the 50% cutoff level. When some pathologists have a tendency to score dubious cases with a TPS that lies around 1% as PD-L1 positive, while other pathologists are more conservative in their PD-L1 scoring, this could easily lead to a high degree of variation in PD-L1 positivity rates between pathology laboratories. Subsequently, this would influence treatment decisions and possibly outcome for individual NSCLC patients.

It is remarkable that PD-L1 testing was performed on cytological material in a considerable number of cases, even though PD-L1 immunostaining on cytological specimens was not validated in clinical trials and only a limited number of studies had assessed histologic-cytologic correlation of PD-L1 immunohistochemistry during our study period[26, 27, 28, 29]. Interestingly, use of cytological material not only resulted in a substantial amount of interlaboratory variation in PD-L1 positivity at the 1% cutoff, but also at the 50% cutoff. Moreover, most of the laboratories differing significantly from the mean in analysis of cytology data were different laboratories than the ones that differed significantly from the mean in analysis of histology data. This suggests that, besides interpathologist variation in scoring of PD-L1 TPS, use of cytological material may also contribute to interlaboratory variation in PD-L1 positivity. We hypothesize that this is most likely a result of the previously mentioned differences in processing of cytology[12, 13, 14]. This may include the use of fixatives other than formalin, such as alcohol-based fixatives, which have been shown to negatively influence immunoreactivity of various IHC antibodies[30, 31, 32, 33, 34]. Use of CytoLyt, a methanol-based fixative, has been shown to negatively affect PD-L1 immunostaining[35]. When laboratories use this fixative on cytological material, this might result in a PD-L1

positivity rate that is substantially lower than that of laboratories that use formalin fixation, for instance. Unfortunately, data extracted from the PALGA database do not contain any information about processing of cytological specimens. Analysis of interlaboratory variation in PD-L1 positivity rate in relation to different methods of processing of cytology of each laboratory, could create more insight into the influence of different processing methods on PD-L1 variation.

Even though the processing of histology samples often shows more similarities between different laboratories than the processing of cytology samples, it is still imaginable that differences in pre-analytical variables in the acquisition and processing of histological material could contribute to interlaboratory variation in PD-L1 positivity. It has been suggested, for instance, that biopsy size and the number of biopsies taken can influence PD-L1 results, with small samples more often resulting in an underestimation of PD-L1 expression[36]. This could potentially lead to variation in PD-L1 positivity between laboratories when some laboratories use smaller core biopsy needles and/or a smaller number of biopsies in their standard biopsy procedures compared to other laboratories.

An analytical factor that might play a role in causing interlaboratory variation in PD-L1 positivity could be the use of different PD-L1 antibodies and protocols within the various laboratories. In our study, funnel plots incorporating information on PD-L1 antibody and protocol per laboratory showed no clustering of specific antibodies or protocols or use of one specific antibody by the laboratories that differed significantly from the mean, which seems promising. Nevertheless, inclusion of the variable IHC antibody/protocol in case mix adjustment resulted in a reduction of maximum variation in PD-L1 positivity between laboratories compared to the analysis that did not include IHC antibody/protocol in case mix adjustment at both cutoffs, although the reduction in maximum variation was small. Adding IHC antibody/protocol to case mix adjustment also resulted in a reduction of the number of laboratories that differed significantly from the mean at both cutoffs. This suggests that use of different PD-L1 antibodies and protocols might have some influence on interlaboratory variation in PD-L1 positivity. It is important to interpret these analyses with caution, since information on specific PD-L1 antibody and protocol or on protocol only was lacking in more than a quarter of laboratories. One study that also used real-world data on PD-L1 testing[37], showed no statistically significant difference in PD-L1 expression between commercial assays 22C3 and 28–8. Yet, the researchers did find a statistically significant difference at the 50% cutoff level between these two commercial assays and all LDTs grouped together, with the LDTs showing more PD-L1 negative results. Another study showed substantial interlaboratory concordance of PD-L1 staining for various commercial assays, but only moderate concordance for LDTs compared to commercial assays[38]. In contrast, a meta-analysis of diagnostic accuracy of PD-L1 IHC assays concluded that properly designed LDTs may in fact achieve higher accuracy than commercial PD-L1 assays, when

both are compared to an appropriate reference standard[11]. Various other studies have shown substantial interlaboratory concordance of PD-L1 staining for several commercial PD-L1 assays[22, 39, 40, 41], while another study stated that equivalence of commercial PD-L1 assays at the 1% and 50% cutoff cannot be assumed[42]. Lastly, a study by Butter et al.[43] showed a similar degree of interlaboratory concordance between laboratories using a 22C3 LDT and laboratories using the 22C3 pharmDx commercial assay (Agilent), but also concluded that interlaboratory variability of immunostaining contributes to discrepancies in PD-L1 positivity between centers. Unfortunately, based on the data in our study, no definitive conclusions can be drawn regarding the actual influence of different antibodies and protocols on interlaboratory variation in PD-L1 positivity in the real-life clinical setting.

A large degree of variation between laboratories, such as seen in our study, is problematic. After all, this implies that a patient could receive different PD-L1 test results depending on the pathology laboratory where his or her material is tested. In turn, this could result in different courses of treatment, and may subsequently influence outcome of individual NSCLC patients. Variation at the 50% cutoff may be especially problematic, since this cutoff is used across the world to differentiate between treatment with immunotherapy alone or a more toxic treatment regimen of immunotherapy combined with chemotherapy for patients with metastatic NSCLC[1, 44, 45]. With this study, we hope to raise awareness among pathologists, but also among pulmonologists, of the existence of interlaboratory variation in PD-L1 positivity. In order to further decrease variation between laboratories, one could think of various measures pathologists could take, such as double reading of difficult cases. Also, laboratories might decide to let all PD-L1 staining be scored by properly trained and experienced pathologists only, since it has been shown that training for PD-L1 scoring and experience in routine pathology practice correlate with higher interpathologists concordance[46]. Unfortunately, our data do not include information about individual laboratories' expertise in scoring PD-L1 on material from NSCLC patients, hence we cannot draw conclusions on the correlation between laboratories' level of experience and the amount of interlaboratory variation in PD-L1. In future, digital image analysis for PD-L1 scoring might improve interobserver concordance[47], although this needs to be studied more extensively before implementation into clinical practice. Parallel to the previously mentioned measures, which mainly focused on reducing interobserver variability, a reduction of technical differences between laboratories might also contribute to a decrease in interlaboratory variation. We do not know how many of the laboratories included in our study participated in PD-L1 external quality assessment (EQA) schemes during the study period. Potentially, participation of all laboratories in such EQA schemes could help in reducing interlaboratory variation in PD-L1 positivity[48]. Also, when LDTs are used for PD-L1 staining, it is paramount that these LDTs are optimized and validated sufficiently[49]. This may not always be the case in all laboratories, which may suggest that use of commercial assays could also contribute to reduction of interlaboratory variability

in PD-L1 positivity. Nevertheless, this remains uncertain, since even laboratories that use the same commercial assay can produce differences in PD-L1 staining results[38, 43] and others have reported inequality of commercial assays at the 1% and 50% cutoff[42]. To help create more awareness among pathologists, results from individual laboratories in our study were sent back to these laboratories as feedback reports. In this way, pathologists are encouraged to discuss and reflect on their own results concerning PD-L1 testing, compared to other laboratories in the Netherlands, and to think of ways to improve their own PD-L1 testing practices.

Our study has some limitations. Notably, some variables were unknown in a large number of patients, which restricted the analyses that we could perform. This included antibody or specific IHC protocol (commercial assay or LDT) used for PD-L1 testing, as discussed above. Variables such as mutational status, smoking status, and stage of disease were unknown in many or all patients, and could therefore not be included in the multivariate logistic regression model for case mix correction either. It would have strengthened our study if we could have included these variables in our logistic regression model. However, the association between mutation status, such as *KRAS* and *EGFR* mutation status, and PD-L1 expression is still controversial, with various studies showing opposing results[50, 51, 52]. Also, while some studies show a significant correlation between smoking status and PD-L1[53], others do not[54, 55]. With regard to stage of disease, various studies have shown a positive correlation between higher stages of disease and high PD-L1 expression[53, 54]. Since PD-L1 expression holds clinical implications for stage III and stage IV NSCLC patients, most of the patients in our data set would have most likely had stage III or stage IV disease. It is thus unlikely that interlaboratory differences in mutation status, smoking status, or stage of disease could have completely explained the high degree of variation in PD-L1 positivity that we found. Another limitation is the lack of information on the number of pathologists scoring PD-L1 expression per laboratory. While this information could have provided valuable insight into the amount of interpathologist variation in PD-L1 positivity rate within each laboratory, we still feel that the insight into the variation in PD-L1 positivity on an interlaboratory level is valuable on its own. To our knowledge, no other studies have provided these kinds of analyses on such a large scale using real-world clinical pathology data. Finally, the influence of different methods of processing of cytology material on interlaboratory variation in PD-L1 positivity could not be studied, as information on processing of cytological specimens was not part of the PALGA database. We intend to retrieve this information in another way, such as through questionnaires sent out to laboratories, so that we will be able to study the relationship between processing of cytological specimens and interlaboratory variation in PD-L1 positivity in the future.

In conclusion, we have shown that in a real-world setting, a considerable amount of variation in PD-L1 positivity rates exists between pathology laboratories on a nationwide level. Most

likely, this is caused to a large extent by discordance between pathologists at the 1% cutoff point. Potentially, various analytical or pre-analytical factors, such as differences in processing of cytological material between laboratories, may contribute to interlaboratory variation in PD-L1 positivity as well. It is important that the amount of variation between laboratories is reduced, since a high degree of variation could result in patients receiving a different course of treatment when PD-L1 is assessed in the one laboratory compared to another. Both pathologists and pulmonologists should be made aware of this risk, and work together to try to reduce the amount of interlaboratory variation in PD-L1 positivity.

ACKNOWLEDGEMENTS

We would like to thank Marion Visscher, Eeke Leerssen and Jasmijn van Es for their help in extracting data from the PALGA dataset.

This work was supported by AstraZeneca, MSD and Roche (grant numbers not applicable). None of the grant suppliers were involved in the study design, collection, analysis and interpretation of data, writing of the manuscript, or the decision to submit the manuscript for publication.

REFERENCES

1. Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(Supplement_4):iv192-iv237.
2. European Society for Medical Oncology (ESMO). Metastatic Non-Small-Cell Lung Cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up, 2019. Available from: https://www.esmo.org/Guidelines/Lung-and-Chest-Tumours/Metastatic-Non-Small-Cell-Lung-Cancer (Accessed 25 November 2019).
3. EMA. Imfinzi (Durvalumab), 2018. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/imfinzi (Accessed 25 November 2019).
4. Kuijpers CC, Sluijter CE, von der Thusen JH, et al. Interlaboratory Variability in the Histologic Grading of Colorectal Adenocarcinomas in a Nationwide Cohort. Am J Surg Pathol. 2016;40(8):1100-1108.
5. Kuijpers CC, Sluijter CE, von der Thusen JH, et al. Interlaboratory variability in the grading of dysplasia in a nationwide cohort of colorectal adenomas. Histopathology. 2016;69(2):187-197.
6. van Dooijeweert C, van Diest PJ, Willems SM, Kuijpers C, Overbeek LIH, Deckers IAG. Significant inter- and intra-laboratory variation in grading of ductal carcinoma in situ of the breast: a nationwide study of 4901 patients in the Netherlands. Breast Cancer Res Treat. 2019;174(2):479-488.
7. van Dooijeweert C, van Diest PJ, Willems SM, et al. Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. Int J Cancer. 2019.
8. Koomen BM, Badrising SK, van den Heuvel MM, Willems SM. Comparability of PD-L1 immunohistochemistry assays for non-small-cell lung cancer: a systematic review. Histopathology. 2020;76(6):793-802.
9. van Dooijeweert C, Deckers IAG, Baas IO, van der Wall E, van Diest PJ. Hormone- and HER2-receptor assessment in 33,046 breast cancer patients: a nationwide comparison of positivity rates between pathology laboratories in the Netherlands. Breast Cancer Res Treat. 2019;175(2):487-497.
10. Kerr KM, Hirsch FR. Programmed Death Ligand-1 Immunohistochemistry: Friend or Foe? Arch Pathol Lab Med. 2016;140(4):326-331.
11. Torlakovic E, Lim HJ, Adam J, et al. "Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Mod Pathol. 2019. DOI: 10.1038/s41379-019-0327-4.
12. Vigliar E, Malapelle U, Iaccarino A, et al. PD-L1 expression on routine samples of non-small cell lung cancer: results and critical issues from a 1-year experience of a centralised laboratory. J Clin Pathol. 2019;72(6):412-417.
13. Nambirajan A, Jain D. Cell blocks in cytopathology: An update. Cytopathology. 2018;29(6):505-524.
14. Jain D, Mathur SR, Iyer VK. Cell blocks in cytopathology: a review of preparative methods, utility in diagnosis and role in ancillary studies. Cytopathology. 2014;25(6):356-371.
15. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol. 2007;29(1):19-24.
16. Netherlands Patients Federation. Rapport Delen van uw data [in Dutch]. 2019. Available from: https://www.patientenfederatie.nl/downloads/rapporten/154-rapport-delen-van-je-data/file (Accessed 12 November 2020).
17. Sheng J, Fang W, Yu J, et al. Expression of programmed death ligand-1 on tumor cells varies pre and post chemotherapy in non-small cell lung cancer. Sci Rep. 2016;6:20090.
18. Rojko L, Reiniger L, Teglasi V, et al. Chemotherapy treatment is associated with altered PD-L1 expression in

lung cancer patients. J Cancer Res Clin Oncol. 2018;144(7):1219-1226.

19. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med. 2005;24(8):1185-1202.
20. Chan AWH, Tong JHM, Kwan JSH, et al. Assessment of programmed cell death ligand-1 expression by 4 diagnostic assays and its clinicopathological correlation in a large cohort of surgical resected non-small cell lung carcinoma. Mod Pathol. 2018;31(9):1381-1390.
21. Brunnström H, Johansson A, Westbom-Fremer S, et al. PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: Inter-pathologist variability is higher than assay variability. Modern Pathology. 2017;30(10):1411-1421.
22. Marchetti A, Barberis M, Franco R, et al. Multicenter Comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) Assays to Test PD-L1 Expression for NSCLC Patients to Be Treated with Immune Checkpoint Inhibitors. J Thorac Oncol. 2017;12(11):1654-1663.
23. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between Programmed Cell Death Ligand-1 Diagnostic Assays across Multiple Protein Expression Cutoffs in Non-Small Cell Lung Cancer. Clin Cancer Res. 2017;23(14):3585-3591.
24. Rimm DL, Han G, Taube JM, et al. A Prospective, Multi-institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. JAMA Oncol. 2017;3(8):1051-1058.
25. Tsao MS, Kerr KM, Kockx M, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of Blueprint phase 2 project. J Thorac Oncol. 2018. DOI: 10.1016/j.jtho.2018.05.013.
26. Russell-Goldman E, Kravets S, Dahlberg SE, Sholl LM, Vivero M. Cytologic-histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. Cancer Cytopathol. 2018;126(4):253-263.
27. Skov BG, Skov T. Paired Comparison of PD-L1 Expression on Cytologic and Histologic Specimens From Malignancies in the Lung Assessed With PD-L1 IHC 28-8pharmDx and PD-L1 IHC 22C3pharmDx. Appl Immunohistochem Mol Morphol. 2017;25(7):453-459.
28. Sakakibara R, Inamura K, Tambo Y, et al. EBUS-TBNA as a Promising Method for the Evaluation of Tumor PD-L1 Expression in Lung Cancer. Clin Lung Cancer. 2017;18(5):527-534 e521.
29. Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death ligand 1 antibody to determine programmed death ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. Cancer Cytopathol. 2018. DOI: 10.1002/cncy.21977.
30. Buonocore DJ, Konno F, Jungbluth AA, et al. CytoLyt fixation significantly inhibits MIB1 immunoreactivity whereas alternative Ki-67 clone 30-9 is not susceptible to the inhibition: Critical diagnostic implications. Cancer Cytopathol. 2019. DOI: 10.1002/cncy.22170.
31. Gorman BK, Kosarac O, Chakraborty S, Schwartz MR, Mody DR. Comparison of breast carcinoma prognostic/predictive biomarkers on cell blocks obtained by various methods: Cellient, formalin and thrombin. Acta Cytol. 2012;56(3):289-296.
32. Gruchy JR, Barnes PJ, Dakin Hache KA. CytoLyt(R) fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. Appl Immunohistochem Mol Morphol. 2015;23(4):297-302.
33. Hanley KZ, Birdsong GG, Cohen C, Siddiqui MT. Immunohistochemical detection of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 expression in breast carcinomas:

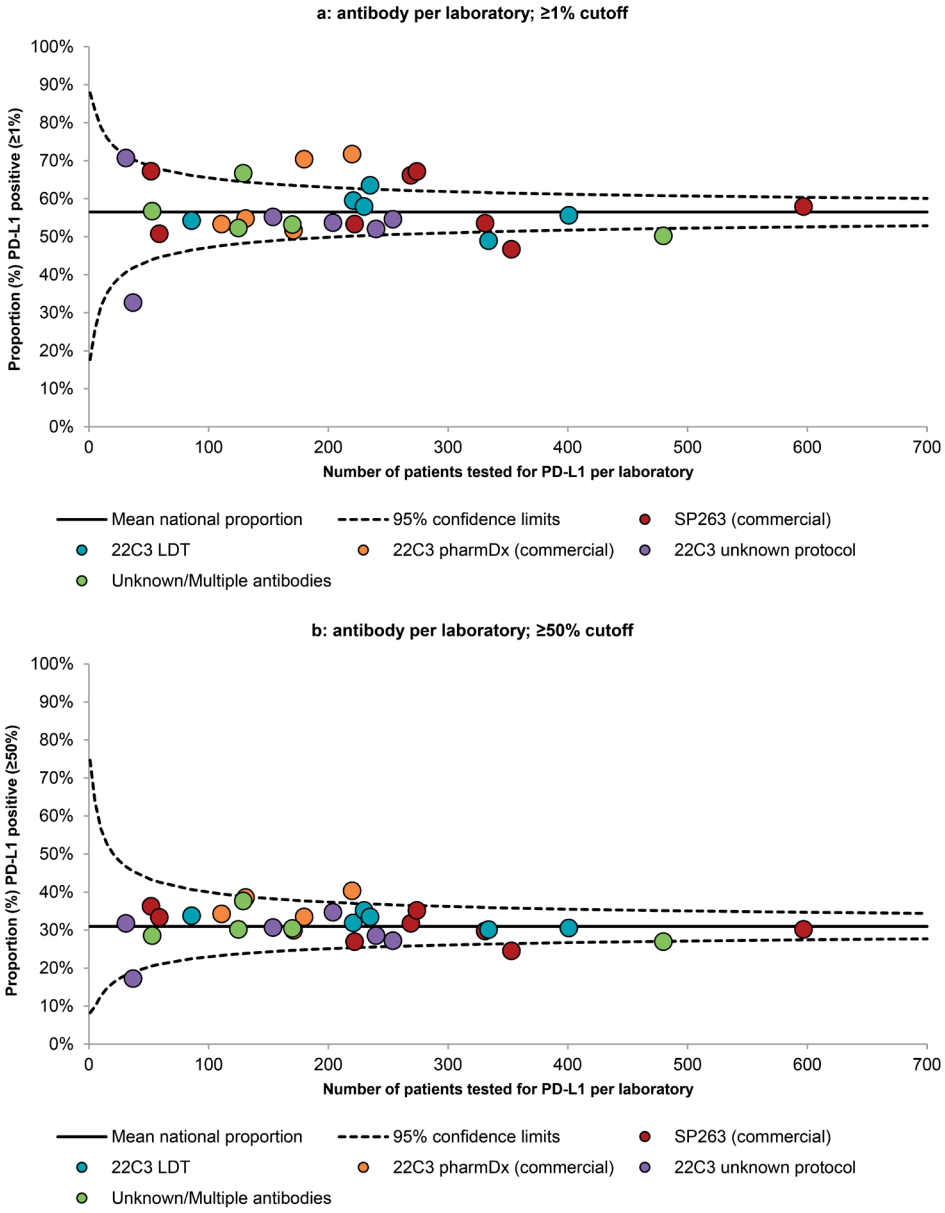
comparison on cell block, needle-core, and tissue block preparations. Cancer. 2009;117(4):279-288.

34. Sauter JL, Grogg KL, Vrana JA, Law ME, Halvorson JL, Henry MR. Young investigator challenge: Validation and optimization of immunohistochemistry protocols for use on cellient cell block specimens. Cancer Cytopathol. 2016;124(2):89-100.
35. Lloyd IE, Zhou W, Witt BL, Chadwick BE. Characterization of PD-L1 Immunohistochemical Expression in Cell Blocks With Different Specimen Fixation and Processing Methods. Appl Immunohistochem Mol Morphol. 2019;27(2):107-113.
36. Thunnissen E, Kerr KM, Dafni U, et al. Programmed death-ligand 1 expression influenced by tissue sample size. Scoring based on tissue microarrays' and cross-validation with resections, in patients with, stage I-III, non-small cell lung carcinoma of the European Thoracic Oncology Platform Lungscape cohort. Mod Pathol. 2020;33(5):792-801.
37. Velcheti V, Patwardhan PD, Liu FX, Chen X, Cao X, Burke T. Real-world PD-L1 testing and distribution of PD-L1 tumor expression by immunohistochemistry assay type among patients with metastatic non-small cell lung cancer in the United States. PLoS One. 2018;13(11):e0206370.
38. Scheel AH, Baenfer G, Baretton G, et al. Interlaboratory concordance of PD-L1 immunohistochemistry for non-small-cell lung cancer. Histopathology. 2018;72(3):449-459.
39. Adam J, Le Stang N, Rouquette I, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. Ann Oncol. 2018;29(4):953-958.
40. Roach C, Zhang N, Corigliano E, et al. Development of a Companion Diagnostic PD-L1 Immunohistochemistry Assay for Pembrolizumab Therapy in Non-Small-cell Lung Cancer. Appl Immunohistochem Mol Morphol. 2016;24(6):392-397.
41. Vennapusa B, Baker B, Kowanetz M, et al. Development of a PD-L1 Complementary Diagnostic Immunohistochemistry Assay (SP142) for Atezolizumab. Appl Immunohistochem Mol Morphol. 2019;27(2):92-100.
42. Dodson A, Parry S, Lissenberg-Witte B, et al. External quality assessment demonstrates that PD-L1 22C3 and SP263 assays are systematically different. J Pathol Clin Res. 2020;6(2):138-145.
43. Butter R, t Hart NA, Hooijer GKJ, et al. Multicentre study on the consistency of PD-L1 immunohistochemistry as predictive test for immunotherapy in non-small cell lung cancer. J Clin Pathol. 2020;73(7):423-430.
44. Wu YL, Planchard D, Lu S, et al. Pan-Asian adapted Clinical Practice Guidelines for the management of patients with metastatic non-small-cell lung cancer: a CSCO-ESMO initiative endorsed by JSMO, KSMO, MOS, SSO and TOS. Ann Oncol. 2019;30(2):171-210.
45. National Cancer Institute. Non-Small Cell Lung Cancer Treatment (PDQ®)–Health Professional Version, 2021. Available from: https://www.cancer.gov/types/lung/hp/non-small-cell-lung-treatment-pdq#_48406 (Accessed 7 June 2021).
46. Adam J, Hofman V, Mansuet-Lupo A, et al. P2.09-17 Real-World Concordance Across Pathologists for PD-L1 Scoring in Non-Small Cell Lung Cancer: Results from a Large Nationwide Initiative. Journal of Thoracic Oncology. 2019;14(10):S775.
47. Koelzer VH, Gisler A, Hanhart JC, et al. Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. Histopathology. 2018;73(3):397-406.
48. Keppens C, Dequeker EM, Pauwels P, Ryska A, t Hart N, von der Thusen JH. PD-L1 immunohistochemistry in

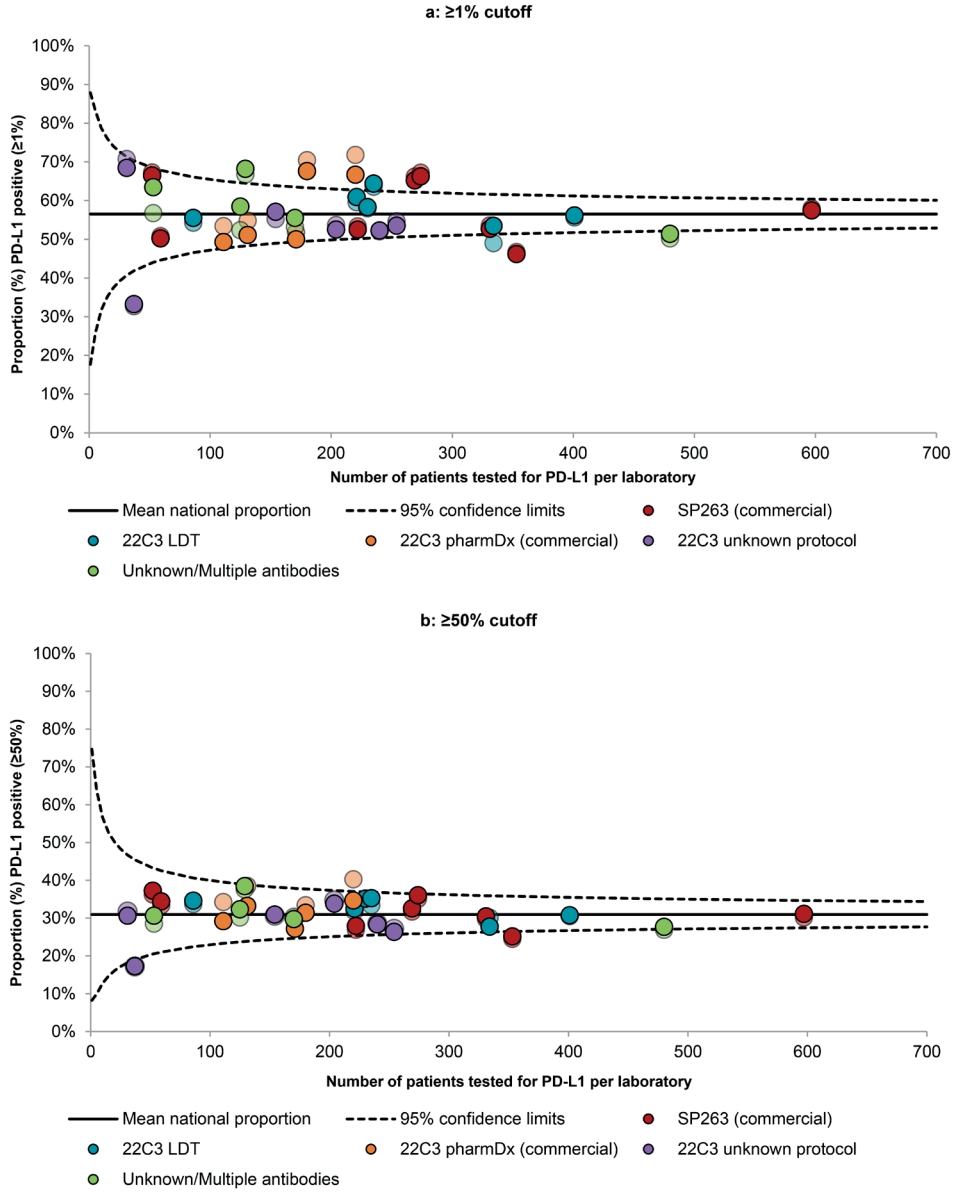
non-small-cell lung cancer: unraveling differences in staining concordance and interpretation. Virchows Arch. 2020. DOI: 10.1007/s00428-020-02976-5.

49. Thunnissen E. How to Validate Predictive Immunohistochemistry Testing in Pathology? A Practical Approach Exploiting the Heterogeneity of Programmed Death Ligand-1 Present in Non-Small Cell Lung Cancer. Arch Pathol Lab Med. 2019;143(1):11-12.
50. Li D, Zhu X, Wang H, Li N. Association between PD-L1 expression and driven gene status in NSCLC: A meta-analysis. Eur J Surg Oncol. 2017;43(7):1372-1379.
51. Petrelli F, Maltese M, Tomasello G, et al. Clinical and Molecular Predictors of PD-L1 Expression in Non-Small-Cell Lung Cancer: Systematic Review and Meta-analysis. Clin Lung Cancer. 2018;19(4):315-322.
52. Zhang M, Li G, Wang Y, et al. PD-L1 expression in lung cancer and its correlation with driver mutations: a meta-analysis. Sci Rep. 2017;7(1):10255.
53. Song P, Guo L, Li W, Zhang F, Ying J, Gao S. Clinicopathologic Correlation With Expression of PD-L1 on Both Tumor Cells and Tumor-infiltrating Immune Cells in Patients With Non-Small Cell Lung Cancer. J Immunother. 2019;42(1):23-28.
54. Skov BG, Rorvig SB, Jensen THL, Skov T. The prevalence of programmed death ligand-1 (PD-L1) expression in non-small cell lung cancer in an unselected, consecutive population. Mod Pathol. 2020;33(1):109-117.
55. Ye L, Leslie C, Jacques A, Mesbah Ardakani N, Amanuel B, Millward M. Programmed death ligand-1 expression in non-small cell lung cancer in a Western Australian population and correlation with clinicopathologic features. Mod Pathol. 2019;32(4):524-531.

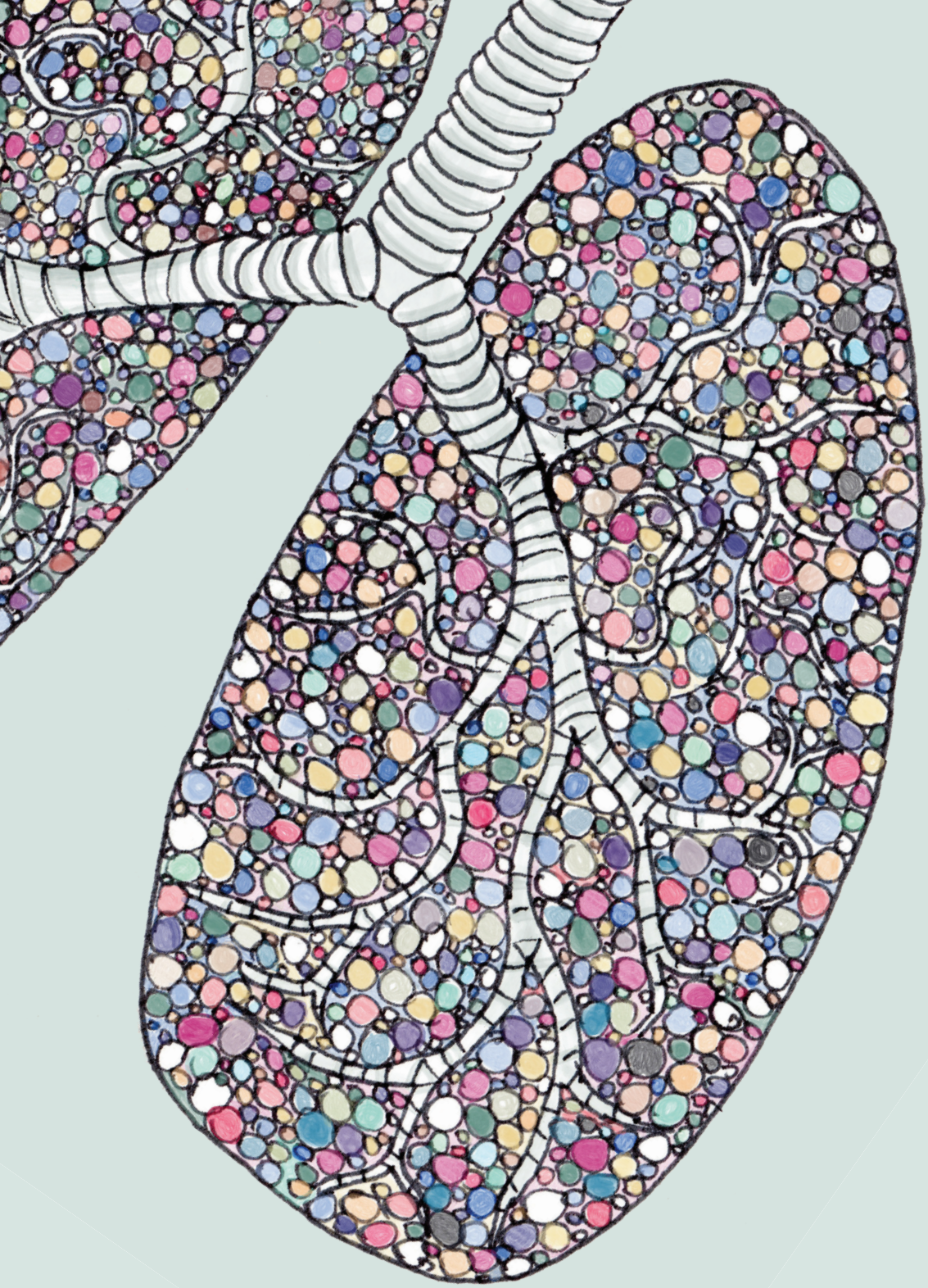
SUPPLEMENTARY DATA



Supplementary figure 1. Funnel plots showing interlaboratory variation in PD-L1 positivity (based on histology data), with incorporation of PD-L1 antibody and protocol used per laboratory. PD-L1 positivity was determined using either a 1% cutoff (**a**) or a 50% cutoff (**b**). Case-mix adjusted positivity rates are displayed for each laboratory (dots), plotted against the total number of patients tested for PD-L1. Color of the dots indicates the PD-L1 antibody used by the laboratory (see legend). The black line shows the mean national proportion of PD-L1 positive patients, surrounded by its 95% confidence limits (dotted lines).



Supplementary figure 2. Funnel plots showing interlaboratory variation in PD-L1 positivity (based on histology data) and displaying case mix adjusted PD-L1 positivity rates with and without incorporation of the variable IHC antibody/protocol. PD-L1 positivity was determined using either a 1% cutoff (**a**) or a 50% cutoff (**b**). Case-mix adjusted positivity rates are displayed for each laboratory (dots), plotted against the total number of patients tested for PD-L1. Positivity rates that were adjusted for case-mix including IHC antibody/protocol are displayed as non-transparent dots, while positivity rates adjusted for case-mix without IHC antibody/protocol are displayed as semi-transparent dots. Color of the dots indicates the PD-L1 antibody used by the laboratory (see legend). The black line shows the mean national proportion of PD-L1 positive patients, surrounded by its 95% confidence limits (dotted lines).



6

NATIONWIDE DIFFERENCES IN CYTOLOGY FIXATION AND PROCESSING METHODS AND THEIR IMPACT ON INTERLABORATORY VARIATION IN PD-L1 POSITIVITY

Koomen BM
de Boer M
van Dooijeweert C
van Lindert ASR
Deckers IAG
Voorham QJM
Willems SM

Virchows Archiv. 2023;482(4):702-720

ABSTRACT

Programmed death-ligand 1 (PD-L1) immunostaining, which aids clinicians in decision-making on immunotherapy for non-small cell lung cancer (NSCLC) patients, is sometimes performed on cytological specimens. In this study, differences in cytology fixation and cell block (CB) processing between pathology laboratories were assessed, and the influence of these differences on interlaboratory variation in PD-L1 positivity was investigated. Questionnaires on cytology processing were sent to all Dutch laboratories. Information gathered from the responses was added to data on all Dutch NSCLC patients with a mention of PD-L1 testing in their cytopathology report from July 2017 to December 2018, retrieved from PALGA (the nationwide network and registry of histo- and cytopathology in the Netherlands). Case mix-adjusted PD-L1 positivity rates were determined for laboratories with known fixation and CB method. The influence of differences in cytology processing on interlaboratory variation in PD-L1 positivity was assessed by comparing positivity rates adjusted for differences in the variables fixative and CB method with positivity rates not adjusted for differences in these variables. Twenty-eight laboratories responded to the survey and reported 19 different combinations of fixation and CB method. Interlaboratory variation in PD-L1 positivity was assessed in 19 laboratories. Correcting for differences in the fixative and CB method resulted in a reduction (from eight (42.1%) to five (26.3%)) in the number of laboratories that differed significantly from the mean in PD-L1 positivity. Substantial variation in cytology fixation and CB processing methods was observed between Dutch pathology laboratories, which partially explains the existing considerable interlaboratory variation in PD-L1 positivity.

INTRODUCTION

Globally, lung cancer is one of the most frequent forms of cancer, with more than 2.2 million new cases in 2020, accounting for 11.4% of all cancer cases worldwide[1]. Lung cancer also accounts for the most cancer-related deaths worldwide[1], with a 5-year survival rate of only 10–20% for patients diagnosed between 2010 and 2014 in most countries[2]. Non-small cell lung cancer (NSCLC), one of the two main histopathological types (the other being small cell lung cancer), accounts for 85% of lung cancer patients[3, 4]. For patients with advanced NSCLC without actionable mutations in driver genes such as *EGFR* or *ALK*, immunotherapy may prove beneficial and has become part of standard treatment. Patients with stage IV NSCLC may qualify for first-line monotherapy with pembrolizumab, a programmed death receptor-1 (PD-1) inhibitor, when at least 50% of their tumor cells show expression of programmed death-ligand 1 (PD-L1)[5, 6]. Similarly, patients with unresectable stage III NSCLC may receive consolidation treatment with durvalumab, a PD-L1 inhibitor, after chemoradiotherapy, for which the European Society for Medical Oncology (ESMO) recommended that patients should have a PD-L1 expression of at least 1% on tumor cells[7].

PD-L1 expression is determined by pathologists through immunohistochemistry, which has been validated on histological specimens. In clinical practice, however, quite often, minimally invasive techniques such as fine needle aspirations (FNA) are used to collect diagnostic material[8]. In those instances, pathologists may be asked to perform PD-L1 immunostaining on cytology samples, for which usually cell blocks (CBs) are prepared. While histological specimens are generally processed into formalin-fixed paraffin-embedded (FFPE) tissue blocks, numerous ways of processing cytology specimens into cell blocks exist[9]. What is more, various non-formaldehyde-based fixatives are used for the fixation of cytological specimens[9, 10]. We and others have shown that the use of alcohol-based fixatives may negatively influence the immunoreactivity of various antibodies[11-14], including PD-L1[15-17]. We have also demonstrated that a considerable amount of variation in PD-L1 positivity rates exists between pathology laboratories in a real-world setting, both in PD-L1 positivity rates based on histological material and in PD-L1 positivity rates based on cytological material[18]. Besides interobserver variability between pathologists, pre-analytical factors such as use of different fixatives and CB methods might play a role in causing this variation, too. In this study, the variation in fixation and CB processing of cytology samples between pathology laboratories in the Netherlands was assessed. Subsequently, we investigated whether variation in fixation and processing methods influences interlaboratory variation in PD-L1 positivity, using real-world clinical pathology data of a large cohort of NSCLC patients.

MATERIALS AND METHODS

Survey on fixation and CB methods

In order to gather information on methods used for fixation and CB processing of cytology samples in each individual laboratory, a questionnaire was sent to all pathology laboratories in the Netherlands. The questionnaire contained questions on how many different processing methods were used within the individual laboratory, which collection media and fixatives were used for various cytology samples, if post-fixation in a different fixative was part of the process, which fixation times were used, and which method was used for processing the cytology sample into a CB (Supplementary information 1). The respondents were asked to specifically answer these questions for cytological samples from NSCLC patients that may had been tested for PD-L1 between 1 July 2017 and 31 December 2018, corresponding with the study period used for extraction of data from the PALGA data set (see "Data source and data extraction").

Data source and data extraction

In order to analyze the impact of fixation and CB methods on interlaboratory variation in PD-L1 positivity, data were extracted from PALGA, the nationwide network and registry of histo- and cytopathology in the Netherlands. The PALGA registry contains all pathology records from all Dutch pathology laboratories since 1991[19]. All pathology laboratories have given consent for the storage of their data by PALGA, and for the scientific use of these data. Patients can opt out of consenting to the use of their data for research purposes. Since this specific study had a national, non-interventional retrospective design and all data were analyzed anonymously, patient consent was waived. This study was approved by PALGA's Scientific Council and Privacy Committee, and all data were handled according to the General Data Protection Regulation (GDPR).

Data on all NSCLC patients in the Netherlands with a mention of PD-L1 testing in their pathology report between 1 July 2017 and 31 December 2018 were retrieved. We have reported on this previously in another manuscript[18], in which we assessed interlaboratory variation in PD-L1 positivity in both histological and cytological material on a nationwide level. Patients with multiple primary lung tumors were excluded from the data set, because treatment of one of these tumors could potentially have influenced PD-L1 expression in the other tumor[20, 21]. For the current study, only the data on cytological specimens were used. Since information on fixation and CB method is not part of standard pathology reporting, the information from the survey was used to enrich the data from PALGA. In order to do so, the various methods described by the respondents were divided into categories within two variables, i.e. fixative and CB method. These variables were then added to the PALGA data set. In order to enable the linking of the information gathered from the survey to the data retrieved from PALGA, the survey was sent to laboratories through PALGA, ensuring

that the laboratories taking part remained anonymous to the researchers.

From all pathology reports concerning cytological specimens with known fixation and CB method, the following data were extracted: age and sex of the patient, histologic subtype of the tumor, number of PD-L1 tests performed, source of material for PD-L1 test(s), reported tumor proportion score (TPS), and number of tumor cells present (<100 or ≥100), if reported.

Analysis of interlaboratory variation in PD-L1 positivity in relation to fixation and CB methods

Equal to what we described in our previous paper[18], variation in PD-L1 positivity was studied by comparing the proportions of reported PD-L1 positive patients between the laboratories that performed PD-L1 testing. PD-L1 positivity was determined according to two clinically relevant cutoffs, i.e., ≥1% and ≥50%. Analysis of interlaboratory variation was performed for each of these cutoffs separately. Only cytology samples with known fixation and CB method were included. Furthermore, only patients from laboratories that had performed PD-L1 tests in ≥30 patients during the study period were included, and for each patient, only one PD-L1 test performed on a cytological sample was included. Patients with discordant results of multiple PD-L1 tests performed on cytological material were excluded, since it was impossible to determine which of the test results could be considered as the "true" result. This concerned results from multiple tests performed on the same tumor focus as well as results from tests on different foci of the same tumor process (e.g., primary tumor and metastasis). Patients with inconclusive test results only and patients with tests without a reported TPS were excluded as well.

Plots displaying interlaboratory variation in PD-L1 positivity in cytological material were created. Information on fixative was incorporated in these plots by using colors to display the fixative that was used most in each laboratory. Some laboratories performed PD-L1 testing for both their own and external laboratories, and fixation and CB methods could also differ within one laboratory, which is why some laboratories performed PD-L1 testing on cytological material fixed in various fixatives. Whenever laboratories used two fixatives in a fairly even distribution (up to 65–35%), two colors were used.

Statistical analysis

Patient and sample characteristics were summarized using counts and proportions. Differences between PD-L1 positive and negative subgroups were tested by using a Pearson's chi-square (χ^2) test for categorical variables and a t-test for continuous variables. Potential associations between PD-L1 positivity and fixative or CB method were assessed using univariable logistic regression analysis. PD-L1 positivity was determined using the 1% cutoff and the 50% cutoff, separately.

The mean PD-L1 positivity rate of all patients included was determined for both the 1% cutoff and the 50% cutoff. Differences in PD-L1 positivity rates between laboratories were displayed in funnel plots, which showed the mean PD-L1 positivity rate with its 95% confidence limits and the percentage of PD-L1 positive patients plotted against the total number of patients tested for each laboratory. All laboratories falling outside the 95% confidence limits were considered to differ significantly from the mean.

The PD-L1 positivity rates used in the funnel plots were adjusted for differences in patient and sample characteristics (i.e., case mix) by performing multivariable logistic regression analysis. Based on the multivariable regression model, case mix-adjusted positivity rates were determined by dividing the observed percentage of PD-L1 positive patients per laboratory by the expected percentage, followed by multiplying with the mean percentage of PD-L1 positivity. The predetermined variables that were included in the adjustment analyses were age, sex, histologic subtype, and source of material used for PD-L1 testing. Additional case mix-adjusted PD-L1 positivity rates were calculated using multivariable logistic regression analysis that also included the variables fixative and CB method. These positivity rates were then compared with the case mix-adjusted positivity rates of the same laboratories without the variables fixative and CB method. Furthermore, the likelihood ratio test (LRT) was used to compare the goodness of fit of both multivariable logistic regression models (with and without the variables fixative and CB method).

All statistical analyses were performed using IBM SPSS Statistics version 26.

RESULTS

Nationwide variation in fixation and CB methods of cytology samples

We received responses from 28 (66.7%) of the 42 laboratories to that the questionnaire was sent to. From the responses, 19 different ways of processing cytology samples could be discerned. Figure 1 provides an overview of the various combinations of collection medium, fixative, post-fixation, and CB method, with the number of times each combination was used by a laboratory in the final column. Sometimes different processing methods were used within one laboratory, depending on the type of cytological material. This explains why the numbers in the final column add up to 37 instead of 28. Variation in mean fixation time ranged from 20 min to 36 h, with a mean of 12 h.

Formalin or Unifix, a substance also containing formaldehyde, was used most often in the various combinations of fixatives and CB methods employed by the different laboratories (12 out of 37 times, 32.4%). Alcohol fixation (methanol- or ethanol-based) followed by formalin fixation was used 11 of 37 times (29.7%). Seven (18.9%) of all 37 processing

methods contained a step involving CytoRich Red fixation, a solution that contains both alcohols and formaldehyde. Finally, in 7 out of 37 cases (18.9%), the cytological material was fixed using CytoLyt and PreservCyt (Hologic, Marlborough, Massachusetts, USA), both methanol-based, without formalin post-fixation.

Four different CB methods were used, i.e., centrifugation of the cytology sample and processing the cell pellet into an FFPE CB, agar embedding, the Thermo Scientific or Shandon CB method (Thermo Fischer Scientific, Waltham, Massachusetts, USA), and the Cellient-automated CB system (Hologic, Marlborough, Massachusetts, USA). Centrifugation and processing into an FFPE CB and the agar-based method were used most often (12 out of 37 times, 32.4%). The Cellient CB system and Thermo Scientific CB method were used in 8 (21.6%) and 5 (13.5%) out of 37 cases, respectively.

Patient selection process from the PALGA data set

Information on fixation methods used by the various laboratories was divided into four categories (formalin fixation, CytoRich Red fixation, alcohol (methanol or ethanol) fixation with formalin post-fixation, and CytoLyt/PreservCyt fixation (without formalin post-fixation)) and added to the PALGA data set. Similarly, information on CB methods used by the various laboratories was also divided into four categories (centrifugation and FFPE CB, agar-based CB, Thermo Scientific CB, and Cellient CB) and added to the PALGA data set.

The PALGA data set showed that during the study period, 10,625 PD-L1 tests were performed in 8,725 patients with NSCLC in the Netherlands. Data from 42 laboratories were included, of which 32 performed PD-L1 testing on cytology samples. Of all tests, 2,665 (25.1%) were performed on cytological material of 2,300 patients. Based on the results from our survey, information on fixation and CB method could be added to the samples of 1,784 patients, resulting in the exclusion of 516 patients. After this, 92 patients from laboratories that performed PD-L1 testing in < 30 patients were excluded. Finally, after the exclusion of patients with discordant results from multiple PD-L1 tests (n = 14), patients with inconclusive test results only (n = 216), and patients with tests with unknown TPS (n = 4), 1,458 patients from 19 laboratories remained for analysis of interlaboratory variation in PD-L1 positivity (Figure 2). The patients with inconclusive test results only had a total of 259 tests performed, of which 195 (75.3%) were reported to have an insufficient amount of viable tumor cells (<100). Characteristics of all included patients and their samples tested for PD-L1 are displayed in Table 1. Proportions significantly differed between PD-L1 positivity and negativity across sex, histologic subtype, fixative, and CB method for both cutoffs.

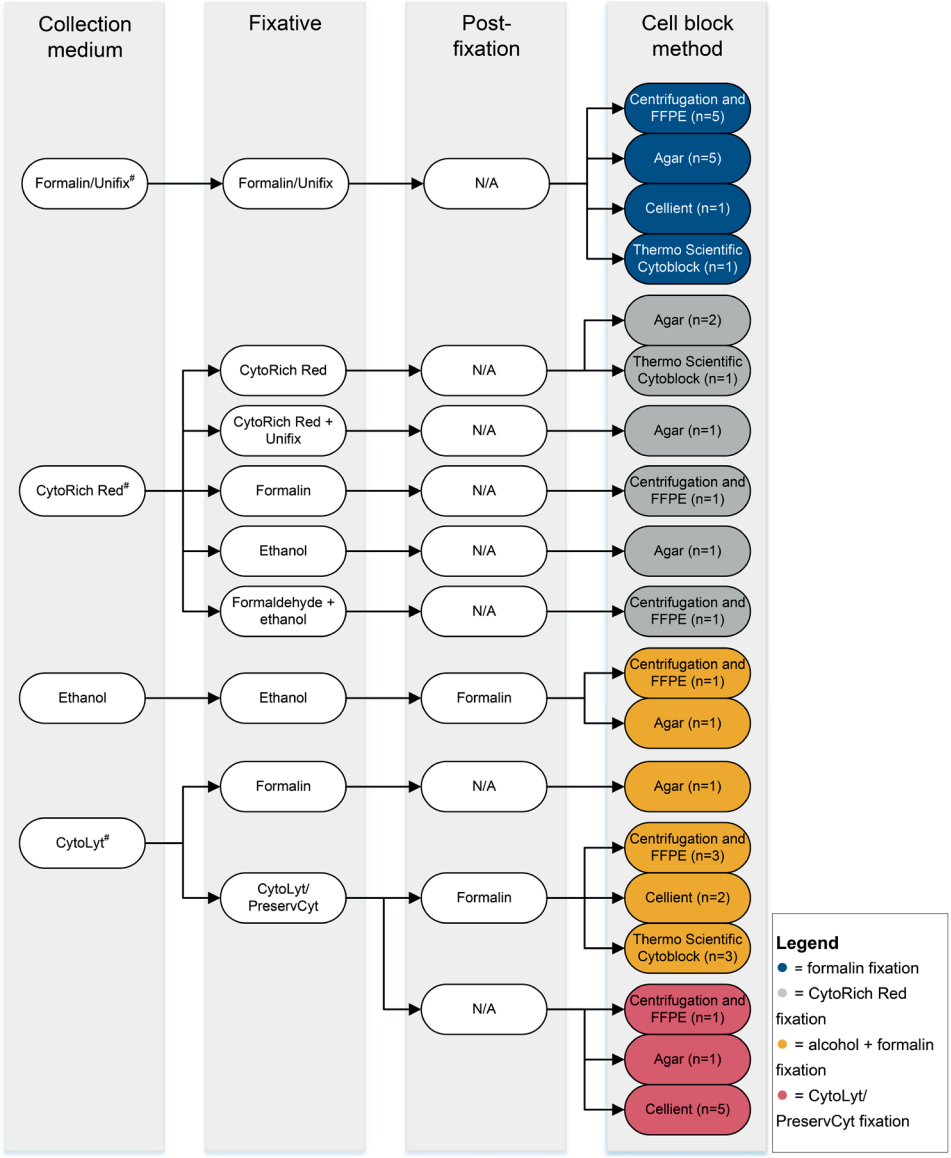


Figure 1. Overview of the various combinations of fixation and cell block methods for cytology samples as described by the survey respondents. The final column displays the number of times each combination is used. Colors depict the overall fixation method for each combination (see legend). #Fluids were not always received in a collection medium, but rather as fresh fluids. Abbreviations: FFPE, formalin-fixed paraffin-embedded; N/A, not applicable.

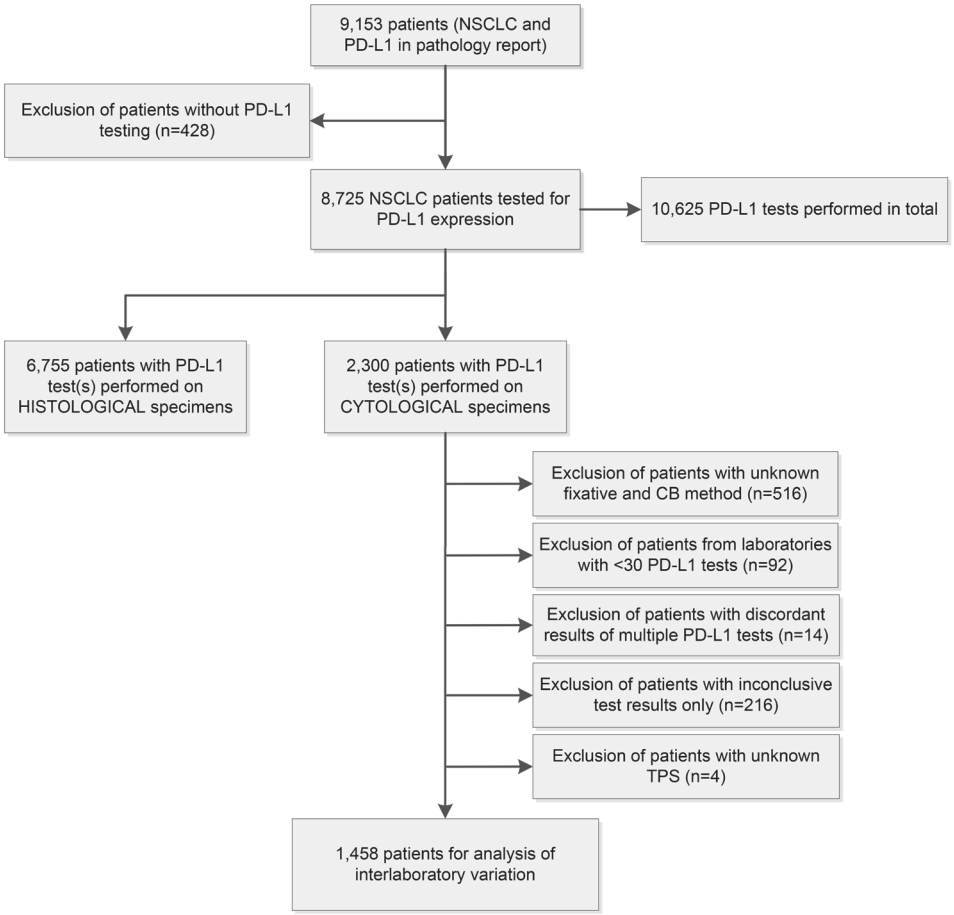


Figure 2. Flowchart of the patient selection process from the PALGA data set. CB, cell block; NSCLC, non-small cell lung cancer; PD-L1, programmed death-ligand 1; TPS, tumor proportion score.

Table 1. Differences in patient and sample characteristics between PD-L1 negativity and PD-L1 positivity (PD-L1 <1% vs. PD-L1 ≥1% and PD-L1 <50% vs. PD-L1 ≥50%).

	Total (n = 1458)	**PD-L1 <1% (n= 705)**	**PD-L1 ≥1% (n = 753)**	***p*-value**	**PD-L1 <50% (n= 1042)**	**PD-L1 ≥50% (n = 416)**	***p*-value**
Age in years (mean (SD))	66.5 (10.1)	66.6 (10.3)	66.5 (9.8)	0.82	66.6 (10.2)	66.3 (9.7)	0.66
Sex							
Male	777 (53.3%)	408 (57.9%)	369 (49.0%)	<0.01	583 (56.0%)	194 (46.6%)	<0.01
Female	681 (46.7%)	297 (42.1%)	384 (51.0%)		459 (44.0%)	222 (53.4%)	
Histologic subtype							
AC	1105 (75.8%)	500 (70.9%)	605 (80.3%)	<0.01	758 (72.7%)	347 (83.4%)	<0.01
SCC	206 (14.1%)	122 (17.3%)	84 (11.2%)		167 (16.0%)	39 (9.4%)	
NSCLC NOS	132 (9.1%)	73 (10.4%)	59 (7.8%)		105 (10.1%)	27 (6.5%)	
Other	15 (1.0%)	10 (1.4%)	5 (0.7%)		12 (1.2%)	3 (0.7%)	
Source of material							
Primary tumor	81 (5.6%)	41 (5.8%)	40 (5.3%)	0.54	60 (5.8%)	21 (5.0%)	0.64
Metastasis	103 (7.1%)	54 (7.7%)	49 (6.5%)		72 (6.9%)	31 (7.5%)	
Lymph node metastasis	873 (59.9%)	430 (61.0%)	443 (58.8%)		618 (59.3%)	255 (61.3%)	
Pleural effusion	342 (23.5%)	154 (21.8%)	188 (25.0%)		245 (23.5%)	97 (23.3%)	
Bronchial brush/fluid	59 (4.0%)	26 (3.7%)	33 (4.4%)		47 (4.5%)	12 (2.9%)	
Fixative							
Formalin	459 (31.5%)	185 (26.2%)	274 (36.4%)	<0.01	306 (29.4%)	153 (36.8%)	<0.05
CytoRich Red	301 (20.6%)	140 (19.9%)	161 (21.4%)		213 (20.4%)	88 (21.2%)	
Alcohol + formalin	346 (23.7%)	175 (24.8%)	171 (22.7%)		255 (24.5%)	91 (21.9%)	
CytoLyt/PreservCyt	352 (24.1%)	205 (29.1%)	147 (19.5%)		268 (25.7%)	84 (20.2%)	
Cell block method							
Centrifugation and FFPE	301 (20.6%)	143 (20.3%)	158 (21.0%)	<0.01	213 (20.4%)	88 (21.2%)	<0.01
Agar	627 (43.0%)	255 (36.2%)	372 (49.4%)		417 (40.0%)	210 (50.5%)	
Thermo Scientific Cytoblock	191 (13.1%)	91 (12.9%)	100 (13.3%)		138 (13.2%)	53 (12.7%)	
Cellient	339 (23.3%)	216 (30.6%)	123 (16.3%)		274 (26.3%)	65 (15.6%)	

Abbreviations: AC = adenocarcinoma; FFPE = formalin-fixed paraffin-embedded; NOS = not otherwise specified; NSCLC = non-small cell lung cancer; PD-L1 = programmed death-ligand 1; SD = standard deviation; SCC = squamous cell carcinoma.

Analysis of the association between fixative and PD-L1 positivity

Within the different categories of the variable fixative, the mean PD-L1 positivity rate based on the 1% cutoff was highest in the samples fixed in formalin (59.7%). After this came CytoRich Red (53.5%) and alcohol fixation with formalin post-fixation (49.4%), followed by CytoLyt/PreservCyt fixation without formalin post-fixation (41.8%). At the 50% cutoff, the mean percentage of PD-L1 positive cases was 33.3% for all samples fixed in formalin. Mean PD-L1 positivity was 29.2% for samples fixed in CytoRich Red, 26.3% for samples fixed in alcohol with formalin post-fixation, and 23.9% for samples fixed in CytoLyt/PreservCyt (Supplementary Figure 1a).

Univariable logistic regression analysis revealed a statistically significant association between fixative and PD-L1 positivity for both the 1% cutoff and the 50% cutoff (Table 2), with the odds of scoring PD-L1 as positive being significantly lower in samples fixed in CytoLyt/PreservCyt (without formalin post-fixation) or in alcohol with formalin post-fixation compared with samples fixed in formalin only. There was no significant difference in the odds of scoring PD-L1 as positive between samples fixed in CytoRich Red and those fixed in formalin.

Table 2. Univariable logistic regression analysis for assessment of association between the variables fixative and cell block method and PD-L1 positivity (defined as ≥1% and ≥50% positive tumor cells).

	PD-L1 ≥1% vs. <1%		PD-L1 ≥50% vs. <50%	
	OR (95% CI)	**overall *p*-value**	**OR (95% CI)**	**overall *p*-value**
Fixative		<0.01		<0.05
Formalin	1.00 *reference*		1.00 *reference*	
CytoRich Red	0.78 (0.58-1.04)		0.83 (0.60-1.13)	
Alcohol + formalin	**0.66 (0.50-0.87)**		**0.71 (0.52-0.97)**	
CytoLyt/PreservCyt	**0.48 (0.37-0.64)**		**0.63 (0.46-0.86)**	
Cell block method		<0.01		<0.01
Centrifugation and FFPE	1.00 *reference*		1.00 *reference*	
Agar	**1.32 (1.00-1.74)**		1.22 (0.90-1.64)	
Thermo Scientific Cytoblock	1.00 (0.69-1.43)		0.93 (0.62-1.39)	
Cellient	**0.52 (0.38-0.71)**		**0.57 (0.40-0.83)**	

Data in bold indicate a statistically significant difference in the odds of scoring PD-L1 as positive compared to the reference category. Abbreviations: CI = confidence interval; FFPE = formalin-fixed paraffin-embedded; OR = odds ratio; PD-L1 = programmed death-ligand 1.

Analysis of the association between CB method and PD-L1 positivity

The mean PD-L1 positivity rate at the 1% cutoff for samples that were centrifuged and processed into FFPE CBs was 52.5%. For the other categories, mean PD-L1 positivity rates were 59.3% (agar CBs), 52.4% (Thermo Scientific CBs), and 36.3% (Cellient CBs). At the 50% cutoff, the mean PD-L1 positivity was 29.2% for samples centrifuged and processed into FFPE CBs. Mean PD-L1 positivity was 33.5% for samples processed into agar CBs, 27.7% for samples processed into Thermo Scientific CBs, and 19.2% for samples processed

into Cellient CBs (Supplementary Figure 1b).

A statistically significant association was found between CB method and PD-L1 positivity for both cutoffs (Table 2). The odds of scoring PD-L1 as positive were significantly lower in samples processed into Cellient CBs in comparison to samples that were centrifuged and processed into FFPE CBs. At the 1% cutoff, the odds of PD-L1 positivity were significantly higher for samples processed into agar-based CBs than for those centrifuged and processed into FFPE CBs. No statistically significant differences in the odds of PD-L1 positivity were found between the Thermo Scientific Cytoblock method and centrifugation and processing into an FFPE CB.

Fixation and CB methods in relation to interlaboratory variation of PD-L1 positivity

The mean PD-L1 positivity rate of all included patients was 51.6% at the 1% cutoff and 28.5% at the 50% cutoff. When positivity rates without any case mix adjustment were plotted against the total number of PD-L1 tests for each laboratory and compared to the overall mean, eight (42.1%) laboratories differed significantly from the mean at the 1% cutoff and nine (47.4%) laboratories differed significantly from the mean at the 50% cutoff (data not shown). After case mix adjustment for sex, age, histologic subtype, and source of material, funnel plots showed eight (42.1%) laboratories differing significantly from the overall mean at both cutoffs (Figure 3). Case mix adjustment with these variables thus resulted in a reduction of the number of laboratories differing significantly from the mean from nine to eight at the 50% cutoff. No reduction in the number of laboratories differing significantly from the mean was seen at the 1% cutoff. The case mix-adjusted positivity rates from the individual laboratories ranged from 26.0 to 72.4% at the 1% cutoff, and from 9.9 to 40.9% at the 50% cutoff.

When looking at Figure 3a, attention is drawn to the seven laboratories that used formalin fixation for at least part of their samples (in blue). Five of these laboratories lie above the overall mean, with four laboratories falling outside the upper 95% confidence limit. In contrast, most of the laboratories that primarily used CytoLyt and PreservCyt fixation without formalin post-fixation (in red) lie below the overall mean, with one of the laboratories that fall far below the lower 95% confidence limit also using CytoLyt/PreservCyt fixation. The other laboratory falling below the lower 95% confidence limit mainly used combined alcohol and formalin fixation, but this fixation method was also used by two of the laboratories that lie above the upper 95% confidence limit. Hence, the mean PD-L1 positivity rate of all samples fixed in alcohol followed by formalin post-fixation lies much closer to the overall mean, while the mean of all samples fixed in CytoLyt/PreservCyt without formalin post-fixation lies further below the overall mean. Conversely, the mean of all samples fixed in formalin lies quite far above the overall mean. The differences between the laboratories that used different fixation methods are less apparent at the 50% cutoff (Figure 3b), although a similar pattern can still be discerned.

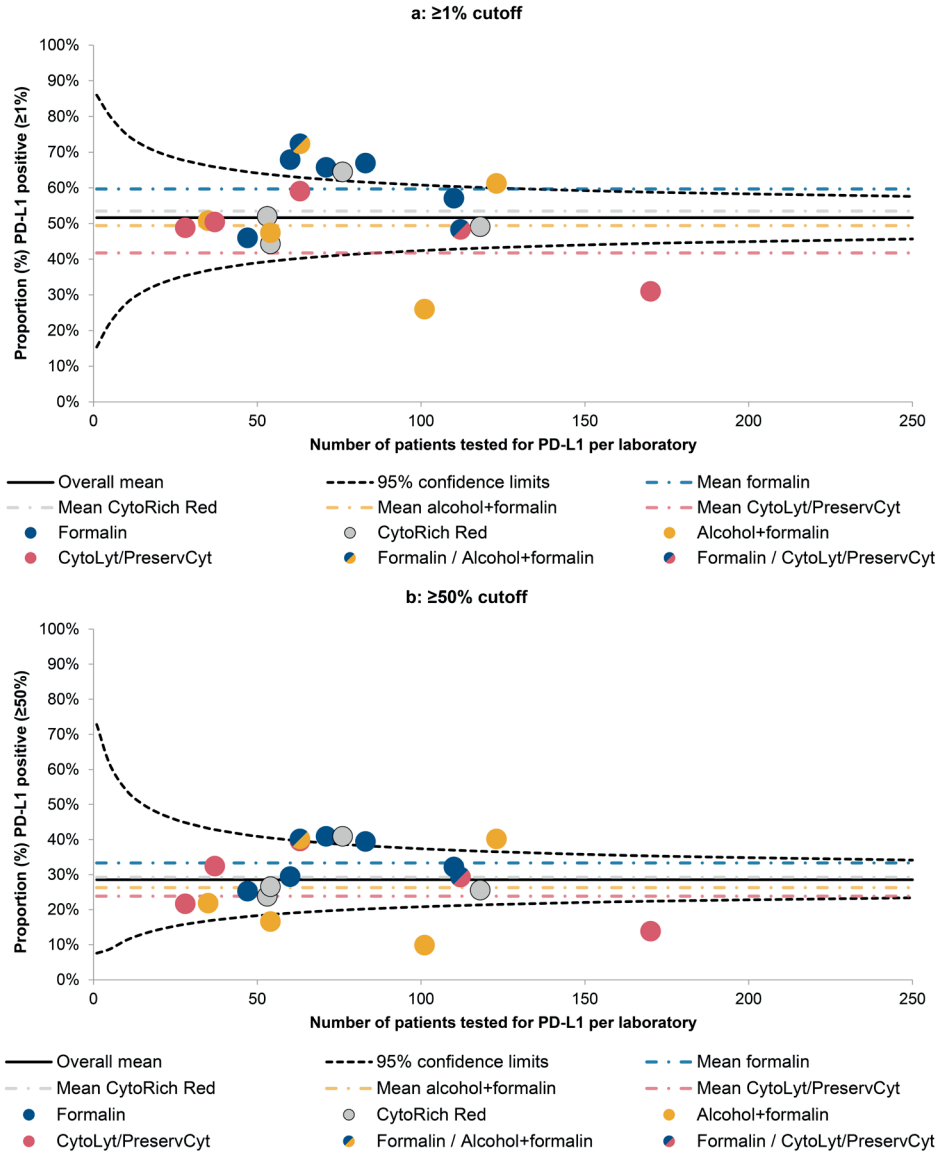


Figure 3. Funnel plots showing interlaboratory variation in programmed death-ligand 1 (PD-L1) positivity. PD-L1 positivity was determined using either a 1% cutoff (**a**) or a 50% cutoff (**b**). For each laboratory, case mix-adjusted positivity rates are displayed against the total number of patients tested for PD-L1 (dots). The variables age, sex, histological subtype, and source of material used for PD-L1 testing were included in the case mix adjustment analysis. Colors are used to indicate the fixative that was used the most in each laboratory (see legend). The black line shows the overall mean proportion of PD-L1 positive patients, surrounded by its 95% confidence limits (black dotted lines). The colored dotted lines display the mean PD-L1 positivity rate for each fixative category.

When the variables fixative and CB method were included in the multivariable logistic regression models for case mix correction, the number of laboratories falling outside the 95% confidence limits decreased from eight (42.1%) to five (26.3%) at the 1% cutoff (Figure 4a). Adding the variables fixative and CB method to the case mix correction analysis resulted in a maximum variation between PD-L1 positivity rates of 47.9% (34.4–82.3%). This was a slightly wider range than the range of PD-L1 positivity rates adjusted for case mix without fixative and CB method, which was 46.4%. This seems mainly driven by a single laboratory that went from falling within the 95% confidence limits when adjusted for case mix without the variables fixative and CB method, to falling far outside the upper 95% confidence limit when adjusted for case mix including fixative and CB method.

At the 50% cutoff, adjusting PD-L1 positivity rates for case mix with the inclusion of the variables fixative and CB method again resulted in a decrease of the number of laboratories differing significantly from the mean from eight (42.1%) to five (26.3%) (Figure 4b). The PD-L1 positivity rates of the individual laboratories ranged from 14.9 to 55.9%, resulting in a maximum variation of 41.0%. This maximum variation was wider than the maximum variation of PD-L1 positivity rates adjusted for the case mix without the fixative and CB method (41.0% vs. 31.0%). Again, this seems mainly driven by the same outlier laboratory as at the 1% cutoff.

At both cutoffs, adding the variables fixative and CB method to the multivariable logistic regression model resulted in a lower log-likelihood value in comparison to the model without the variables fixative and CB method, indicating a better fit to the data of the extended model. The difference in fit between the models was statistically significant based on the LRT (p-value < 0.01).

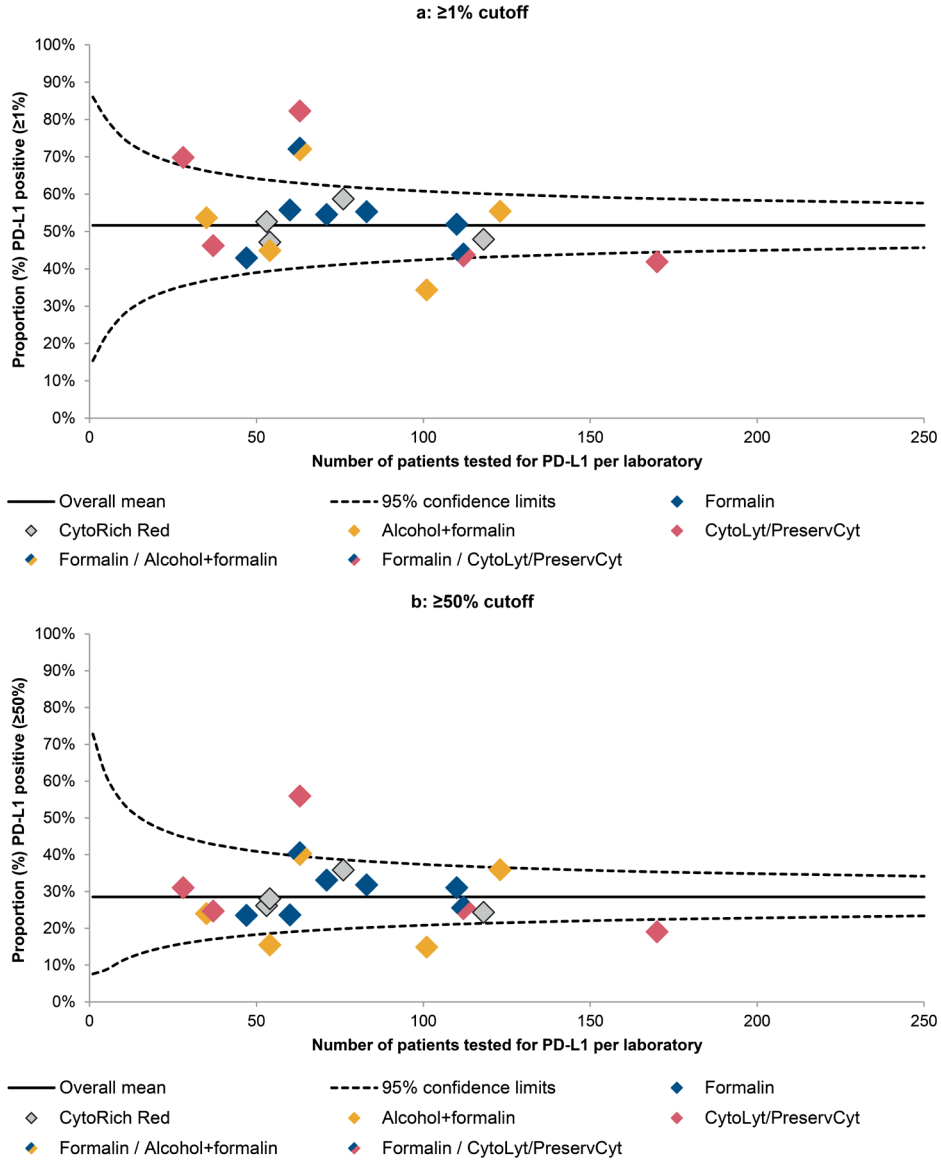


Figure 4. Funnel plots showing interlaboratory variation in programmed death-ligand 1 (PD-L1) positivity, with PD-L1 positivity rates adjusted for case mix including the variables fixative and cell block method. PD-L1 positivity was determined using either a 1% cutoff (**a**) or a 50% cutoff (**b**). For each laboratory, case mix-adjusted positivity rates are displayed against the total number of patients tested for PD-L1 (diamonds). The color of the diamonds indicates the fixative that was used the most in each laboratory (see legend). The black line shows the overall mean proportion of PD-L1 positive patients, surrounded by its 95% confidence limits (dotted lines).

DISCUSSION

In this cohort study based on real-world data, the variation in fixation and CB processing of cytology samples by pathology laboratories in the Netherlands was assessed. We revealed that many differences exist in both the use of fixatives and of CB methods, sometimes including multiple methods within one laboratory. Correcting PD-L1 positivity rates of individual laboratories for differences in the use of fixative and CB method resulted in a reduction of the number of laboratories that differed significantly from the mean PD-L1 positivity. Moreover, the observed decrease in interlaboratory variation was considerably greater than the decrease that was seen when PD-L1 positivity rates were corrected for differences in patient and sample characteristics without the variables fixative and CB method.

First of all, the number of different fixation and processing methods and reported combinations among the laboratories that responded to our survey is enormous: within a total of 28 laboratories, 19 different combinations of fixation and processing cytological material into a CB could be discerned. These results are comparable to those of other studies that used surveys to assess interlaboratory variation in both fixation and CB methods, which also showed large amounts of variation in cytology processing methods between laboratories[8, 22-24].

Both methanol-based and ethanol-based fixatives have a potentially deleterious effect on PD-L1 immunostaining performed on CBs[15-17], with a risk of false-negative PD-L1 immunostaining results. Indeed, correcting for differences in cytology fixation and CB processing methods between laboratories resulted in a reduction in the number of laboratories differing significantly from the mean in PD-L1 positivity. Formalin post-fixation may reverse the negative effects of alcohol fixation to some degree[15], with some studies showing good concordance in PD-L1 positivity between histology and cytological specimens from the same tumor fixed in an alcohol-based fixative followed by formalin fixation[25-27]. It is unclear, however, what the maximum duration of alcohol fixation is after which formalin post-fixation is still effective, and what the most optimal formalin post-fixation time would be. CytoRich Red, containing both alcohols and formaldehyde, did not seem to have a negative effect on PD-L1 immunostaining in various studies[15, 26, 28]. Likewise, univariable logistic regression analyses in our study did not show a statistically significant difference in the odds of finding PD-L1 positivity between CytoRich Red fixation and formalin fixation, while the odds of scoring PD-L1 as positive were significantly lower in samples fixed in CytoLyt/PreservCyt without formalin post-fixation or in alcohol with formalin post-fixation compared with samples fixed in formalin.

It is possible that differences in the CB method also influence interlaboratory variation in PD-L1 positivity, regardless of the fixation method. Yet, very few studies have been published

that assessed the influence of the CB method on immunostaining independently of the fixation method. In a study by Lloyd et al.[17], cytology samples processed into CBs with the Cellient-automated CB system showed optimal PD-L1 staining results compared with CB preparation according to the plasma-thromboplastin method. However, the authors advise against the use of CytoLyt as a collection medium due to the poor performance of PD-L1 immunostaining in samples collected in CytoLyt. Remarkably, CytoLyt is the collection medium of choice recommended for use with the Cellient system by the manufacturer. In our study, nearly 75% of the Cellient processed samples were fixed in CytoLyt/PreservCyt, and the remainder were fixed in an alcohol-based fixative with formalin post-fixation. None of the Cellient processed samples were fixed in formalin only. All in all, based on the available literature, it is very likely that the influence of differences in cytology processing methods on interlaboratory variation in PD-L1 positivity can be attributed mostly to differences in fixation methods.

We have reported previously that a large degree of variation in PD-L1 positivity between laboratories is problematic. Indeed, this could result in patients receiving different PD-L1 test results depending on the pathology laboratory where their material is tested[18]. Thus, efforts should be taken to keep interlaboratory variation in PD-L1 positivity to a minimum. Based on the current study, an important step to take would be to create more uniformity between laboratories in the way that cytology samples are fixed and processed, using a method that does not negatively influence immunostaining results. This desire for uniformity has been expressed by others[8, 29], too, and could prove beneficial not only to results from PD-L1 immunostaining but also to results from other immunohistochemical assays that show adverse effects of alcohol fixation, such as for progesterone receptor[30] and MIB1[31]. External quality assessment (EQA) schemes specifically designed to assess immunocytochemistry could perhaps aid in uncovering possible technical issues and in promoting standardization[32]. Future studies should investigate which method is the preferred (combination of) cytology processing method(s) for PD-L1 testing.

Potentially, rigorous validation and optimization of immunostaining protocols that are used on cytology samples but have originally been validated on FFPE tissue samples could aid in diminishing variation as well. Unfortunately, it has been shown that validation and optimization of immunostaining protocols for cytology samples are not common practice[23, 24], even though organizations such as the College of American Pathologists (CAP) recommend that a sufficient number of cases should be tested to ensure that immunohistochemical assays achieve similar results when performed on cytological material compared to histological material[33]. No advice is given, however, on the criteria and number of specimens needed for validation, and it is stated that “separate validation of all markers on all potential cytologic specimens is generally not feasible”[33]. The type of material that should be used for validation may often not be clear either, or it may be

difficult to collect enough material, especially when dealing with small cytology samples. On top of that, some laboratories receive cytological specimens from external laboratories for immunocytochemical testing, which may have been fixed and processed in a variety of ways. Moreover, with PD-L1 immunostaining, a decrease in staining intensity could result in false-negative staining results, but it is hard to determine what level of decrease in staining can still be accepted and what level would actually cause problems in clinical practice. All these factors may complicate the proper validation and optimization of PD-L1 immunohistochemical stains that are used on cytological specimens. On top of that, laboratories may use commercial assays for PD-L1 immunostaining, which use standardized protocols developed by the manufacturer that cannot simply be adjusted.

After correction for case mix including variation in the fixative and CB method, the amount of interlaboratory variation in PD-L1 positivity was still substantial at both cutoffs. Compared to histology, tissue architecture is disrupted in cytology samples, which can complicate the recognition of tumor cells. Also, it may be a lot harder to distinguish tumor cells from inflammatory cells, especially macrophages, which may lie adjacent to or intermixed with isolated tumor cells[34, 35]. The level of experience and training of the pathologists scoring PD-L1 immunostaining on cytology in a routine clinical pathology setting may not be the same in all laboratories, all the more so because scoring of PD-L1 on cytology requires adequate training in both cytopathology and PD-L1 scoring. Structural differences between laboratories could arise, for instance, when inflammatory cells are often mistaken for tumor cells. Moreover, small tissue samples can lead to an underestimation of PD-L1 expression[36], which probably also applies to cytology samples. In fact, it has been shown that PD-L1 staining results of CBs and resection specimens are more concordant when a greater number of tumor cells were present in the CB[37]. Perhaps laboratories that structurally receive cytological samples that contain more tumor cells, for instance, because multiple passes or bigger needles are used to collect material, have higher PD-L1 positivity rates based on cytology than other laboratories. Our study, however, does not provide the data to properly investigate this hypothesis.

Of note, in our study, an association was found between sex and PD-L1 expression, with PD-L1 positivity being more likely in samples from women than from men (Table 1). While similar results have been shown by others[38, 39], various other studies did not find any association between PD-L1 expression and sex[40-45] or found that PD-L1 was more likely to be positive in men than in women[46-48]. These studies, however, primarily used FFPE material, mostly from surgical resections or biopsies. Our study only included cytological samples, many of which were not fixed in formalin or embedded in paraffin, which might explain the differences in results. Similarly, while some studies did not find a statistically significant association between PD-L1 expression and sampling site[49, 50], comparable to our results (Table 1), others showed that pleural and nodal metastases were more likely

to express PD-L1 than primary tumors[42]. Again, the difference in results could potentially be explained by the latter study using FFPE material which largely came from biopsies or surgical resections, while our study only used cytological material fixed and processed in a variety of ways. Moreover, the differences in proportions between PD-L1 positivity and negativity of the various characteristics in Table 1 were tested through univariate analysis, which does not account for potential confounding factors. Also, since we used a large cohort in our study, small and maybe even clinically insignificant differences might be statistically significant, whereas they might not have been in studies with smaller sample sizes. These factors should be taken into account when interpreting these results.

This study has some limitations. Most importantly, even though a considerable number of laboratories responded to our survey, we did not receive answers from all laboratories. This resulted in the exclusion of 516 patients from the PALGA data set, for whom the fixative and CB method were unknown. Given the current variation in the fixation and CB methods, it is to be expected that the overall number of methods used would only be larger, potentially resulting in a larger baseline variation among laboratories to start with. This should be considered when interpreting the analyses of the influence of variation in fixation and CB processing on interlaboratory variation in PD-L1 positivity. Second, the respondents reported varying mean fixation times. Vigliar et al.[51] showed that formalin fixation time influences PD-L1 immunostaining results on CBs. This could be the case with other fixatives, too. Unfortunately, we did not know the fixation times for individual samples in the PALGA data set and, thus, could not incorporate information on fixation time in our analyses. Third, due to a large amount of variation in cytology processing methods, especially in fixation methods, it was quite difficult to divide these various methods into larger categories. In fact, if numbers had allowed to include all methods as they were, a better correction could have been performed. However, we do feel that the distribution that we used is compatible with the currently available literature. Finally, since our study is based on real-world pathology data, we were dependent on the way that pathologists report their findings. For instance, while it would have been interesting to include an analysis of TPS on a continuous scale, the fact that various laboratories only reported TPS in categories did not allow us to do so. Also, some potentially relevant information, such as information on previous treatment, is not regularly part of pathology reports, meaning that we could not correct for potential differences between laboratories within these areas. Regarding treatment status, however, in many patients, PD-L1 testing was performed on the initial diagnostic material, either at the time of diagnosis or at a later time. We also excluded patients with more than one primary lung tumor, to avoid including data from PD-L1 tests that might have been influenced by previous treatment. We therefore expect the number of patients in which PD-L1 testing was performed solely on material collected after administration of chemotherapy to be too small to influence our results in a significant way.

To conclude, this study shows that a lot of variation exists between laboratories in the methods used for fixation and CB processing of cytological samples. We have demonstrated that these differences influence interlaboratory variation in PD-L1 positivity in NSCLC patients, with a decrease in the amount of variation when PD-L1 positivity rates are corrected for differences in fixation and CB methods. A high degree of variation in PD-L1 positivity between laboratories is problematic, because this will almost inevitably lead to patients receiving different courses of treatment depending on the laboratory where their cytological material is stained and scored for PD-L1. These results warrant the need for more research to determine the best methods of fixation and CB processing of cytology samples on which PD-L1 immunostaining is to be performed, and for harmonization of these methods between laboratories.

ACKNOWLEDGEMENTS

We would like to thank all laboratories that took part in this study. We also thank Chantal Epskamp-Kuijpers, Marion Visscher, Eeke Leerssen, and Jasmijn van Es for their help in extracting data from the PALGA data set.

This work was funded with research grants from AstraZeneca, MSD, and Roche (grant numbers not applicable). None of the grant suppliers were involved in the study design, collection, analysis and interpretation of data, writing of the manuscript, or the decision to submit the manuscript for publication.

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. Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 2018;391(10125):1023–1075.
3. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. Cancer Manag Res. 2019;11:943–953.
4. Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clin Proc. 2019;94(8):1623–1640. doi: 10.1016/j.mayocp.2019.01.013.
5. Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(Suppl 4):iv192–iv237.
6. Pai-Scherf L, Blumenthal GM, Li H, et al. FDA approval summary: pembrolizumab for treatment of metastatic non-small cell lung cancer: first-line therapy and beyond. Oncologist. 2017;22(11):1392–1399.
7. Pentheroudakis G, Committee EG. Recent eUpdate to the ESMO Clinical Practice Guidelines on early and locally advanced non-small-cell lung cancer (NSCLC) Ann Oncol. 2020;31(9):1265–1266.
8. Crapanzano JP, Heymann JJ, Monaco S, Nassar A, Saqi A. The state of cell block variation and satisfaction in the era of molecular diagnostics and personalized medicine. Cytojournal. 2014;11:7.
9. Nambirajan A, Jain D. Cell blocks in cytopathology: an update. Cytopathology. 2018;29(6):505–524.
10. Krogerus L, Kholova I. Cell block in cytological diagnostics: review of preparatory techniques. Acta Cytol. 2018;62(4):237–243.
11. Sauter JL, Grogg KL, Vrana JA, Law ME, Halvorson JL, Henry MR. Young investigator challenge: validation and optimization of immunohistochemistry protocols for use on cellient cell block specimens. Cancer Cytopathol. 2016;124(2):89–100.
12. Ireka Y, Agustina H, Aziz A, Hernowo BS, Suryanti S. Comparison of fixation methods for preservation cytology specimens of cell block preparation using 10% neutral buffer formalin and 96% alcohol fixation in E-cadherin and Ki-67 immunohistochemical examination. Open Access Maced J Med Sci. 2019;7(19):3139–3144.
13. Gruchy JR, Barnes PJ, Dakin Hache KA. CytoLyt(R) fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. Appl Immunohistochem Mol Morphol. 2015;23(4):297–302.
14. Kruger AM, Stevens MW, Kerley KJ, Carter CD. Comparison of the Cellient() automated cell block system and agar cell block method. Cytopathology. 2014;25(6):381–388.
15. Koomen BM, van der Starre-Gaal J, Vonk JM, et al. Formalin fixation for optimal concordance of programmed death-ligand 1 immunostaining between cytologic and histologic specimens from patients with non-small cell lung cancer. Cancer Cytopathol. 2021;129(4):304–317.
16. Koomen BM, Vreuls W, de Boer M, et al. False-negative programmed death-ligand 1 immunostaining in ethanol-fixed endobronchial ultrasound-guided transbronchial needle aspiration specimens of non-small-cell lung cancer patients. Histopathology. 2021;79(4):480–490.
17. Lloyd IE, Zhou W, Witt BL, Chadwick BE. Characterization of PD-L1 Immunohistochemical expression in cell

blocks with different specimen fixation and processing methods. Appl Immunohistochem Mol Morphol. 2019;27(2):107–113.

18. Koomen BM, Voorham QJM, Epskamp-Kuijpers C, et al. Considerable interlaboratory variation in PD-L1 positivity in a nationwide cohort of non-small cell lung cancer patients. Lung Cancer. 2021;159:117–126.
19. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol. 2007;29(1):19–24.
20. Sheng J, Fang W, Yu J, et al. Expression of programmed death ligand-1 on tumor cells varies pre and post chemotherapy in non-small cell lung cancer. Sci Rep. 2016;6:20090.
21. Rojko L, Reiniger L, Teglasi V, et al. Chemotherapy treatment is associated with altered PD-L1 expression in lung cancer patients. J Cancer Res Clin Oncol. 2018;144(7):1219–1226.
22. Moriarty AT, Nayar R, Auger M, et al. Nongynecologic cytology practice patterns: a survey of participants in the College of American Pathologists interlaboratory comparison program in nongynecologic cytopathology. Arch Pathol Lab Med. 2014;138(7):885–889.
23. Srebotnik Kirbis I, Rodrigues Roque R, Bongiovanni M, Strojan Flezar M, Cochand-Priollet B. Immunocytochemistry practices in European cytopathology laboratories-Review of European Federation of Cytology Societies (EFCS) online survey results with best practice recommendations. Cancer Cytopathol. 2020;128(10):757–766.
24. Fischer AH, Schwartz MR, Moriarty AT, et al. Immunohistochemistry practices of cytopathology laboratories: a survey of participants in the College of American Pathologists Nongynecologic Cytopathology Education Program. Arch Pathol Lab Med. 2014;138(9):1167–1172.
25. Lou SK, Ko HM, Kinoshita T, et al. Implementation of PD-L1 22C3 IHC pharmDxTM in cell block preparations of lung cancer: concordance with surgical resections and technical validation of CytoLyt(R) Prefixation. Acta Cytol. 2020;64(6):577–587.
26. Gosney JR, Haragan A, Chadwick C, et al. Programmed death ligand 1 expression in EBUS aspirates of non-small cell lung cancer: is interpretation affected by type of fixation? Cancer Cytopathol. 2020;128(2):100–106.
27. Arriola AGP, Bashover E, Joseph C, Staerkel G, Wang WL, Roy-Chowdhuri S. The usefulness of various cytologic specimen preparations for PD-L1 immunostaining in non-small cell lung carcinoma. J Am Soc Cytopathol. 2018;7(6):324–332.
28. Song SG, Lee J, Koh J, Kim S, Chung DH, Jeon YK. Utility of PD-L1 immunocytochemistry using body-fluid cell blocks in patients with non-small-cell lung cancer. Diagn Cytopathol. 2020;48(4):291–299.
29. Vigliar E, Malapelle U, Iaccarino A, et al. PD-L1 expression on routine samples of non-small cell lung cancer: results and critical issues from a 1-year experience of a centralised laboratory. J Clin Pathol. 2019;72(6):412–417.
30. Gorman BK, Kosarac O, Chakraborty S, Schwartz MR, Mody DR. Comparison of breast carcinoma prognostic/predictive biomarkers on cell blocks obtained by various methods: cellient, formalin and thrombin. Acta Cytol. 2012;56(3):289–296.
31. Buonocore DJ, Konno F, Jungbluth AA, et al. CytoLyt fixation significantly inhibits MIB1 immunoreactivity whereas alternative Ki-67 clone 30–9 is not susceptible to the inhibition: critical diagnostic implications. Cancer Cytopathol. 2019;127(10):643–649.

32. Kirbis IS, Maxwell P, Flezar MS, Miller K, Ibrahim M. External quality control for immunocytochemistry on cytology samples: a review of UK NEQAS ICC (cytology module) results. Cytopathology. 2011;22(4):230–237.
33. Fitzgibbons PL, Bradley LA, Fatheree LA, et al. Principles of analytic validation of immunohistochemical assays: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med. 2014;138(11):1432–1443.
34. Sinclair W, Kobalka P, Ren R, et al. Interobserver agreement in programmed cell death-ligand 1 immunohistochemistry scoring in nonsmall cell lung carcinoma cytologic specimens. Diagn Cytopathol. 2021;49(2):219–225.
35. Tejerina E, Garca Tobar L, Echeveste JI, de Andrea CE, Vigliar E, Lozano MD. PD-L1 in cytological samples: a review and a practical approach. Front Med (Lausanne). 2021;8:668612.
36. Thunnissen E, Kerr KM, Dafni U, et al. Programmed death-ligand 1 expression influenced by tissue sample size. Scoring based on tissue microarrays' and cross-validation with resections, in patients with, stage I-III, non-small cell lung carcinoma of the European Thoracic Oncology Platform Lungscape cohort. Mod Pathol. 2020;33(5):792–801.
37. Dong Z, Liu Y, Jiang T, et al. Cell block as a surrogate for programmed death-ligand 1 staining testing in patients of non-small cell lung cancer. J Cancer. 2020;11(3):551–558.
38. Azuma K, Ota K, Kawahara A, et al. Association of PD-L1 overexpression with activating EGFR mutations in surgically resected nonsmall-cell lung cancer. Ann Oncol. 2014;25(10):1935–1940.
39. Ota K, Azuma K, Kawahara A, et al. Induction of PD-L1 expression by the EML4-ALK oncoprotein and downstream signaling pathways in non-small cell lung cancer. Clin Cancer Res. 2015;21(17):4014–4021.
40. Karatrasoglou EA, Chatziandreou I, Sakellariou S, et al. Association between PD-L1 expression and driver gene mutations in non-small cell lung cancer patients: correlation with clinical data. Virchows Arch. 2020;477(2):207–217.
41. Jain E, Sharma S, Aggarwal A, et al. PD-L1 expression and its clinicopathologic and genomic correlation in the non-small cell lung carcinoma patients: an Indian perspective. Pathol Res Pract. 2021;228:153497.
42. Evans M, O'Sullivan B, Hughes F, et al. The clinicopathological and molecular associations of PD-L1 expression in non-small cell lung cancer: analysis of a series of 10,005 cases tested with the 22C3 assay. Pathol Oncol Res. 2020;26(1):79–89.
43. Kato Y, Kashima J, Watanabe K, et al. Association between clinicopathological features and programmed death ligand 1 expression in non-small cell lung cancer. Anticancer Res. 2018;38(2):1077–1083.
44. Holmes M, Mahar A, Lum T, Kao S, Cooper WA. Real-world programmed death-ligand 1 prevalence rates in non-small cell lung cancer: correlation with clinicopathological features and tumour mutation status. J Clin Pathol. 2021;74(2):123–128.
45. Sorensen SF, Zhou W, Dolled-Filhart M, et al. PD-L1 Expression and survival among patients with advanced non-small cell lung cancer treated with chemotherapy. Transl Oncol. 2016;9(1):64–69.
46. Li H, Xu Y, Wan B, et al. The clinicopathological and prognostic significance of PD-L1 expression assessed by immunohistochemistry in lung cancer: a meta-analysis of 50 studies with 11,383 patients. Transl Lung Cancer Res. 2019;8(4):429–449.
47. Sun JM, Zhou W, Choi YL, et al. Prognostic significance of PD-L1 in patients with non-small cell lung cancer: a large cohort study of surgically resected cases. J Thorac Oncol. 2016;11(7):1003–1011.

48. Pan Y, Zheng D, Li Y, et al. Unique distribution of programmed death ligand 1 (PD-L1) expression in East Asian non-small cell lung cancer. J Thorac Dis. 2017;9(8):2579–2586.
49. Mansour MSI, Hejny K, Johansson F, et al. Factors influencing concordance of PD-L1 expression between biopsies and cytological specimens in non-small cell lung cancer. Diagnostics (Basel). 2021;11(10):1927.
50. Zhu Y, Lin S, Wang Y, et al. Study on PD-L1 expression in NSCLC patients and related influencing factors in the real world. Comput Math Methods Med. 2021;2021:3273347.
51. Vigliar E, Iaccarino A, Campione S, et al. PD-L1 expression in cell-blocks of non-small cell lung cancer: the impact of prolonged fixation. Diagn Cytopathol. 2020;48(7):595–603.

SUPPLEMENTARY DATA

Supplementary information 1

Questionnaire concerning fixation and cell block processing methods of cytological material used for PD-L1 immunostaining in NSCLC patients

1. Is PD-L1 immunostaining for NSCLC patients performed in your own laboratory or in an external laboratory?
   - Own laboratory
   - External laboratory
   - Other, namely: ...

2. How many ways of processing cytology samples before performance of PD-L1 immunostaining are used within your laboratory? (e.g. different methods for pleural effusions and fine needle aspirations (FNA))
   - 1 way
   - 2 ways
   - 3 ways
   - Other, namely: ...

In case of >1 method, respondents were asked to answer question 3-8 for each method separately.

3. For which type of material are you answering the following questions? (e.g. pleural effusions, FNA, bronchial lavage)
...

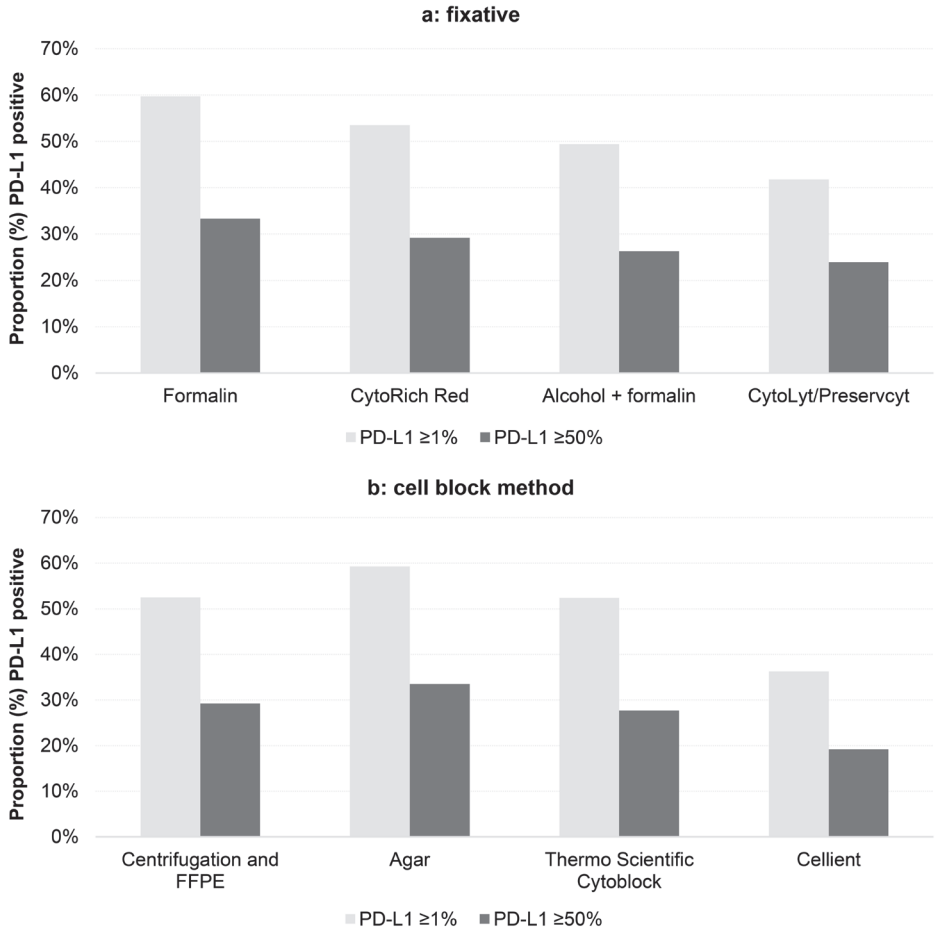
4. What medium is the cytological material collected in? (please be as specific as you can)
...

5. Which fixative is used? (please be as specific as you can)
...

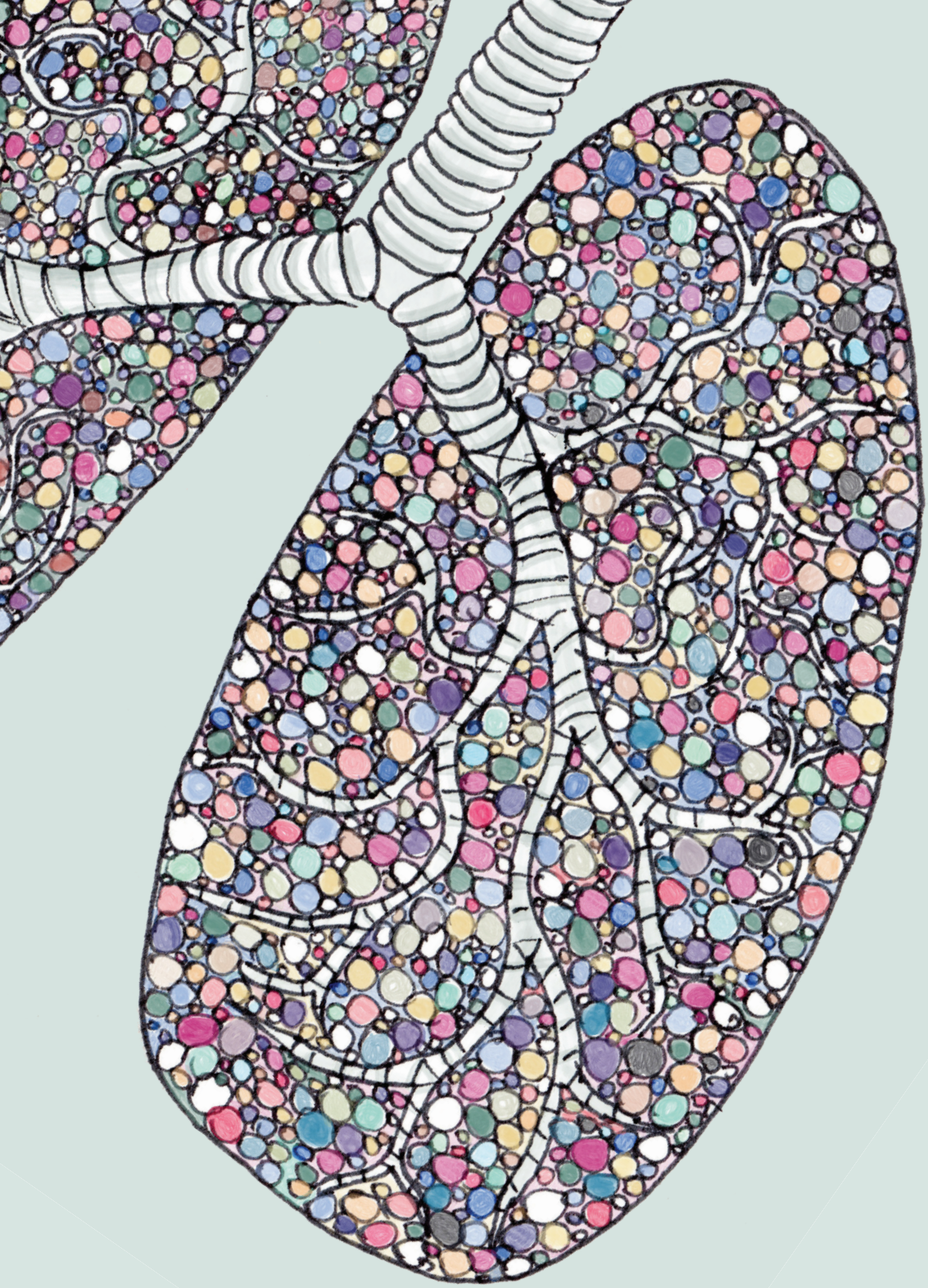
6. What is the estimated fixation time?
   Minimum: ...
   Maximum: ...
   Mean: ...

7. Which intermediate steps are used, if any? (e.g. post-fixation, rinse step, etcetera)
...

8. Which method is used to create a cell block? (please be as specific as you can)
...



Supplementary figure 1. Mean PD-L1 positivity rates per fixative (**a**) and per cell block method (**b**). PD-L1 positivity is determined using either a ≥1% cutoff or a ≥50% cutoff. Abbreviations: FFPE = formalin-fixed paraffin-embedded; PD-L1 = programmed death-ligand 1.



7

SUMMARIZING DISCUSSION AND FUTURE PERSPECTIVES

SUMMARIZING DISCUSSION

Cancer immunotherapy is rapidly gaining importance in clinical care, and various interactions between the tumor and the immune system have been suggested as grounds for development of biomarkers that predict response to immunotherapy[1]. One of those biomarkers is expression of PD-L1 assessed through IHC, which, although certainly not perfect, is one of the most common immune-based biomarkers currently used in clinical practice[2]. For patients with NSCLC, immunohistochemical expression of PD-L1 on tumor cells is crucial in guiding decisions on treatment with PD-1/PD-L1 blockade agents, and accurate and reproducible testing of PD-L1 is therefore of the utmost importance. In general, immunostaining results are influenced by pre-analytic, analytic, and post-analytic factors[3, 4]. For PD-L1 IHC specifically, these may include, for instance, sample size[5] and type of fixation[6-8] (pre-analytic), choice of IHC assay or LDT[9, 10] (analytic), and reproducibility of the scoring performed by the pathologist[9] (post-analytic). Differences in pre-analytic, analytic, and post-analytic factors between laboratories could influence reproducibility of PD-L1 IHC assessment in clinical practice. In this thesis, we assessed the impact of various pre-analytic, analytic, and post-analytic variables on PD-L1 immunostaining results. Subsequently, we used real-world clinical pathology data of a nationwide cohort of NSCLC patients to analyze interlaboratory variation in PD-L1 positivity in both histological and cytological material, and investigated if this variation is influenced by differences in cytology fixation and processing methods.

In **chapter 2** we performed a systematic search to identify all available literature on interassay, interobserver and interlaboratory comparability of PD-L1 IHC assays and LDTs. We showed that interassay agreement is generally moderate to high between commercial assays 22C3, 28-8 and SP263[11-19], while comparisons between these assays and SP142 show much lower concordance rates[11, 13-17, 19-22]. Comparisons of LDTs with commercial assays also resulted in high concordance levels in various studies, provided that the LDTs were properly optimized and validated[13-15, 20, 23-27]. Use of cutoffs to classify PD-L1 TPS as positive or negative, however, resulted in lower concordance levels in some studies[11, 13, 16-18, 26]. In current clinical practice, a 1% cutoff and a 50% cutoff are relevant for NSCLC patients, seeing that, based on various clinical trials, treatment guidelines recommend first-line prescription of pembrolizumab, cemiplimab or atezolizumab as monotherapy to patients with metastatic NSCLC with a PD-L1 expression of ≥50%[28-30] and prescription of durvalumab as consolidation treatment after chemoradiotherapy to patients with locally advanced unresectable NSCLC with a PD-L1 expression of ≥1%[31]. Discordance in PD-L1 results at these clinically relevant cutoffs could therefore impact treatment decisions for NSCLC patients, and thus it seems that simply interchanging PD-L1 IHC assays could be problematic in clinical practice. A similar conclusion was drawn by others[10, 13, 26]. In fact, a recent study showed that the commercial PD-L1 assays actually represent different levels

of analytic sensitivity, explaining why some tissue samples are scored as positive by one assay and as negative by another[32].

Based on our systematic search, interobserver variability in PD-L1 scoring proved to be high for all commercial assays and various LDTs[12, 15, 16, 19-21, 24, 26, 27, 33-37]. Multiple studies, however, reported lower concordance levels when PD-L1 positivity was scored using a 1% cutoff compared to a 50% cutoff[12, 15, 16, 18-20, 33], suggesting that determining PD-L1 positivity using a 1% cutoff is harder than determining PD-L1 positivity at the 50% cutoff level. The use of the 1% cutoff to guide treatment decisions on durvalumab as consolidation treatment can be disputed, since the decision of the EMA to indicate durvalumab for the treatment of unresectable stage III NSCLC in patients whose tumors express PD-L1 on ≥ 1% of tumor cells was based on a post hoc analysis in a trial that, according to the researchers, "was not designed to evaluate the efficacy of durvalumab based on PD-L1 status"[38]. For this reason, in the Netherlands durvalumab can be prescribed to patients with stage III NSCLC regardless of PD-L1 expression[39, 40]. Even so, in various European countries the 1% cutoff is indeed relevant for prescription of durvalumab, meaning that accurate scoring of PD-L1 around this cutoff is still of the utmost importance.

Interlaboratory concordance of PD-L1 IHC assays and LDTs, finally, was assessed by a limited number of studies according to our systematic search, which showed that concordance is generally high between laboratories for each of the individual commercial assays[14, 18, 35, 37, 41], but only moderate for LDTs compared to commercial assays[41]. None of these studies, however, assessed variation in PD-L1 expression between laboratories in a real-world clinical setting.

We did evaluate interlaboratory variation in PD-L1 positivity in a real-world clinical setting and on a nationwide level, for which we used real-world clinical pathology data extracted from PALGA (the nationwide network and registry of histo- and cytopathology in the Netherlands). Results from this study are discussed in **chapter 5**. We discovered that variation in PD-L1 positivity rates between laboratories is substantial, even after correction for differences in patient and sample characteristics. The interlaboratory variation in PD-L1 positivity based on histology data was a lot greater at the 1% cutoff than at the 50% cutoff. Most likely, this is brought about by interobserver variability in PD-L1 scoring to a great extent. After all, as we demonstrated in **chapter 2**, interpathologist variation is often reported to be higher at the 1% cutoff than at the 50% cutoff. We also evaluated if differences in PD-L1 IHC assays and LDTs between laboratories play a role in causing interlaboratory variation in PD-L1 positivity, although it should be emphasized that our data set was not entirely sufficient for this analysis. Correcting for these differences resulted in a decrease in variation in PD-L1 positivity between laboratories, suggesting that differences in PD-L1 IHC assays and LDTs between laboratories do have an impact on interlaboratory variation

in PD-L1 positivity in a clinical setting. Another study that assessed the distribution of PD-L1 expression by assay type in a real-world clinical setting found no significant difference between two commercial assays (22C3 and 28-8), but did find statistically significant differences between these commercial assays and LDTs[42]. Similarly, a study that evaluated results from EQA schemes found better concordance of PD-L1 immunostaining results for commercial assays than for LDTs[43]. All in all, even though no definitive conclusions can be drawn on the impact of differences in PD-L1 IHC assay use on interlaboratory variation in PD-L1 expression based on our study results, the available evidence seems to suggest that interchangeability of assays and LDTs cannot be assumed.

Remarkably, in **chapter 5**, interlaboratory variation in PD-L1 positivity based on cytology data was considerable at both the 1% cutoff and the 50% cutoff, suggesting that other factors specifically related to PD-L1 testing on cytology might play a role in causing variation between laboratories, too. In **chapter 3** we assessed concordance of PD-L1 immunostaining results between cytological cell blocks (CBs) and matching tissue specimens, by using FNA samples and histologic tissue from the same resected lung tumor. We found that overall concordance was moderate to substantial between all CBs and histology. Analyzing agar-based CBs and CBs processed with the Cellient Automated Cell Block System separately, however, resulted in much lower concordance levels for Cellient CBs compared with histology than for agar CBs compared with histology. We hypothesized that these differences in concordance levels might be explained by use of different fixatives. After all, all agar-based cytology samples in our study were fixed in formalin, whereas all Cellient-processed samples were fixed in CytoLyt and PreservCyt, which are methanol-based fixatives. We therefore used PD-L1 expressing cell lines to study the influence of various alcohol-based fixatives, and found that use of methanol-based and ethanol-based fixatives resulted in lower PD-L1 staining intensity compared to formalin. This effect was seen with use of SP263 and 22C3 commercial assays and with the 22C3 antibody used in an LDT. Similarly, in **chapter 4**, we discovered that ethanol fixation of EBUS-TBNA samples prior to formalin fixation resulted in a substantial amount of false-negative PD-L1 immunostaining results compared to pure formalin fixation, at both the 1% cutoff (22C3 LDT and SP263 commercial assay) and the 50% cutoff (22C3 LDT).

A negative effect of alcohol-fixation on PD-L1 immunostaining results has been demonstrated previously[44], and similar effects have been demonstrated on other antibodies, too[6, 7, 45]. Contrastingly, some studies have concluded that alcohol fixation does not affect PD-L1 immunostaining. These studies, however, mainly used samples that were post-fixed in formalin[46-49] or samples that were fixed in CytoRich Red[49, 50], a fixative that also contains formaldehyde. As we demonstrated in **chapter 3**, CytoRich Red fixation did not result in lower PD-L1 immunostaining compared to formalin. Moreover, we showed that the negative effects of alcohol-based fixatives on PD-L1 immunostaining may be reversed

by formalin post-fixation to some degree, although this effect was much stronger when immunostaining was performed with the 22C3 antibody than with the SP263 antibody. In **chapter 4**, all EBUS-TBNA samples that were fixed in an ethanol-based fixative received formalin post-fixation prior to processing into agar-based CBs, yet we still observed false-negative PD-L1 immunostaining results in these samples compared to the ones purely fixed in formalin. Unfortunately, we were unable to analyze whether there was any difference in PD-L1 results between specimens with shorter and longer formalin post-fixation times, since we did not know the individual fixation times for the samples in our study. This goes to show that even though the use of formalin post-fixation to counteract negative effects of alcohol-based fixatives on PD-L1 immunostaining seems promising, more research is needed to determine variables such as the most optimal fixation time and whether the effect differs between the various PD-L1 IHC antibodies.

In **chapter 6** we investigated how many different ways of fixing and processing cytology samples into a CB exist in the Netherlands, by sending out questionnaires to all Dutch pathology laboratories. The amount of variation between laboratories was striking, with 19 different combinations of fixation and CB methods within the 28 laboratories that responded to our questionnaire. Such wide variation between laboratories does not seem uncommon, since other studies have reported comparable findings from surveys sent to laboratories[51-53]. Based on the negative influence of alcohol-based fixatives on PD-L1 immunostaining, which we demonstrated in **chapter 3** and **chapter 4**, it seems likely that differences between laboratories in the way that cytology samples are fixed and processed can influence interlaboratory variation in PD-L1 positivity. Hence, in **chapter 6**, we combined the data from the survey sent to all Dutch laboratories with data on PD-L1 testing performed on cytological samples of NSCLC patients retrieved from PALGA. We discovered that correcting for differences in cytology fixation and CB processing methods resulted in decreased interlaboratory variation in PD-L1 positivity, indicating that differences in cytology fixation and CB processing methods can explain part of the considerable variation in PD-L1 positivity that we found. The use of alcohol-based fixatives without any formalin (post-)fixation especially seems to be problematic when compared with formalin as reference.

In conclusion, with this thesis we have shown that in a real-world clinical setting considerable variation in PD-L1 positivity exists between pathology laboratories, both in PD-L1 positivity rates based on histology samples and in PD-L1 positivity rates based on cytology samples. Most likely, interobserver variability in PD-L1 scoring plays an important part in causing this, especially at the 1% cutoff. When it comes to PD-L1 scoring performed on cytology samples, differences in cytology processing methods and especially fixation methods between laboratories partially explain the considerable interlaboratory variation in PD-L1 positivity, too. Moreover, it is likely that differences between laboratories in the

use of PD-L1 IHC assays and LDTs influence real-world interlaboratory variation in PD-L1 positivity as well, although the extent of this relationship remains uncertain based on this thesis. The degree of interlaboratory variation in PD-L1 positivity that we found is problematic for NSCLC patients, since it could lead to different PD-L1 results depending on the laboratory where their PD-L1 test is performed. This would in turn result in different courses of treatment, seeing that PD-L1 IHC results guide decisions on immunotherapy. We feel that raising awareness of this considerable interlaboratory variation in PD-L1 positivity among all clinicians involved in NSCLC patient care is of the utmost importance, and that efforts should be taken to diminish this degree of variation.

FUTURE PERSPECTIVES

In order to try to diminish the degree of variation in PD-L1 positivity between laboratories, we believe that a first important step to take is to create awareness. This is why the results from individual laboratories in our study on interlaboratory variation in PD-L1 positivity were sent back to these laboratories as feedback reports, encouraging pathologists to discuss and reflect on their own results concerning PD-L1 testing compared to other laboratories in the Netherlands. When laboratories discover that they differ quite strongly from the overall mean and from other laboratories, this might stimulate them to investigate what could cause these differences and to think of ways to improve their own PD-L1 results. A similar initiative with feedback reports on grading of invasive breast cancer has in fact been shown to lead to an encouraging decrease in grading variation between laboratories[54]. Another example of such an initiative is one by the Netherlands Expertise Network Cytology (part of the Dutch Society for Pathology (NVVP)), which provided participating laboratories with information on the performance of their own PD-L1 IHC stain on cytological samples fixed and processed according to their own routinely used methods. Giving laboratories an insight into their own performance might be a more powerful tool to instigate quality improvement than simply sharing results from studies conducted with anonymous data, so we believe that initiatives like the ones mentioned here could be very valuable in this regard.

Since interobserver variability in PD-L1 scoring between pathologists most likely contributes greatly to interlaboratory variation in PD-L1 positivity, it would be worthwhile to investigate ways to diminish the amount of interobserver variability. Several possibilities have been suggested in literature. Proper training for PD-L1 scoring, for instance, might lead to higher interobserver concordance[55, 56], although others showed little to no effect of training on reproducibility of PD-L1 scoring[34]. Perhaps digital image analysis could also improve interobserver concordance, based on promising results from various studies[57-61]. Future studies, however, are needed to determine whether the use of artificial intelligence tools for PD-L1 scoring actually leads to increased reproducibility of PD-L1 results in the real-world setting[62]. Moreover, in this thesis, variation in PD-L1 positivity rates was assessed at an interlaboratory level and not at an interpathologist level, since information on PD-L1 scoring by individual pathologists was not part of the PALGA data set. Such assessment at an interpathologist level, however, could help in determining the degree of interobserver variability in PD-L1 scoring in clinical practice, and feedback reports that contain this kind of information could provide pathologists with insightful information on their own scoring results compared to those of others. In order for these kinds of analyses to succeed, willingness of enough pathologists to participate is crucial.

When it comes to PD-L1 testing on cytology specimens, standardization of methods for fixation and processing, using a (combination of) method(s) that does not negatively

influence PD-L1 immunostaining, could aid in decreasing interlaboratory variation in PD-L1 positivity. In order to achieve such standardization, future studies should be conducted to determine the method that leads to the most optimal PD-L1 staining results. Based on the available literature and this thesis, this method will likely contain some form of formalin (post-)fixation, although the specific amount and duration of fixation is as yet unclear. Incorporation of clinical outcome data in studies assessing methods for cytology fixation and processing would also be valuable, since it would allow for assessment of the predictive value of PD-L1 immunostaining performed on cytology samples fixed and processed in different ways. Linking of clinical pathology data, such as the data derived from PALGA in this thesis, to data from registries that record information on treatment and follow-up, such as the Netherlands Cancer Registry by the Netherlands Comprehensive Cancer Organization (IKNL), could prove very interesting in this regard.

Standardization of other processes besides cytology fixation and CB processing might also prove effective in diminishing variation in PD-L1 positivity between laboratories, and it could be worthwhile to explore this further. It has been shown, for instance, that performance of PD-L1 immunostaining on small samples may lead to underestimation of PD-L1 expression[5], which likely applies to both histology and cytology. Differences in PD-L1 positivity rates between laboratories could arise, for example, when some laboratories structurally receive smaller samples than others. Assessment of differences between laboratories in such pre-analytic factors may prove insightful, and could spark a discussion between pathologists and other clinicians on the best ways to obtain material for PD-L1 testing. Standardization based on the use of the same PD-L1 IHC assay by all laboratories might also help in diminishing interlaboratory variation, but may not be entirely feasible. Participation in EQA schemes, however, could perhaps help in ensuring that the quality of individual laboratories' PD-L1 IHC assays and LDTs adheres to the same standard, thereby improving interlaboratory concordance[43]. Also, implementation of quantitative analytic standards, such as IHC calibrators measuring the lowest analyte concentration that produces a visual stain in individual IHC tests, could potentially prove to be helpful in harmonizing PD-L1 IHC between laboratories[32, 63]. Nonetheless, the fact that each PD-1/PD-L1 blockade agent was developed with its own PD-L1 IHC assay has complicated PD-L1 testing in daily pathology practice. For example, specific assays only run best on specific staining platforms that might not be widely available. This important lesson should be taken into account when new drugs that require companion diagnostics are developed and assessed in clinical trials.

In order to achieve implementation of measures aimed at quality improvement, it is crucial that laboratories discuss their pitfalls and successes in PD-L1 testing with each other and join together in thinking of ways to diminish interlaboratory variability. We feel that it is important to facilitate such dialogue, since it may not come natural to everyone to open a

conversation about issues one might encounter in routine clinical practice. Regional and national initiatives could prove vital in this regard. An example of such an initiative is the project 'National Implementation of Predictive Analysis in NSCLC', a Dutch collaboration of pharmaceutical and health insurance companies, patient associations, scientific associations and health care professionals. This initiative facilitates dialogue between health care professionals to improve quality of predictive diagnostics in NSCLC patients, including PD-L1 testing. Thus far, meetings within this project have been held within the northern region of the Netherlands, but this will be extended to other regions in the future. Of course, re-evaluating the degree of interlaboratory variation in PD-L1 positivity after implementation of quality improvement measures is important to assess if these measures have the intended effect. Regularly sending feedback reports on PD-L1 testing and reports on improvement ideas from joint discussions to individual laboratories could aid in monitoring the degree of interlaboratory variation and in continuously encouraging laboratories to deliver high quality results.

Although this thesis focusses on PD-L1 testing in material from NSCLC patients, its conclusions are relevant to patients with other types of cancer, too. In the past few years PD-L1 immunohistochemistry has become relevant for decisions on immunotherapy in various cancer types besides NSCLC, such as squamous cell carcinoma of head and neck[64], urothelial cell carcinoma[65], esophageal squamous cell carcinoma[66], triple-negative breast cancer[67], and cervical cancer[68]. Moreover, while PD-L1 scoring in NSCLC patients is based on the percentage of tumor cells that show expression of PD-L1, PD-L1 scoring for other tumors may be based on the percentage of positively staining immune cells[67], or the ratio of all positively staining cells (tumor cells and immune cells) relative to all viable tumor cells (combined positive score (CPS))[64, 68-70]. Also, cutoffs for PD-L1 positivity differ between tumor types, with a CPS of ≥10 being relevant for metastatic urothelial cell carcinoma[69], for instance. Proper training and ongoing education are therefore crucial for accurate PD-L1 interpretation in the various tumor types[71]. Additional analysis of interlaboratory variation in PD-L1 results in tumor types other than NSCLC could prove useful in this regard. This could initiate dialogue on quality improvement of PD-L1 testing in these areas, too, and help in achieving high quality PD-L1 results across all settings in which PD-L1 testing is relevant.

Finally, in this era of precision medicine, it is likely that predictive markers will be used more and more to guide treatment decisions for individual patients[72]. Potentially, other IHC assays than those for PD-L1 will be used as companion diagnostics for specific therapies[73, 74], requiring quantitative assessment in a similar fashion to PD-L1 IHC. The issues in pre-analytic, analytic and post-analytic phases that were discussed in this thesis, may also influence results of other IHC assays used as predictive biomarkers. Thus, we hope that this thesis may serve as a starting point for quality improvement of PD-L1 testing, and potentially of other predictive IHC markers, too.

REFERENCES

1. Blank CU, Haanen JB, Ribas A, Schumacher TN. CANCER IMMUNOLOGY. The “cancer immunogram”. Science. 2016;352(6286):658-660.
2. Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. J Immunother Cancer. 2019;7(1):278.
3. Bogen SA. A Root Cause Analysis Into the High Error Rate in Clinical Immunohistochemistry. Appl Immunohistochem Mol Morphol. 2019;27(5):329-338.
4. Lin F, Shi J. Standardization of Diagnostic Immunohistochemistry. In: Lin F, Pritchard J, editors. Handbook of Practical Immunohistochemistry. New York, NY, USA: Springer; 2015.
5. Thunnissen E, Kerr KM, Dafni U, et al. Programmed death-ligand 1 expression influenced by tissue sample size. Scoring based on tissue microarrays’ and cross-validation with resections, in patients with, stage I-III, non-small cell lung carcinoma of the European Thoracic Oncology Platform Lungscape cohort. Mod Pathol. 2020;33(5):792-801.
6. Sauter JL, Grogg KL, Vrana JA, Law ME, Halvorson JL, Henry MR. Young investigator challenge: Validation and optimization of immunohistochemistry protocols for use on cellient cell block specimens. Cancer Cytopathol. 2016;124(2):89-100.
7. Gruchy JR, Barnes PJ, Dakin Hache KA. CytoLyt(R) fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. Appl Immunohistochem Mol Morphol. 2015;23(4):297-302.
8. Kruger AM, Stevens MW, Kerley KJ, Carter CD. Comparison of the Cellient() automated cell block system and agar cell block method. Cytopathology. 2014;25(6):381-388.
9. Lantuejoul S, Sound-Tsao M, Cooper WA, et al. PD-L1 Testing for Lung Cancer in 2019: Perspective From the IASLC Pathology Committee. J Thorac Oncol. 2020;15(4):499-519.
10. Torlakovic E, Lim HJ, Adam J, et al. “Interchangeability” of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Mod Pathol. 2020;33(1):4-17.
11. Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project. J Thorac Oncol. 2017;12(2):208-222.
12. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between Programmed Cell Death Ligand-1 Diagnostic Assays across Multiple Protein Expression Cutoffs in Non-Small Cell Lung Cancer. Clin Cancer Res. 2017;23(14):3585-3591.
13. Hendry S, Byrne DJ, Wright GM, et al. Comparison of Four PD-L1 Immunohistochemical Assays in Lung Cancer. J Thorac Oncol. 2018;13(3):367-376.
14. Adam J, Le Stang N, Rouquette I, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. Ann Oncol. 2018;29(4):953-958.
15. Brunnström H, Johansson A, Westbom-Fremer S, et al. PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: Inter-pathologist variability is higher than assay variability. Modern Pathology. 2017;30(10):1411-1421.
16. Chan AWH, Tong JHM, Kwan JSH, et al. Assessment of programmed cell death ligand-1 expression by 4 diagnostic assays and its clinicopathological correlation in a large cohort of surgical resected non-small cell

lung carcinoma. Mod Pathol. 2018. DOI: 10.1038/s41379-018-0053-3.

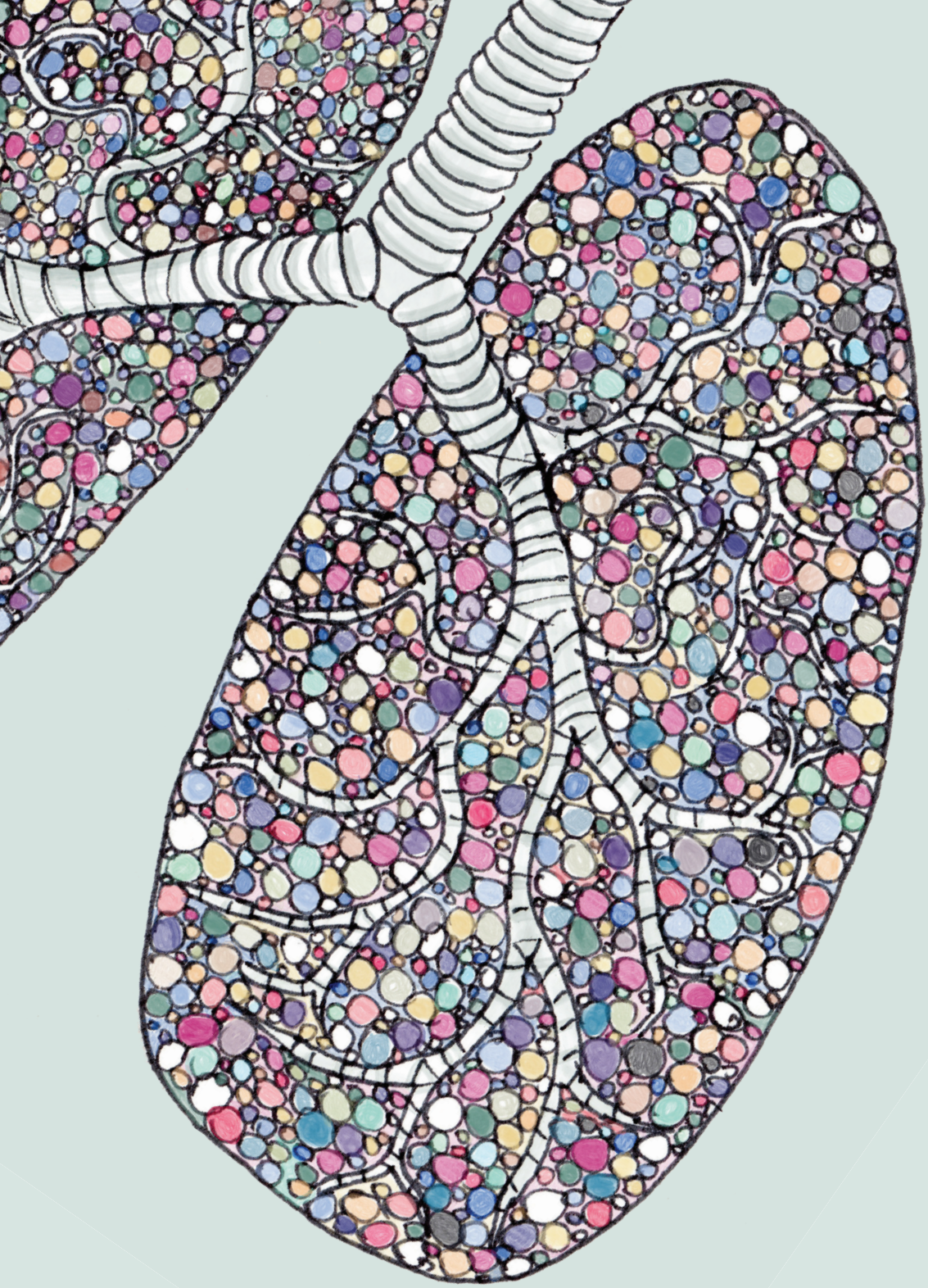
17. Fujimoto D, Sato Y, Uehara K, et al. Predictive Performance of Four Programmed Cell Death Ligand 1 Assay Systems on Nivolumab Response in Previously Treated Patients with Non-Small Cell Lung Cancer. J Thorac Oncol. 2018;13(3):377-386.
18. Marchetti A, Barberis M, Franco R, et al. Multicenter Comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) Assays to Test PD-L1 Expression for NSCLC Patients to Be Treated with Immune Checkpoint Inhibitors. J Thorac Oncol. 2017;12(11):1654-1663.
19. Tsao MS, Kerr KM, Kockx M, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of Blueprint phase 2 project. J Thorac Oncol. 2018. DOI: 10.1016/j.jtho.2018.05.013.
20. Rimm DL, Han G, Taube JM, et al. A Prospective, Multi-institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. JAMA Oncol. 2017;3(8):1051-1058.
21. Ilie M, Falk AT, Butori C, et al. PD-L1 expression in basaloid squamous cell lung carcinoma: Relationship to PD-1(+) and CD8(+) tumor-infiltrating T cells and outcome. Mod Pathol. 2016;29(12):1552-1564.
22. Xu H, Lin G, Huang C, et al. Assessment of Concordance between 22C3 and SP142 Immunohistochemistry Assays regarding PD-L1 Expression in Non-Small Cell Lung Cancer. Sci Rep. 2017;7(1):16956.
23. Soo RA, Lim JSY, Asuncion BR, et al. Determinants of variability of five programmed death ligand-1 immunohistochemistry assays in non-small cell lung cancer samples. Oncotarget. 2018;9(6):6841-6851.
24. Ilie M, Khambata-Ford S, Copie-Bergman C, et al. Use of the 22C3 anti-PD-L1 antibody to determine PD-L1 expression in multiple automated immunohistochemistry platforms. PLoS One. 2017;12(8):e0183023.
25. Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death ligand 1 antibody to determine programmed death ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. Cancer Cytopathol. 2018. DOI: 10.1002/cncy.21977.
26. Munari E, Rossi G, Zamboni G, et al. PD-L1 Assays 22C3 and SP263 are Not Interchangeable in Non-Small Cell Lung Cancer When Considering Clinically Relevant Cutoffs: An Interclone Evaluation by Differently Trained Pathologists. Am J Surg Pathol. 2018. DOI: 10.1097/pas.0000000000001105.
27. Conde E, Caminoa A, Dominguez C, et al. Aligning digital CD8(+) scoring and targeted next-generation sequencing with programmed death ligand 1 expression: a pragmatic approach in early-stage squamous cell lung carcinoma. Histopathology. 2018;72(2):270-284.
28. Hendriks LE, Kerr KM, Menis J, et al. Non-oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023;34(4):358-376.
29. Wu YL, Planchard D, Lu S, et al. Pan-Asian adapted Clinical Practice Guidelines for the management of patients with metastatic non-small-cell lung cancer: a CSCO-ESMO initiative endorsed by JSMO, KSMO, MOS, SSO and TOS. Ann Oncol. 2019;30(2):171-210.
30. Akinboro O, Larkins E, Pai-Scherf L, et al. FDA Approval Summary: Pembrolizumab, Atezolizumab, and Cemiplimab-rwlc as Single Agents for First-Line Treatment of Advanced/Metastatic PD-L1–High NSCLC. Clin Cancer Res. 2022;28(11):2221-2228.
31. Remon J, Soria JC, Peters S, Esmo Guidelines Committee. Early and locally advanced non-small-cell lung cancer: an update of the ESMO Clinical Practice Guidelines focusing on diagnosis, staging, systemic and local therapy. Ann Oncol. 2021;32(12):1637-1642.

32. Sompuram SR, Torlakovic EE, t Hart NA, Vani K, Bogen SA. Quantitative comparison of PD-L1 IHC assays against NIST standard reference material 1934. Mod Pathol. 2021. DOI: 10.1038/s41379-021-00884-w.
33. Russell-Goldman E, Kravets S, Dahlberg SE, Sholl LM, Vivero M. Cytologic-histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. Cancer Cytopathol. 2018;126(4):253-263.
34. Cooper WA, Russell PA, Cherian M, et al. Intra- and Interobserver Reproducibility Assessment of PD-L1 Biomarker in Non-Small Cell Lung Cancer. Clin Cancer Res. 2017;23(16):4569-4577.
35. Roach C, Zhang N, Corigliano E, et al. Development of a Companion Diagnostic PD-L1 Immunohistochemistry Assay for Pembrolizumab Therapy in Non-Small-cell Lung Cancer. Appl Immunohistochem Mol Morphol. 2016;24(6):392-397.
36. Rehman JA, Han G, Carvajal-Hausdorf DE, et al. Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. Mod Pathol. 2017;30(3):340-349.
37. Vennapusa B, Baker B, Kowanetz M, et al. Development of a PD-L1 Complementary Diagnostic Immunohistochemistry Assay (SP142) for Atezolizumab. Appl Immunohistochem Mol Morphol. 2018. DOI: 10.1097/pai.0000000000000594.
38. Gray JE, Villegas A, Daniel D, et al. Three-Year Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC-Update from PACIFIC. J Thorac Oncol. 2020;15(2):288-293.
39. NVMO-commissie BOM. Durvalumab als consolidatietherapie na chemoradiatie bij stadium III NSCLC [in Dutch]. Medische Oncologie. 2020;23(10):31-35.
40. Federatie Medisch Specialisten Richtlijnendatabase. Niet-kleincellig longcarcinoom - Adjuvante immunotherapie [in Dutch], 2020. Available from: https://richtlijnendatabase.nl/richtlijn/niet_kleincellig_longcarcinoom/stadium_iii_nsclc/nklc_adjuvante_immunotherapie.html (Accessed 7 February 2022).
41. Scheel AH, Baenfer G, Baretton G, et al. Interlaboratory concordance of PD-L1 immunohistochemistry for non-small-cell lung cancer. Histopathology. 2018;72(3):449-459.
42. Velcheti V, Patwardhan PD, Liu FX, Chen X, Cao X, Burke T. Real-world PD-L1 testing and distribution of PD-L1 tumor expression by immunohistochemistry assay type among patients with metastatic non-small cell lung cancer in the United States. PLoS One. 2018;13(11):e0206370.
43. Keppens C, Dequeker EM, Pauwels P, Ryska A, t Hart N, von der Thusen JH. PD-L1 immunohistochemistry in non-small-cell lung cancer: unraveling differences in staining concordance and interpretation. Virchows Arch. 2021;478(5):827-839.
44. Lloyd IE, Zhou W, Witt BL, Chadwick BE. Characterization of PD-L1 Immunohistochemical Expression in Cell Blocks With Different Specimen Fixation and Processing Methods. Appl Immunohistochem Mol Morphol. 2019;27(2):107-113.
45. Buonocore DJ, Konno F, Jungbluth AA, et al. CytoLyt fixation significantly inhibits MIB1 immunoreactivity whereas alternative Ki-67 clone 30-9 is not susceptible to the inhibition: Critical diagnostic implications. Cancer Cytopathol. 2019;127(10):643-649.
46. Wang H, Agulnik J, Kasymjanova G, et al. Cytology cell blocks are suitable for immunohistochemical testing for PD-L1 in lung cancer. Ann Oncol. 2018;29(6):1417-1422.
47. Lou SK, Ko HM, Kinoshita T, et al. Implementation of PD-L1 22C3 IHC pharmDxTM in Cell Block Preparations of Lung Cancer: Concordance with Surgical Resections and Technical Validation of CytoLyt(R) Prefixation. Acta

Cytol. 2020;64(6):577-587.

48. Arriola AGP, Bashover E, Joseph C, Staerkel G, Wang WL, Roy-Chowdhuri S. The usefulness of various cytologic specimen preparations for PD-L1 immunostaining in non-small cell lung carcinoma. J Am Soc Cytopathol. 2018;7(6):324-332.
49. Gosney JR, Haragan A, Chadwick C, et al. Programmed death ligand 1 expression in EBUS aspirates of non-small cell lung cancer: Is interpretation affected by type of fixation? Cancer Cytopathol. 2020;128(2):100-106.
50. Song SG, Lee J, Koh J, Kim S, Chung DH, Jeon YK. Utility of PD-L1 immunocytochemistry using body-fluid cell blocks in patients with non-small-cell lung cancer. Diagn Cytopathol. 2020;48(4):291-299.
51. Crapanzano JP, Heymann JJ, Monaco S, Nassar A, Saqi A. The state of cell block variation and satisfaction in the era of molecular diagnostics and personalized medicine. Cytojournal. 2014;11:7.
52. Srebotnik Kirbis I, Rodrigues Roque R, Bongiovanni M, Strojan Flezar M, Cochand-Priollet B. Immunocytochemistry practices in European cytopathology laboratories-Review of European Federation of Cytology Societies (EFCS) online survey results with best practice recommendations. Cancer Cytopathol. 2020;128(10):757-766.
53. Fischer AH, Schwartz MR, Moriarty AT, et al. Immunohistochemistry practices of cytopathology laboratories: a survey of participants in the College of American Pathologists Nongynecologic Cytopathology Education Program. Arch Pathol Lab Med. 2014;138(9):1167-1172.
54. van Dooijeweert C, van Diest PJ, Baas IO, van der Wall E, Deckers IA. Variation in breast cancer grading: the effect of creating awareness through laboratory-specific and pathologist-specific feedback reports in 16 734 patients with breast cancer. J Clin Pathol. 2020;73(12):793-799.
55. Adam J, Hofman V, Mansuet-Lupo A, et al. Real-World Concordance Across Pathologists for PD-L1 Scoring in Non-Small Cell Lung Cancer: Results from a Large Nationwide Initiative. Journal of Thoracic Oncology. 2019;14(10):S775.
56. Chang S, Park HK, Choi YL, Jang SJ. Interobserver Reproducibility of PD-L1 Biomarker in Non-small Cell Lung Cancer: A Multi-Institutional Study by 27 Pathologists. J Pathol Transl Med. 2019;53(6):347-353.
57. Koelzer VH, Gisler A, Hanhart JC, et al. Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. Histopathology. 2018;73(3):397-406.
58. Kapil A, Meier A, Zuraw A, et al. Deep Semi Supervised Generative Learning for Automated Tumor Proportion Scoring on NSCLC Tissue Needle Biopsies. Sci Rep. 2018;8(1):17343.
59. Taylor CR, Jadhav AP, Gholap A, et al. A Multi-Institutional Study to Evaluate Automated Whole Slide Scoring of Immunohistochemistry for Assessment of Programmed Death-Ligand 1 (PD-L1) Expression in Non-Small Cell Lung Cancer. Appl Immunohistochem Mol Morphol. 2019;27(4):263-269.
60. Wu J, Liu C, Liu X, et al. Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer. Mod Pathol. 2021.
61. Cheng G, Zhang F, Xing Y. Artificial Intelligence-Assisted Score Analysis for Predicting the Expression of the Immunotherapy Biomarker PD-L1 in Lung Cancer. Front Immunol. 2022;13:893198.
62. Wu J, Lin D. A Review of Artificial Intelligence in Precise Assessment of Programmed Cell Death-ligand 1 and Tumor-infiltrating Lymphocytes in Non-Small Cell Lung Cancer. Adv Anat Pathol. 2021;28(6):439-445.
63. Bogen SA, Dabbs DJ, Miller KD, et al. A Consortium for Analytic Standardization in Immunohistochemistry. Arch Pathol Lab Med. 2022;147(5):584-590.

64. Machiels JP, Rene Leemans C, Golusinski W, et al. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS-ESMO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020;31(11):1462-1475.
65. Powles T, Bellmunt J, Comperat E, et al. Bladder cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2022;33(3):244-258.
66. U.S. Food & Drug Administration. FDA approves pembrolizumab for advanced esophageal squamous cell cancer, 2019. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-pembrolizumab-advanced-esophageal-squamous-cell-cancer (Accessed 10 November 2023).
67. Gennari A, Andre F, Barrios CH, et al. ESMO Clinical Practice Guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer. Ann Oncol. 2021;32(12):1475-1495.
68. U.S. Food & Drug Administration. FDA approves pembrolizumab combination for the first-line treatment of cervical cancer, 2021. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-pembrolizumab-combination-first-line-treatment-cervical-cancer?utm_medium=email&utm_source=govdelivery (Accessed 9 November 2023).
69. PD-L1 IHC 22C3 pharmDx Interpretation Manual – Urothelial Carcinoma. Agilent Dako, 2021. Available from: https://www.agilent.com/cs/library/usermanuals/public/29276_22C3_pharmdx_uc_interpretation_manual_us.pdf (Accessed 16 December 2021).
70. PD-L1 IHC 22C3 pharmDx Interpretation Manual - Esophageal Squamous Cell Carcinoma (ESCC), 2022. Available from: https://www.agilent.com/cs/library/usermanuals/public/D54358%20rev01%20KN181%20ESCC%20Interpretation%20Manual.pdf (Accessed 10 November 2023).
71. Savic Prince S, Bubendorf L. Predictive potential and need for standardization of PD-L1 immunohistochemistry. Virchows Arch. 2019;474(4):475-484.
72. Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer. 2008;44(7):946-953.
73. Taylor CR. Predictive biomarkers and companion diagnostics. The future of immunohistochemistry: “in situ proteomics,” or just a “stain”? Appl Immunohistochem Mol Morphol. 2014;22(8):555-561.
74. Gatalica Z, Feldman R, Vranic S, Spetzler D. Immunohistochemistry-Enabled Precision Medicine. Cancer Treat Res. 2019;178:111-135.



&

APPENDICES

List of abbreviations

Summary in Dutch / Nederlandse samenvatting

Acknowledgements / Dankwoord

Curriculum Vitae

List of publications

LIST OF ABBREVIATIONS

AC	adenocarcinoma
ASC	adenosquamous carcinoma
CAP	College of American Pathologists
CB	cell block
CCMO	Centrale Commissie Mensgebonden Onderzoek / Central Committee on Research involving Human Subjects
CE	Conformité Européene
CI	confidence interval
CMA	cell-microarray
CPS	combined positive score
CWZ	Canisius Wilhelmina Hospital
EBUS-TBNA	endobronchial ultrasound-guided transbronchial needle aspiration
EMA	European Medicines Agency
EQA	external quality assessment
ESMO	European Society for Medical Oncology
FDA	Food and Drug Administration
FFPE	formalin-fixed paraffin-embedded
FNA	fine needle aspiration
GDPR	General Data Protection Regulation
H&E	haemotoxylin and eosin
IC	immune cell
ICC	immunocytochemistry
ICC	intraclass correlation coefficient
IHC	immunohistochemistry
IKNL	Integraal Kankercentrum Nederland / Netherlands Comprehensive Cancer Organization
κ	kappa
LCC	large cell carcinoma
LCNEC	large cell neuro-endocrine carcinoma
LDT	laboratory-developed test
LELC	lymphoepithelioma-like carcinoma
LRT	likelihood ratio test
NBF	neutral buffered formalin
NOS	not otherwise specified
NPA	negative percentage agreement
NSCLC	non-small cell lung cancer
NVVP	Nederlandse Vereniging Voor Pathologie / Dutch Society for Pathology
OPA	overall percentage agreement

OR	odds ratio
ρ	Pearson's correlation coefficient
PALGA	Pathologisch-Anatomisch Landelijk Geautomatiseerd Archief / the nationwide network and registry of histo- and cytopathology in the Netherlands
PD-1	programmed death-1
PD-L1	programmed death-ligand 1
PPA	positive percentage agreement
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2
QUIPS	Quality in Prognosis Studies
RoB	risk of bias
SC	sarcomatoid carcinoma
SCC	squamous cell carcinoma
SCLC	small cell lung cancer
SD	standard deviation
TBNA	transbronchial needle aspiration
TMA	tissue-microarray
TC	tumor cell
TPS	tumor proportion score
TTF-1	transcription factor-1
UMCU	University Medical Center Utrecht
χ2	chi-square

SUMMARY IN DUTCH / NEDERLANDSE SAMENVATTING

Niet-kleincellig longcarcinoom

Longkanker is de op één na meest voorkomende vorm van kanker wereldwijd en is de veroorzaker van de meeste kankergerelateerde sterfgevallen van mannen en vrouwen samen. Op basis van weefselonderzoek worden twee hoofdvormen van longkanker onderscheiden, namelijk het kleincellig longcarcinoom (small cell lung cancer; SCLC) en het niet-kleincellig longcarcinoom (non-small cell lung cancer; NSCLC). Het grootste deel van de gevallen van longkanker betreft het niet-kleincellig longcarcinoom. In Nederland wordt jaarlijks bij zo'n 10.000 patiënten de diagnose NSCLC gesteld, vaak pas wanneer de ziekte al is uitgezaaid (stadium IV). De prognose voor deze patiënten is slecht, met een 5-jaarsoverleving van slechts 6% op basis van cijfers tussen 2017 en 2021. Voorheen werden deze patiënten met stadium IV ziekte alleen behandeld met chemotherapie, wat als doel heeft om de tumorcellen te doden, maar vaak ook veel bijwerkingen geeft. Sinds een aantal jaren bestaat er echter ook de optie om deze patiënten te behandelen met immuuntherapie of een combinatie van chemotherapie en immuuntherapie. Immuuntherapie heeft als doel om het eigen afweersysteem van de patiënt te ondersteunen bij het opruimen van tumorcellen. De toevoeging van deze behandelingsopties heeft geleid tot een verbetering van de 5-jaarsoverleving naar 25% voor patiënten die worden behandeld met immuuntherapie en 17% voor patiënten die worden behandeld met een combinatie van chemotherapie en immuuntherapie.

Programmed death-1 receptor en programmed death-ligand 1

Immuuntherapie voor patiënten met NSCLC is gebaseerd op blokkade van de receptor 'programmed death-1' (PD-1) of de bijbehorende ligand 'programmed death-ligand 1' (PD-L1). PD-1 is een zogenaamd 'immune checkpoint', aangezien deze receptor een rol speelt bij de regulatie van de immuunrespons. PD-1 komt tot expressie op het celmembraan van T-cellen. Dit zijn witte bloedcellen die een belangrijk onderdeel vormen van het afweersysteem en ook in actie kunnen komen tegen tumorcellen in het lichaam. De ligand PD-L1 komt normaalgesproken tot expressie op het celmembraan van verschillende cellen in het menselijk lichaam, bijvoorbeeld op macrofagen. Wanneer PD-1 en PD-L1 met elkaar binden, zorgt dit voor inactivatie van de T-cel, waardoor een overmatige immuunreactie kan worden voorkomen. De interactie tussen PD-1 en PD-L1 speelt daarmee in een normale situatie een belangrijke rol in het voorkomen van auto-immuniteit. Verschillende kankersoorten, echter, blijken ook PD-L1 op hun cellen tot expressie te brengen. Wanneer PD-L1 op tumorcellen bindt met PD-1 op T-cellen, zorgt dit eveneens voor inactivatie van de T-cel. De tumorcellen kunnen zo ontsnappen aan de afweerreactie van de T-cellen en dus blijven leven. Het doel van immuuntherapie gebruikt bij patiënten met NSCLC is het voorkomen van de binding tussen PD-1 en PD-L1, waardoor de T-cellen hun werk kunnen

blijven doen en de tumorcellen kunnen doden.

Immuuntherapie gericht tegen PD-1 en PD-L1

Er bestaan verschillende medicijnen die als immuuntherapie voor patiënten met NSCLC kunnen worden ingezet, namelijk nivolumab, pembrolizumab en cemiplimab (gebaseerd op blokkade van PD-1) en atezolizumab en durvalumab (gebaseerd op blokkade van PD-L1). Deze medicijnen worden ook wel 'immune checkpoint inhibitors' genoemd. Deze immune checkpoint inhibitors zijn getest in verschillende klinische trials en lieten hierin een gunstig effect zien op overleving vergeleken met chemotherapie, met over het algemeen ook minder bijwerkingen vergeleken met chemotherapie. Bij sommige medicijnen werd het gunstige effect op de overleving echter voornamelijk gezien bij patiënten wiens tumoren een bepaalde mate van PD-L1 tot expressie brachten. Dit werd bepaald met behulp van immuunhistochemie (IHC), waarbij aanwezigheid van PD-L1 op cellen zichtbaar wordt gemaakt middels met kleur gelabelde antilichamen gericht tegen PD-L1. Door pathologen kan vervolgens onder de microscoop (of digitaal) bepaald worden welk percentage van alle aanwezige tumorcellen aankleurt en dus PD-L1 tot expressie brengt. Dit percentage van PD-L1 positieve tumorcellen wordt ook wel de 'tumor proportion score' (TPS) genoemd. Bij één van de medicijnen (pembrolizumab) bleek monotherapie met immuuntherapie alleen een significante verbetering in overleving te tonen ten opzichte van chemotherapie wanneer de TPS ten minste 50% bedroeg. Bij weer een ander medicijn (durvalumab) bleek het gunstige effect van immuuntherapie sterker aanwezig bij patiënten wiens tumoren ten minste 1% PD-L1 expressie toonden. In de klinische praktijk wordt bij patiënten met NSCLC die mogelijk in aanmerking komen voor immuuntherapie daarom de PD-L1 expressie van de tumor bepaald middels IHC, waarna de behandelend arts mede aan de hand van de TPS kan bepalen welke behandeling het beste aan de patiënt kan worden gegeven.

PD-L1 immuunhistochemie en beïnvloedende factoren

Aangezien het meten van PD-L1 expressie middels IHC een belangrijke rol speelt in het bepalen van de therapie voor patiënten met NSCLC, is het van groot belang dat PD-L1 expressie accuraat wordt beoordeeld. Bij het uitvoeren van IHC om PD-L1 expressie te bepalen, kunnen verschillende factoren van invloed zijn op de nauwkeurigheid. Dit kunnen pre-analytische factoren zijn, zoals het type materiaal en de grootte van het weefsel dat wordt getest, maar bijvoorbeeld ook het fixatief dat wordt gebruikt om het materiaal te fixeren. Ook analytische factoren kunnen een rol spelen, zoals het type antilichaam dat wordt gebruikt voor het uitvoeren van het kleuringsprotocol. Een zeer belangrijke post-analytische factor die een rol speelt, is het scoren van de PD-L1 expressie door de patholoog. Verschillen in al deze factoren tussen pathologielaboratoria zouden kunnen zorgen voor een grote mate aan variatie in PD-L1 positiviteit tussen deze laboratoria. Dit zou zeer onwenselijk zijn, aangezien een patiënt dan wellicht in het ene laboratorium een andere uitslag van PD-L1 IHC zou krijgen dan in een ander laboratorium, wat belangrijke

behandelingsconsequenties zou kunnen hebben. In dit proefschrift is daarom van verschillende factoren onderzocht welke invloed zij hebben op het resultaat van PD-L1 immunokleuringen, en welke mate aan variatie in gemiddelde PD-L1 positiviteit er bestaat tussen pathologielaboratoria in Nederland.

Analytisch: interassayvariatie

In **hoofdstuk 2** hebben we een literatuurstudie verricht naar de vergelijkbaarheid van verschillende antilichamen die gebruikt kunnen worden voor PD-L1 IHC. Bij iedere initiële checkpoint inhibitor gericht tegen PD-1 of PD-L1, werd namelijk ook een bijbehorend antilichaam bedoeld voor PD-L1 IHC ontwikkeld. Deze antilichamen kunnen worden gebruikt in een gestandaardiseerd protocol ontwikkeld door de fabrikant, in dat geval ook wel een commerciële assay genoemd. Er zijn echter ook laboratoria die hun eigen protocol opzetten en dus gebruik maken van een antilichaam in een 'laboratory-developed test' (LDT). **Hoofdstuk 2** beschrijft een systematische review waarin we de beschikbare literatuur hebben onderzocht op interassay-, interobserver- en interlaboratoriumvariatie van PD-L1 IHC commerciële assays en LDTs. Hieruit bleek dat de overeenkomst in resultaten, ook wel concordantie genoemd, tussen de meeste commerciële assays gemiddeld tot hoog is, met uitzondering van één assay (SP142). Ook LDTs lieten bij goede optimalisatie en validatie in meerdere studies hoge concordantie zien met commerciële assays. Wanneer echter afkapwaarden werden toegepast voor het onderverdelen van TPS in PD-L1 positiviteit en PD-L1 negativiteit, bleek de concordantie in verschillende studies te dalen. Aangezien in de praktijk de afkapwaarden van 1% en 50% relevant zijn voor de therapiekeuze, zou discordantie tussen verschillende PD-L1 IHC assays bij het gebruik van deze afkapwaarden belangrijke klinische consequenties kunnen hebben. De verschillende PD-L1 commerciële assays en LDTs lijken dus niet simpelweg inwisselbaar te zijn in de klinische praktijk.

Post-analytisch: interobservervariatie

Daarnaast keken we in **hoofdstuk 2** ook naar interobservervariatie van de verschillende PD-L1 commerciële assays en LDTs, oftewel naar de overeenstemming tussen pathologen in het bepalen van de TPS voor iedere individuele assay of LDT. Over het algemeen was de overeenstemming hoog, echter bleek ook in dit geval het toepassen van afkapwaarden voor het bepalen van PD-L1 positiviteit problematisch. Met name een afkapwaarde van 1% leidde tot lagere concordantie tussen pathologen vergeleken met een afkapwaarde van 50%. Blijkbaar wordt het bepalen of een score positief of negatief is op basis van een afkapwaarde van 1% als lastiger ervaren dan wanneer dit gebeurt op basis van een afkapwaarde van 50%.

Interlaboratoriumvariatie in PD-L1 positiviteit

Op basis van onze literatuurstudie uit **hoofdstuk 2** bleek verder dat interlaboratoriumvariatie

voor de individuele commerciële assays laag is, maar dat de variatie groter is wanneer LDTs met commerciële assays worden vergeleken. Het betrof hier echter maar een beperkt aantal studies en geen van deze studies onderzocht variatie in PD-L1 expressie tussen laboratoria in een 'real-world' setting. In **hoofdstuk 5** beschrijven wij een studie waarin wij wel onderzoek hebben gedaan naar interlaboratoriumvariatie in PD-L1 positiviteit in een real-world klinische setting en op nationaal niveau, waarbij we de pathologielaboratoria in Nederland met elkaar hebben vergeleken. Hiertoe hebben we gebruik gemaakt van data van PALGA, het pathologisch-anatomisch landelijk geautomatiseerd archief, dat sinds 1991 alle pathologieverslagen in Nederland beheert. Data werden verzameld van alle patiënten in Nederland met NSCLC en bij wie PD-L1 in het pathologieverslag werd genoemd gedurende de periode juli 2017 t/m december 2018. Op basis van deze data werd voor ieder laboratorium dat PD-L1 bepalingen verrichtte de gemiddelde PD-L1 positiviteit berekend, gecorrigeerd voor verschillende patiënt- en weefselkarakteristieken (case-mix). De PD-L1 positiviteit werd apart berekend voor histologisch materiaal (verkregen door het verrichten van biopten en resecties) en cytologisch materiaal (verkregen door het verrichten van bronchusspoeling/brush, pleurapunctie, of dunne naald aspiratie van primaire tumor, lymfklier- of afstandsmetastase) en bepaald aan de hand van de twee klinisch relevante afkapwaarden, te weten 1% en 50%. We ontdekten dat de variatie in gemiddelde PD-L1 positiviteit tussen laboratoria substantieel was. Bij het gebruik van histologisch materiaal voor het testen voor PD-L1 expressie, was de variatie voornamelijk groot bij gebruik van 1% als afkapwaarde voor PD-L1 positiviteit. Vermoedelijk heeft dit te maken met de grotere interobservervariatie bij gebruik van 1% als afkapwaarde ten opzichte van 50%, zoals we in **hoofdstuk 2** beschreven. Mogelijk zou ook het gebruik van verschillende PD-L1 IHC commerciële assays en LDTs door de verschillende laboratoria onderdeel kunnen zijn van de verklaring, hoewel de data in onze studie helaas niet toereikend waren om hier een definitief antwoord op te geven.

Pre-analytisch: invloed van fixatief

Opvallend genoeg zagen we in **hoofdstuk 5** dat de variatie in PD-L1 positiviteit tussen laboratoria bij gebruik van cytologisch materiaal niet alleen substantieel was bij gebruik van 1% als afkapwaarde, maar ook bij gebruik van 50% als afkapwaarde. Dit doet vermoeden dat er mogelijk ook specifieke factoren gerelateerd aan cytologisch materiaal zijn die een rol spelen bij het ontstaan van interlaboratoriumvariatie in PD-L1 positiviteit. In **hoofdstuk 3** onderzochten we de concordantie in PD-L1 immunokleuringen tussen ingeblokt cytologisch materiaal en corresponderend histologisch materiaal, beide afgenomen van dezelfde longtumor. We zagen dat de overkoepelende concordantie gemiddeld tot substantieel was. Wanneer we het cytologisch materiaal echter uitsplitsten in twee groepen op basis van de verwerkingsmethode van het materiaal, zagen we dat de concordantie met histologie hoger was voor materiaal dat was ingeblokt in agar dan voor materiaal dat was verwerkt volgens een geautomatiseerd systeem, Cellient genaamd. We

vermoedden dat dit wellicht te maken zou kunnen hebben met de manier waarop het materiaal was gefixeerd. Al het cytologisch materiaal dat was ingeblokt in agar, was in formaline gefixeerd, wat ook wordt gebruikt voor het fixeren van histologisch materiaal. Het materiaal dat werd verwerkt volgens de Cellient-methode, was echter gefixeerd in CytoLyt en PreservCyt, twee media op methanol-basis. Middels cellijnen die PD-L1 tot expressie brachten en die met verschillende methoden waren gefixeerd, toonden we aan dat fixatieven met methanol of ethanol zorgden voor een lagere intensiteit van PD-L1 immunokleuringen vergeleken met formalinefixatie. Ook in **hoofdstuk 4** beschrijven we dat het gebruik van een ethanolfixatief in materiaal verkregen middels naaldaspiratie leidde tot fout-negatieve PD-L1 resultaten vergeleken met formalinefixatie, bij gebruik van zowel 1% als 50% als afkapwaarden. Met beide studies toonden we dus aan dat het gebruik van alcohol-gebaseerde fixatieven negatieve effecten kan hebben op de resultaten van PD-L1 immunokleuringen. Mogelijk zou het negatieve effect (deels) teniet kunnen worden gedaan door nafixatie in formaline, zoals we in **hoofdstuk 3** zagen in PD-L1 cellijnen die waren gefixeerd in een alcohol-gebaseerd fixatief gevolgd door formaline. Of dit echter ook toepasbaar is in de klinische praktijk, moet nog verder worden onderzocht.

In **hoofdstuk 6** hebben we vervolgens onderzocht hoeveel verschillende manieren van fixeren en verwerken van cytologisch materiaal er worden gebruikt in de klinische praktijk in Nederland, door het versturen van vragenlijsten naar alle Nederlandse pathologielaboratoria. De variatie die we vonden tussen de laboratoria, was enorm: 19 verschillende combinaties van fixatieven en inblokmethoden binnen 28 laboratoria. Om te onderzoeken of deze variatie in cytologieverwerking ook invloed heeft op de interlaboratoriumvariatie in PD-L1 positiviteit, combineerden we de data vergregen uit de vragenlijsten met de data over PD-L1 bepalingen op cytologisch materiaal van NSCLC patiënten verkregen via PALGA. Wanneer we corrigeerden voor verschillen in fixatie- en inblokmethoden, resulteerde dit in een vermindering van de interlaboratoriumvariatie in PD-L1 positiviteit. Verschillen in verwerking van cytologisch materiaal tussen laboratoria lijken dus inderdaad een deel van de aanzienlijke variatie in PD-L1 positiviteit tussen laboratoria te kunnen verklaren, waarbij voornamelijk het gebruik van alcohol-gebaseerde fixatieven zonder enige vorm van formalinefixatie problematisch lijkt te zijn.

Conclusie

In dit proefschrift brachten we de variatie in het testen voor PD-L1 expressie bij NSCLC patiënten binnen Nederlandse pathologielaboratoria in kaart, en onderzochten we een aantal factoren die van invloed kunnen zijn op de resultaten van PD-L1 immunokleuringen. We toonden aan dat er in de klinische praktijk aanzienlijke variatie bestaat in gemiddelde PD-L1 positiviteit tussen pathologielaboratoria, zowel bij gebruik van histologisch materiaal als van cytologisch materiaal voor het uitvoeren van de PD-L1 bepaling. Zeer waarschijnlijk speelt interobservervariatie hierin een belangrijke rol, zeker wanneer 1% als afkapwaarde

wordt gebruikt voor het bepalen van PD-L1 positiviteit. Bij het gebruik van cytologisch materiaal voor het beoordelen van de PD-L1 expressie, spelen verschillen in verwerking van het materiaal en met name in type fixatief een belangrijke rol in het verklaren van de grote interlaboratoriumvariatie in PD-L1 positiviteit. Daarnaast is het waarschijnlijk dat ook verschillen in gebruik van PD-L1 IHC assays en LDTs invloed hebben op de interlaboratoriumvariatie in PD-L1 positiviteit, hoewel de daadwerkelijke omvang van deze invloed in dit proefschrift niet goed kon worden onderzocht.

De mate van variatie in PD-L1 positiviteit die we met ons onderzoek aantoonden, is problematisch voor patiënten met NSCLC. PD-L1 expressie beoordeeld middels immunokleuring speelt namelijk een belangrijke rol in het bepalen van de therapie voor deze patiënten, en verschillen in uitkomst tussen laboratoria kunnen derhalve leiden tot verschillen in behandeling. Het is belangrijk dat clinici betrokken bij de behandeling van patiënten met NSCLC, maar ook bij patiënten met andere kankersoorten waarbij het bepalen van PD-L1 expressie relevant is, op de hoogte zijn van de variatie die tussen laboratoria bestaat, zodat gezamenlijk kan worden nagedacht hoe dergelijke variatie verminderd kan worden. We hopen dat we daar met dit proefschrift een bijdrage aan hebben kunnen leveren.

ACKNOWLEDGEMENTS / DANKWOORD

Het is zover: na heel wat jaren is mijn proefschrift nu toch écht af! Het voltooien van dit werk was niet mogelijk geweest zonder de hulp en steun van een heleboel mensen, die ik daarvoor graag enorm wil bedanken.

Geachte prof. dr. Willems, beste Stefan. Toen wij elkaar eind 2017 voor het eerst spraken over een eventueel onderzoeksproject, dacht ik nog niet dat ik aan een promotietraject zou gaan beginnen. Jouw enthousiasme en je denken in mogelijkheden werkten echter aanstekelijk, en een aantal maanden verder besloot ik toch om voor een PhD-traject te gaan onder jouw begeleiding. Ik ben heel dankbaar dat je me die kans hebt gegeven en kijk met veel plezier terug op onze samenwerking. Ik heb erg veel van je geleerd, niet alleen op inhoudelijk vlak, maar ook van je positieve instelling, je relativeringsvermogen en je overtuiging dat dingen goed komen, ook als ze wellicht anders lopen dan gepland. Veel dank daarvoor!

Geachte prof. dr. van Dijk, beste Marijke. Wat ben ik blij dat ik laat in mijn zesde jaar geneeskunde besloot om een coschap pathologie te gaan doen in plaats van het reeds geplande coschap huisartsgeneeskunde! Ik werd aangestoken door jouw enorme enthousiasme en vond eindelijk het vak dat ik wilde gaan uitoefenen. Jij was ook degene die Stefan en mij in contact bracht voor het onderzoeksproject dat hij had liggen, van waaruit mijn PhD-traject volgde. Ik ben je enorm dankbaar voor alle steun die je me biedt, niet alleen bij het afronden van mijn proefschrift binnen mijn opleiding, maar ook bij het zoeken naar mijn eigen weg in mijn carrière. Jouw betrokkenheid en je vertrouwen zijn van grote waarde voor me, en ik prijs mezelf gelukkig met jou als promotor én opleider.

Geachte leden van de leescommissie, prof. dr. Paul van Diest, prof. dr. Carla van Gils, prof. dr. Egbert Smit, prof. dr. Karijn Suijkerbuijk, en dr. Jan von der Thüsen, hartelijk dank voor het zitting nemen in de leescommissie en voor het beoordelen van mijn proefschrift.

Anne van Lindert en Aryan Vink, bedankt voor het vormen van mijn AIO-begeleidingscommissie en voor het meedenken over de voortgang van mijn PhD-traject. Anne, bedankt voor jouw deskundige en waardevolle feedback vanuit jouw klinische blik. Aryan, bedankt voor al jouw kennis en deskundigheid op het gebied van longpathologie en voor je enthousiasme en aanmoediging. Ik had heel graag gewild dat je de afronding van mijn PhD-traject had kunnen meemaken.

Veel dank ook aan een aantal andere pathologen die van grote waarde zijn geweest binnen mijn PhD-traject. Nils ’t Hart, bedankt voor de prettige samenwerking die resulteerde in

hoofdstuk 3 van dit proefschrift. Dank je voor het vertrouwen om met jouw mooie dataset aan de slag te mogen gaan en voor het kijkje in de keuken van de pathologie in het Isala in Zwolle. Willem Vreuls, ook jij bedankt voor de fijne samenwerking die resulteerde in hoofdstuk 4 van dit proefschrift. Ik heb genoten van onze gezellige en leerzame PD-L1 kijksessies, waarbij je me tussendoor ook een heleboel leerde over andere diagnostiek. Ik was altijd weer extra enthousiast over de pathologie als ik bij jou vandaan kwam! Mirthe de Boer, bedankt dat je me wegwijs hebt gemaakt in de wereld van cytologie. Ik vond het heel leerzaam en gezellig om met jou samen te werken en ben je enorm dankbaar voor al het werk dat je in ons PD-L1 project vanuit het expertisenetwerk cytologie hebt gestoken. Dank ook voor de vele fijne gesprekken die wij hebben gevoerd. Ik ben blij dat we tegenwoordig nog steeds af en toe kunnen samenwerken!

Ook wil ik Stichting PALGA bedanken, die met haar enorme schat aan klinische pathologiedata zo ontzettend waardevol is voor wetenschappelijk onderzoek. In het bijzonder dank aan Rinus Voorham, Ivette Deckers en Chantal Epskamp-Kuijpers, voor jullie hulp bij het verzamelen, analyseren en interpreteren van de data voor dit proefschrift.

Beste pathologen, AIOS en andere medewerkers van de afdeling pathologie in het UMC Utrecht, bedankt voor de prettige samenwerking en de fijne tijd die ik bij jullie als AIOS mag beleven. De combinatie van het afronden van mijn proefschrift naast het in opleiding zijn vond ik niet altijd makkelijk, maar werd een stuk draaglijker dankzij jullie steun, interesse en begrip. Daar ben ik jullie zeer dankbaar voor!

Lieve collega-onderzoekers van de PRL, wat heb ik genoten van mijn tijd samen met jullie in de overvolle, maar o zo gezellige kamer op de 3e verdieping, waar hard werken en inhoudelijke discussies moeiteloos werden afgewisseld met koffiepauzes, voetbalpoules en borrels. Ik kijk met heel veel plezier terug op de vele leuke dingen die we samen hebben gedaan, waaronder natuurlijk onze skivakantie in Flachau en de conferentie in Nice. We hebben samen veel gelachen, maar er was ook altijd wel een luisterend oor voor wie daar behoefte aan had. Ik ben jullie zeer dankbaar voor de toptijd die ik als PhD'er heb beleefd! Carmen, wat ben ik blij dat jij mij als paranimf wil bijstaan bij het afronden van mijn PhD-traject! Heel erg bedankt voor al je hulp op inhoudelijk vlak, zoals bij het maken van funnel plots of het uitvoeren van case-mix correcties, maar zeker ook voor alle gezellige, fijne en soms meer serieuze gesprekken die we voerden over allerlei levenszaken. Je bent een geweldige collega en met jou aan mijn zijde weet ik zeker dat ik me door die verdediging heen zal slaan! Emma, R-wonder, dank je voor je altijd scherpe wetenschappelijke blik en voor het meedenken over allerhande statistische uitdagingen, maar zeker ook voor alle gezelligheid! Ik heb erg genoten van onze gezamenlijke trip naar Wenen en vind het super tof dat we inmiddels samen in opleiding zijn tot patholoog. Lilian, dank je wel voor de vele fijne gesprekken die we hebben gevoerd en voor het kunnen delen van onze

liefde voor schattige dieren. Ik heb veel bewondering voor hoe jij in het leven staat en wens je alle goeds en geluk in de nieuwe uitdaging die jij bent aangegaan! Quirine, lieve PRL-buurvrouw, bedankt voor al je gezelligheid! Wat was het heerlijk om naast jou te zitten en ter afwisseling van het soms toch behoorlijk eentonige werk samen wedstrijden van de Olympische Spelen, geitenfilmpjes of Beleef de Lente-filmpjes te kijken. Folkert, pater familias van de PRL, altijd goedlachs en altijd in voor een praatje. Dank je voor je vrolijkheid, je gezelligheid en voor je interesse in mijn muzikale activiteiten naast het werk. Aernoud, niet altijd op de PRL te vinden, maar als je er was altijd in voor gezelligheid met koffie, lunch of bier. Bedankt dat je zelfs deze niet-voetbalfan enthousiast wist te krijgen voor de voetbalpoule! Wenzel, PRL-buurman na het vertrek van Quirine, wat kun jij ontzettend hard werken! Respect voor hoe je dat in die volle AIO-kamer voor elkaar kreeg en hoe je ook nog eens altijd bij alle borrels en andere activiteiten aanwezig was. Natalie, mijn niet-skimaatje tijdens de skivakantie, bedankt voor al jouw gezelligheid, je luisterend oor en al je hulp, ook nu nog tijdens de diagnostiek. Sebastiaan, PRL'er op de 4e verdieping, fijn dat ik met jou ervaringen kon uitwisselen over het doen van een PhD naast je opleiding. Dank je voor de koffiemomentjes en voor je geduld als ik weer eens loop te klagen! Gwen, ook wij werkten vooral samen op de 4e verdieping, maar je hebt daar veel meegekregen van mijn PhD-gerelateerde ups-and-downs. Dank je voor je steun, je gezelligheid en de vele fijne gesprekken die wij niet alleen over onze PhD's, maar ook over een heleboel andere zaken konden voeren! Liling, thank you for coming to one of my concerts together with your husband. Shuang, thank you for all the nice chats we had, even after I had left the 3rd floor to work as a resident on the 4th floor. And thanks to both of you for introducing us to proper Chinese food, it was quite the adventure! En dan zijn er nog een heleboel andere collega's die hebben bijgedragen aan de fijne tijd die ik als PRL-onderzoeker heb gehad: Anna Vera, Betzabel, Cathy, Floris, Jan, Niels, Rachel, Roel, Sangeeta, Susana, Willem, Wisse, enorm bedankt allemaal!

Ook buiten mijn werk weet ik mij gesteund door een heleboel mensen, die er allemaal op een eigen manier aan hebben bijgedragen dat ik dit proefschrift heb kunnen afronden. Een aantal van hen wil ik hier graag in het bijzonder noemen.

Lieve studievrienden, Maaike, Sabrin, Herke, Ruben en Taco-Jan, samen met jullie ben ik in 2009 het geneeskunde-avontuur gestart. Ik ben enorm dankbaar dat ik die jaren met jullie heb mogen doormaken en voor de vriendschap die daaruit is ontstaan. Ik vind het mooi en bijzonder om te zien hoe we vanuit die gezamenlijke studie allemaal een eigen weg in zijn gegaan. Ik heb veel bewondering voor ieder van jullie en ben benieuwd welke paden jullie in de toekomst nog allemaal gaan bewandelen!

Zonder muziek zou het leven maar saai en kleurloos zijn. Veel dank daarom aan alle geweldige musici met wie ik met regelmaat muziek mag maken! Een speciaal woord

van dank aan de liefste tweede violen van het USKO. Lieve meiden, door de jaren heen hebben we samen heel veel mooie muziek gemaakt en veel lief en leed gedeeld. Bedankt voor de enorm waardevolle vriendschap die we hebben opgebouwd en voor de steun die jullie me geven!

Lieve Uden-girls, wat ben ik blij met jullie in mijn leven en wat hebben we samen al veel meegemaakt! Dank jullie wel dat jullie er altijd zijn om alles mee te delen, zowel de mooie als de mindere momenten. Lieve Margot, dank je dat jij als paranimf aan mijn zijde wil staan tijdens de laatste “hindernis” van dit traject. Hoe bijzonder dat wij ooit samen ons profielwerkstuk schreven en dat we dadelijk allebei op dezelfde plek zijn gepromoveerd! Ik voel me vereerd dat jij me daarbij tot steun wil zijn.

Lieve Dorien en Rien, de liefste schoonouders die ik me maar zou kunnen wensen. Bedankt voor al jullie interesse in mij en mijn werk, voor de fijne gesprekken en etentjes, voor het vele gelach, en voor de ontspanning tijdens gezellige activiteiten, zoals het maken van pasta, brood of adventskransen.

Lieve Teuntje, mijn grote zus, dank je wel voor al je steun, je luisterend oor, je betrokkenheid ondanks de afstand en alle heerlijke gekte die vaak alleen wij begrijpen. Ik ben blij dat ik zoveel met jou kan delen! Ik wens jou samen met Arjan en Ella alle geluk en liefde toe. Lieve Arjan, ook jij bedankt voor al jouw betrokkenheid vanaf de andere kant van de wereld. Mega lief dat jij de cover voor mijn proefschrift hebt willen ontwerpen, dankjewel!

Lieve pap, ik had heel graag gewild dat je hier was geweest om dit mee te maken. Ik denk je erbij en hoop dat je trots op me bent!

Lieve mam en Steven, bedankt voor jullie niet-aflatende steun, liefde, enthousiasme, interesse en betrokkenheid bij alles wat ik doe ik mijn leven. Jullie staan altijd voor me klaar en daar ben ik jullie enorm dankbaar voor. Ik vind het heel fijn dat ik nu ook deze mijlpaal met jullie kan vieren!

Lieve Niels, mijn grootste steun, wat ben ik dankbaar en blij dat ik jou in mijn leven heb. Alles kan ik met je delen, zowel vreugde als verdriet. Jij viert met mij de mooie dingen (ik zal de champagne-traditie van ons beider PhD’s gaan missen!) en vangt me op als het even minder gaat. Dankjewel voor al je liefde, je blijdschap, je vertrouwen in mijn kunnen, je enthousiasme, en al het andere waarmee je mijn leven zoveel mooier maakt!

CURRICULUM VITAE

Bregje Maria Koomen was born on November 10, 1990, in Veghel, the Netherlands. She grew up in Uden, where she graduated cum laude from the gymnasium department of the Udens College in 2008. After this, she spent the first six months of 2009 in Oxford (United Kingdom), improving her British-English language skills and working in a traditional English pub. In September 2009 she started medical school at Utrecht University. During her studies, Bregje participated in various extracurricular activities, including being a fulltime board member of the Utrecht Student Choir and Orchestra (USKO; 2013-2014) and playing the violin in the Dutch National Student Orchestra (NSO; 2015) and the Dutch National Student Chamber Orchestra (NESKO; 2016). She obtained her Master's degree in Medicine in November 2017 (cum laude), after which she worked at the office for Medical Education of the University Medical Center Utrecht. In February 2018, she started her PhD research at the department of Pathology of the University Medical Center Utrecht under the supervision of prof. dr. Stefan Willems and prof. dr. Marijke van Dijk, resulting in this thesis. During her PhD-track, she attended congresses in Utrecht (the Netherlands), Vienna (Austria) and Nice (France), where she presented parts of this thesis either with posters or with oral presentations. From December 2019, Bregje has been working as a pathology resident at the department of Pathology of the University Medical Center Utrecht.

LIST OF PUBLICATIONS

THESIS

Koomen BM, Badrising SK, van den Heuvel MM, Willems SM. Comparability of PD-L1 immunohistochemistry assays for non-small-cell lung cancer: a systematic review. *Histopathology.* 2020;76(6):793-802.

Koomen BM, van der Starre-Gaal J, Vonk JM, von der Thüsen JH, van der Meij JJC, Monkhorst K, Willems SM, Timens W, 't Hart NA. Formalin fixation for optimal concordance of programmed death-ligand 1 immunostaining between cytologic and histologic specimens from patients with non–small cell lung cancer. *Cancer Cytopathology.* 2021;129(4):304-317.

Koomen BM, Vreuls W, de Boer M, de Ruiter EJ, Hoelters J, Vink A, Willems SM. False-negative programmed death-ligand 1 immunostaining in ethanol-fixed endobronchial ultrasound-guided transbronchial needle aspiration specimens of non-small-cell lung cancer patients. *Histopathology.* 2021;79(4):480-490.

Koomen BM, Voorham QJM, Epskamp-Kuijpers CCHJ, van Dooijeweert C, van Lindert ASR, Deckers IAG, Willems SM. Considerable interlaboratory variation in PD-L1 positivity in a nationwide cohort of non-small cell lung cancer patients. *Lung Cancer.* 2021;159:117-126.

Koomen BM, de Boer M, van Dooijeweert C, van Lindert ASR, Deckers IAG, Voorham QJM, Willems SM. Nationwide differences in cytology fixation and processing methods and their impact on interlaboratory variation in PD-L1 positivity. *Virchows Archiv.* 2023;482(4):707-720.

OTHER

Mostert M, **Koomen BM**, van Delden JJM, Bredenoord AL. Privacy in Big Data psychiatric and behavioural research: a multiple-case study. *International Journal of Law and Psychiatry.* 2018;60:40-44.

De Ruiter EJ, Mulder FJ, **Koomen BM**, Speel EJ, van den Hout MFCM, de Roest RH, Bloemena E, Devriese LA, Willems SM. Comparison of three PD-L1 immunohistochemical assays in head and neck squamous cell carcinoma (HNSCC). *Modern Pathology.* 2021;34(6):1125-1132.

Van Velthuysen MLF, van Eeden S, le Cessie S, de Boer M, van Boven H, **Koomen BM**, Roozekrans F, Bart J, Timens W, Voorham QJM. Impact of COVID-19 pandemic on diagnostic pathology in the Netherlands.
BMC Health Services Research. 2022;22(1):166.

Hempenius MA, **Koomen BM**, Deckers IAG, Oosting SF, Willems SM, van der Vegt B. Considerable interlaboratory variation in PD-L1 positivity for head and neck squamous cell carcinoma in the Netherlands – a nationwide evaluation study.
Manuscript submitted.