SUPPLEMENTARY MATERIAL - METHODS USED TO PRODUCE THE GUIDELINE

Group Composition

The Guideline Panel included all current members of the The Canadian Association of Pathologists – Association canadienne des pathologistes (CAP-ACP) National Standards Committee for High Complexity Testing (CAP-ACP NSCHCT). Additionally, the Committee invited national and international experts in the field as external consultants. A Steering Committee was formed in order to develop the scope of the Guidelines as well as key questions.

Conflict of Interest (COI) Policy

All members of the Guideline Panel declared potential COI for the period 01/2013 - Present including following categories:

- board membership or consultancy
- employment
- expert testimony
- grants/grants pending
- payments for lectures with educational/scientific content
- payment of speakers' bureau
- payment for manuscript preparation
- patents (planned, pending, issued)
- royalties
- stock/stock options
- other (travel/accommodations/meeting expenses not related to any of the above)
- other (err on the side of full disclosure)

They also needed to reply separately whether there are other relationships of activities that readers could perceive to have influenced, or that give the appearance of potentially influencing, work on the CAP-ACP PD-L1 guidelines.
Declared potential COI is presented in Appendix A.

Systematic Evidence Review (SER)

The objective of the SER was to develop an evidence-based guideline to help pathologists and clinical immunohistochemistry laboratories in Canada choose fit-for-purpose predictive PD-L1 assay/biomarker when required for any Health Canada-approved immunotherapy, harmonize reporting of the results of predictive PD-L1 assay/biomarker, and endorse national standards for PD-L1 quality assurance and quality control.

Key Questions

Key questions were developed by the Steering Committee with the methodologist prior to beginning the literature searches. All key questions were stated in PICO format as well as modified for adaptability/ease of use to clinical IHC laboratory perspective.

1. Which PD-L1 assay(s) should be used to predict potential response to anti-PD-1/PD-L1 immunotherapies?

P: patients with cancer for which immunotherapy exists
I: PD-L1 IHC predictive assay
C: comparisons of different PD-L1 IHC predictive assays (both CDx and LDTs)
O: same diagnostic accuracy (assays identify the same population of patients with specific cancer type as current reference standard for specific purpose)
T: type of study - "interchangeability studies of PD-L1 assays"

PICO format questions:
"Which PD-L1 IHC assays are clinically "interchangeable"? or
"Which PD-L1 assay(s) should be used/selected/performed by IHC laboratory for patients with cancer for which immunotherapy exists and it requires PD-L1 as predictive biomarker?"

Explanatory Notes:
After systematic review of published evidence, there was no high-quality evidence that could guide the recommendations and Key Question 2 was added to the scope of the project. Furthermore, additional data was requested from authors of interchangeability papers to enable assessment of diagnostic accuracy and is published as a meta-analysis (see manuscript reference 90).

2. What is the quality of statistical methodologies employed to evaluate PD-L1 assay performance in interchangeability assessments?

P: published papers on interchangeability of predictive PD-L1 assays
I: statistical methodology/analysis used in published papers on interchangeability of PD-L1 assays
C: comparison of statistical methodology
O: published evidence of high-quality based on statistical methodology
T: type of study - "interchangeability studies of PD-L1 assays"

PICO format question:
"Are published conclusions on interchangeability of PD-L1 assays supported by appropriate statistical methods, i.e. can be graded as "high quality evidence" if all other criteria are fulfilled?"

Explanatory Notes:
No specific guideline statements were issued regarding this key question. The evidence gathered by this systematic review led to a meta-analysis of additional collected results (see manuscript reference 90).

3. Were specific diagnostic assays (IHC protocol conditions and specific readout) used and stated by clinical trials where a specific drug and a specific disease were evaluated?

P: clinical trials for immunotherapy including various types of cancer
I: selection of patients eligible for immunotherapy
C: comparison of predictive PD-L1 biomarker selection for respective clinical trial(s)
O: clinical outcome of patients identified by specific PD-L1 assay (protocol and readout)

PICO format question:
"Is there any clinical trial showing that it is acceptable to select the patients for immunotherapy using any or different PD-L1 IHC protocol(s) and readout(s) for specific diagnostic indication/disease and specific drug?"

Explanatory note:
This question was selected to address whether there is necessity to reinforce recommendations based on "fit-for-purpose" principle in selection of predictive assays, with exploring whether it is possible to disconnect the 3D axis (Disease, Drug, Diagnostic assay).

4. How should the results of predictive PD-L1 assays be reported?

P: reporting of predictive biomarker results
I: use of systematic reporting of pathology diagnosis or biomarker results
C: systematic/harmonized reporting compared to non-systematic, free-text reporting
O: improved patient safety, and "customer/oncologist satisfaction"

PICO format question:
"Will the use of systematic reporting of PD-L1 predictive IHC assays improve patient safety and oncologist satisfaction?"

5. What measures/practices are necessary to ensure the quality of PD-L1 testing for patient selection in immunotherapy?

P: quality assurance for immunohistochemistry
I: measures/parameters of quality assurance
C: quality assurance measures for test development and maintenance
O: selection of patients for immunotherapy expected to have outcomes similar/same as in clinical trials

PICO format question:
"Which quality assurance measures are required to be implemented in clinical IHC laboratories to ensure that patients selected for immunotherapy by the PD-L1 IHC assay developed and performed in the clinical IHC laboratory will results in patient selection closely comparable to that in the given clinical trial?"

Literature Review

Both systematic and targeted review of literature was conducted as a part of a national project for developing Canadian guidelines for PD-L1 testing. The Canadian Association of Pathologists – Association canadienne des pathologistes (CAP-ACP) National Standards Committee for High Complexity Testing (CAP-ACP NSCHCT) initiated development of CAP-ACP Guidelines for PD-L1 testing in order to facilitate introduction of PD-L1 testing for various purposes to Canadian clinical IHC laboratories. The systematic review was performed for key questions 1, 2, and 3. The CAP-ACP NSCHCT also conducted targeted literature review for key questions 4 and 5.

Search and Selection

A search for literature was performed in MEDLINE using the PubMed interface. Last search was performed on August 31st, 2018.

- **Search for Assay Selection (Key Questions 1,2, and 3):** Search strategy using keyword "PD-L1" only was performed for the period of 01/2015 to 08/2016 in order to exclude the possibility of unintentional exclusion of articles based on mismatch of any more specific search terms. Search limits included: "human", and "English". This revealed 2,515 articles, which were downloaded to Zotero reference manager, for which abstracts were reviewed to exclude review papers, case reports, editorials, letters to editor, and any other low level of evidence publication. 106 publications were selected for full text review if they either included the results of clinical trials where a PD-L1 assay was employed as a potential predictive biomarker for immunotherapy or where comparison of performance of different PD-L1 predictive biomarker assays was evaluated. Clinical trial publications, publications on assay development for clinical trials, and FDA and other regulatory agency approval(s) were the source of evidence for selection of "designated reference/gold standard" for various clinical applications of the predictive PD-L1 assays.

- **Search for Assay Reporting (Key Question 4):** The overall strategy consisted of multiple targeted searches involving Pubmed and Google. The first targeted search was conducted in Pubmed for the period of 01/2000 to 08/2018 using keywords ("synoptic" OR "template" OR "structured") AND ("reporting") AND ("cancer" OR "biomarkers") AND ("pathology"); search limits included "human" and "English". This revealed 180 articles, which were downloaded to Zotero reference manager. The second targeted search was conducted in Pubmed for the period of 01/2000 to 08/2018 using keywords ("PD-L1") AND ("image analysis" OR "computer assisted"); search limits included "human" and "English". This revealed 14 articles, which were downloaded to Zotero reference manager. The third targeted search was conducted in Pubmed for the period of 01/2000 to 08/2018 using keywords ("PD-L1") AND ("reproducibility" OR "variability"); search limits included "human" and "English". This revealed 55 articles, which were downloaded to Zotero reference manager. Of all of the targeted Pubmed searches, 88 publications were selected for full text review. Additionally, a search in Google was performed in order to identify recommended reporting elements as described in the interpretation guides from the manufacturers of fit-for-purpose, commercially available, regulatory body-approved, PD-L1 assays retrieving 10 instruction guides.

- **Search for Assay Quality Assurance (Key Question 5):** Search strategy using keywords ("immunohistochemistry" AND (("quality control" OR "quality assurance") AND ("laboratories" OR "laboratory")) was performed for the period of 01/2000 to 08/2018. Search limits included: "human", and "English". This revealed 423 articles, which were downloaded to Zotero reference manager, for which abstracts were reviewed to exclude case reports, editorials, letters to editor, and articles that were deemed as being not relevant to the subject matter. With this approach 280 were excluded outright. In total, 123 publications were selected for full text review if they were international/national standards/guidelines/recommendations, international/national consensus opinion review articles, peer-reviewed publications presenting primary data, published reports

from recognized EQA providers, regulatory agency guidance documents, peer-reviewed published conference reports, and other publications that were deemed relevant to the subject matter. Of these 123, total of 26 were cited as relevant in the manuscript. A targeted search of Google did not reveal any additional contributing non-duplicate publications.

Review Process
All reviewers received Instructions for review. The instructions detailed methodology and criteria for grading published evidence (See Appendix B for full text of Instructions for Reviewers).

Data Extraction & Management of Evidence Tables
A bibliographic database was established in Zotero in order to select and track all publications. Two expert panel members reviewed all titles and abstracts identified by the initial search strategy and selected articles for full review using eligibility criteria as defined above (See "Search and Selection"). Data extraction was performed by expert reviewers who submitted the reviews through specially designed questionnaire on Survey Monkey. Reviewers had to answer twenty-nine questions for each publication that related to Key Question 1; eleven questions for each publication related to Key Question 2; twenty-one questions for each publication that related to Key Question 3 (see Appendix C, D, E, F, and G for full list of questions). All data extractions were audited by a methodologist.

Assessment of Quality of Evidence

- **Assay Selection:** Expert reviewers extracted data and assessed quality of evidence by using specially designed Survey Monkey questionnaire that followed published guidelines for the assessment of quality of evidence. Detailed instructions were provided to reviewers in order to employ the same criteria between different reviewers and different publications (See Appendix B).

- **Reporting:** Not applicable.

- **Assay quality assurance:** Given the nature of the subject matter, the Steering Committee developed a grading scheme to assess the quality of evidence for assay quality assurance that was based on how reviewers would classify each source publication/document. The following grading scheme was employed:
    - International/national standards/guidelines/recommendations [High]
    - International/national consensus opinion review article [Moderate]
    - Peer-reviewed review articles [Low]
    - Peer-reviewed publications presenting primary data [High]
    - Peer-reviewed published reports from recognized EQA providers [High]
    - Self-published reports from recognized EQA providers [Moderate]
    - Regulatory agency guidance documents [High]
    - Peer-reviewed published conference reports [Low]

Results from Assessment of Quality of Evidence
Quality of the evidence for assay selection was documented in Evidence Tables.

Drafting of Guideline Statements

The guideline statements were drafted following review of evidence tables and steering committee discussions.

Based on the Key Questions, 38 PD-L1 Guideline Statements were drafted to include three different sets of recommendations as follows:

- 15 recommendations for assay and sample selection
- 7 recommendations for harmonized reporting
- 16 recommendations for quality assurance (assay introduction/development and monitoring)

The draft guideline statements were disseminated to the main Expert Panel group for review and comment prior to a face-to-face consensus meeting, which was held on September 5th, 2018 in Toronto, Canada.

Assessing the Strength of Recommendations

At the face-to-face consensus meeting held on September 5th, 2018 in Toronto, Canada, the Expert Panel group reviewed and discussed terminology (see "Terminology" below) then separated into 3 breakout groups. Each breakout group was assigned one of the three main areas for assessment: i) assay and sample selection, ii) reporting, iii) quality assurance. Each breakout group discussed the pre-drafted guideline statements (with submitted comments) for their respective sections. The entire group re-convened after the breakout sessions and each breakout group presented their review of the guideline statements and supporting evidence to the group for discussion and final consensus.

Strength of recommendations was initially designated by consensus at the face-to-face meeting of the Expert Panel using a modified GRADE and QUADAS-2 approach; the Expert Panel reached consensus for guideline statements where evidence was lacking or only low-grade evidence was available. Instructions for grading the strength of recommendations was disseminated to the members of the Expert Panel ahead of the face-to-face meeting in Toronto on September 5th, 2018 (see Appendix H). Final agreement was obtained before submission for publication and after the public open review/comment period upon consideration of input received from CAP-ACP members and various societies.

Drafting of Manuscript

All drafts of the manuscript generated by the Steering Committee were reviewed by all co-authors. The final draft submitted for publication was approved by all authors.

Peer Review

A public open comment period was held from April 15th, 2019 to April 30th, 2019. All 38 recommendations were posted on the Canadian Association of Pathologists-Association canadienne des pathologistes (CAP-ACP) web site www.cap-acp.org. An invitation for public review and feedback was disseminated to all members of the CAP-ACP as well as professional societies with potential interest in this subject. These include:

- Canadian Partnership Against Cancer (CPAC)
- Canadian Chairs of Pathology and Laboratory Medicine (CCPLM)
- Canadian Society for Medical Laboratory Science (CSMLS)

- College of American Pathologists (CAP)
- American Society for Clinical Pathology (ASCP)
- United States and Canadian Academy of Pathology (USCAP)
- European Society for Medical Oncology (ESMO)
- European Society for Pathology (ESP)
- International Network for Quality in Pathology (IQN Path)
- International Society for Immunohistochemistry and Molecular Morphology (ISIMM)

The CAP-ACP website received 85 comments in total. As a result of the comments received, clarifications were added to 15 of the guideline statements (or their accompanying explanatory notes) and 5 definitions from the Terminology section.

The final draft was reviewed and approved by the CAP-ACP Executive Committee on June 21st, 2019.

Dissemination Plan

The Dissemination Plan for this work includes:

- Publication in a peer-reviewed journal;
- Direct dissemination to all members of the CAP-ACP;
- Posting of manuscript and supplementary files on the CAP-ACP resource web page;
- Presentation at various society meetings.

Appendix A

Conflict of Interest for PD-L1 Guidelines

Your NAME (Last, First)	Institution (Name, City, Province/State, Country)	I declare potential conflict of interest (enter "NONE" for negative reply or company name(s) for positive reply) for the period of 2013 - Present: Are there other relationships or activities that readers could perceive to have influenced, or that give the appearance of potentially influencing, work on the CAP-ACP PD-L1 guidelines?													
		Board membership or consultancy	Employment	Expert testimony	Grants/grants pending	Payments for lectures with educational/ scientific content	Payments for speakers bureau	Payment for manuscript preparation	Patents (planned, pending, issued)	Royalties	Stock/stock options	Other (travel/ accommodations/meeting expenses not related to any of the above)	Other (err on the side of full disclosure)	No other relationships/conditions/ circumstances that present a potential conflict of interest	Yes, the following relationships/ conditions/circumstances are present (explain below):
Barnes, Penny	Dalhousie University and Nova Scotia Health Authority, Halifax, Nova Scotia, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Bigras, Gilbert	Cross Cancer Institute, Edmonton, Alberta, Canada	Merck, Pfizer, Roche, Bayer, Bristol-Myers Squibb	none	Institut national d'excellence en santé et en services sociaux (INESSS) Quebec, Canada	Merck, Roche	none	Merck, Roche	none	none	none	Merck, Roche	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Boerner, Scott	University Health Network, Toronto, Ontario, Canada	none	none	none	none	Merck	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Butany Jagdish	University Health Network, Toronto, ON, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Calabrese, Fiorella	University Of Padova, Padova, Italy, Veneto Region	none	none	none	MERCK grants pending	Boehringer; Pzizer	none	none	none	none	none	Biotest, Transmedics	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Cheung, Carol	University of Toronto, Toronto, Ontario, Canada	Agilent, Astra-Zeneca, Bristol-Myers-Squibb, Cell Marque/Sigma, Merck, Roche	none	none	Roche	Agilent, Astra-Zeneca, Bristol-Myers-Squibb.	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Couture, Christian	Institut universitaire de cardiologie et de pneumologie de Québec - Université Laval (IUCPQ-UL), Quebec City, QC, Canada	Merck, Pfizer, Astra Zeneca, Boehringer-Ingelheim, Bristol-Meyer-Squibb, AbbVie, Roche, Qiagen	NONE	NONE	Merck	Merck, Pfizer, Novartis	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
El-Zimaity, Hala	Dynacare, Toronto/ON, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Fischer, Gabor															
Fiset, Pierre Olivier	McGill University, Montreal, Quebec, Canada	Consultancy for AstraZeneca Canada, Pfizer Canada and Merck Canada	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Garratt, John	clQc, Vancouver, British Columbia, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Geldenhuys, Laurette	Dalhousie University, Halifax, Nova Scotia, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Gilks, C Blake	University of British Columbia, Vancouver BC Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Ilie, Marius	Université Côte d'Azur, Nice, France	none	none	none	Boehringer Ingelheim	AstraZeneca, Roche, BMS	none	none	none	none	none	AstraZeneca, Roche, BMS, Merck	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Ionescu, Diana N	BC Cancer, Vancouver, BC, Canada	ASTRA-ZENECA, PFIZER, ROCHE, ELI-LILLY, NOVARTIS, BMS, MERCK	none	none	ASTRA-ZENECA	ASTRA-ZENECA, PFIZER, ROCHE, ELI-LILLY, NOVARTIS, BMS, MERCK	none	none	none	none	none	Pfizer	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Jean Deschenes	University of Alberta, Edmonton, Alberta, Canada	PFIZER, ROCHE, JANSSEN, MERCK,	NONE	NONE	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Lim, Hyun J.	University of Saskatchewan, Saskatoon, SK, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Lisa Manning	Shared Health, Winnipeg, Manitoba, Canada	Digital Pathology Association Board of Directors	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Mansoor, Adnan	University of Calgary, Calgary, Alberta Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Riddell,Robert	Mt.Sinai Hospital, Toronto, ON, Canada	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Ross,Catherine	McMaster University Hamilton Ontario Canada	none	none	none	none	Alexion,Celgene	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Roy-Chowdhuri, Sinchita	none	none	none	none	none	none	none	none	none	none	none	none	none		Committee memberships on CAP, ASC, PPS and other professional societies that could potentially be perceived as a position of influence, but no financial conflicts of interest
Spatz, Alan	McGill University Health Center and Lady Davis Institute, Montreal, QC, Canada	Merck, Roche, Janssen, Pfizer, BMS, Bayer, AstraZeneca	none	none	Janssen, Pfizer	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Swanson, Paul	Cumming School of Medicine, University of Calgary, Calgary Alberta CANADA	none	none	none	none	Pacific Northwest Society of Surgical Pathologists (spring 2018)	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Torlakovic, Emina	University of Saskatchewan and Saskatchewan Health Authority	Roche, Merck, Pfizer, BMS, Janssen, AstraZeneca	none	none	Roche, Merck, Pfizer, BMS, Janssen, AstraZeneca	Roche, Merck, Pfizer, BMS, Janssen, AstraZeneca	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Tron, Victor	St Mikes, Toronto	none	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Tsao, Ming	University Health Network, Toronto, Ontario, Canada	Merck, AstraZeneca, BMS, Ventana/Roche	none	none	Merck	Merck, AstraZeneca	none	none	none	none	none	none	none		IASLC Pathology Committee member
Wang, Hangjun	Divisions of Pathology and Molecular genetics, McGill University Health Center and McGill University, Montreal, QC, Canada	Astrazeneca, Merck, Pfizer	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	
Xu, Zhaolin	QEII Health Sciences Centre and Dalhousie University, Halifax, Nova Scotia, Canada	Merck Canada National Lung Cancer Medical Advisory Board	none	none	none	none	none	none	none	none	none	none	none	No other relationships/conditions/ circumstances that present a potential conflict of interest	

Appendix B

CAP-ACP Guidelines for PD-L1 [NSCHCT]

INSTRUCTIONS FOR REVIEWERS

GENERAL INSTRUCTIONS

Thank you for agreeing to review published evidence for the CAP-ACP PD-L1 Guidelines project. The Reviewer's Package consists of pdf files of the papers sent to you to be reviewed, Instructions for Reviewers, and pdf file of the survey (entitled "QUESTIONS") with questions that you will answer as you conduct your review of the published paper(s)/abstracts sent to you.

Tips for efficient review of the papers:

1. Before you access the actual web-page with questions, please review the pdf file with the copy of the questions (entitled "QUESTIONS") so that you are familiar what to specifically look for in the paper(s).
2. Review all five (5) pages of these instructions before you start answering the questions on the web-page (see below for the link).
3. Printing the instructions and the tables for grading evidence on page 4 and 5 may be handy when you start reviewing papers.
4. If you need clarification regarding the instructions for grading evidence that are included below, please contact study leader/principal investigator (PI).
5. You will be asked for the password in order to access the survey; the password is provided on this page (see below).
6. You will not be able to make any changes to your submitted review after submission; in case you need to make changes to already submitted review, you will need to resubmit the whole review again. When you resubmit, please label the resubmission as such (answer Question 1 with the paper number and add text that will make clear that this is the corrected version). Alternatively, if corrections are minor, please contact study leader/PI. It is possible to us to make corrections after the submission, but only after all responses are collected and exported to an Excel file.

SUBMIT YOUR REVIEW AT:
https://www.surveymonkey.com/ [deleted]

DECLARE POTENTIAL CONFILICT OF INTEREST (COI) AT:
https://www.surveymonkey.com/ [deleted]

PASSWORD TO ACCESS THE SURVEY: [deleted]

INSTRUCTIONS FOR GRADING EVIDENCE

Published evidence will be evaluated for its validity, reliability, consistency, and overall risk of bias. Panel members will apply the modified GRADE scheme to grade the strength of evidence by using modified pre-specified GRADE criteria related to study design, methodology, and risk of bias.
The summary rating will be used an indication of the Panel's confidence in the available evidence.
Every study is designated a score of 10 at the start. Scores are deducted as per Table 1 (see below).
A final grade of strength of evidence is designated as per Table 2 (see below). Table 1A and 2A are prepared for grading evidence of interchangeability of the IHC assays, Table 1B and 2B for grading so-called "3D evidence" (evidence is support of selection of a specific Diagnostic test for specific Disease and specific Drug), and Table 1C and 2C for grading evidence for QA/QC recommendations.

1. **Risk of Bias/study limitations.** The study design and execution should be assessed, and if the study is not well-designed and executed, the evidence can be downgraded by one or two levels depending on how serious the problems are. Examples relevant to studies in which two tests are compared:
    a. The IHC biomarkers always have some purpose. If the specific purpose for developing and using the IHC biomarker is not stated in the comparison study, the study may not be valid. The specific purpose could be that the markers will be used as diagnostic markers (e.g. DOG1 and CD34 in GIST), prognostic markers (c-myc and Bcl-2 protein expression in DLBCL), or predictive markers for a specific therapy and clinical setting (e.g. PD-L1 IHC for pembroluzimab first line therapy for NSCLC). There is no "general" use of predictive biomarkers. This includes PD-L1; therefore, "testing for PD-L1 as a predictive IHC biomarker" does not exist if it is not specified for which drug and which clinical setting. The purpose of any predictive biomarker is defined using the so-called "3D approach": the **D**iagnostic test is used for a defined **D**isease, for a specific **D**rug/therapy.
    b. Did the study include and properly designate a relevant reference standard test for specific use ("fit-for-purpose" reference standard)?
    c. Studies comparing an index test (test in question or new test) to a reference standard should have the operators blinded to the results of the other test.
    d. Valid studies of diagnostic test accuracy should include representative and consecutive patients.
    e. Did the study use reasonable "acceptance criteria" for diagnostic accuracy (e.g ≥ 90% agreement with positive results [diagnostic sensitivity] and ≥ 95% or at least 90% agreement for negative results [diagnostic specificity])? If diagnostic accuracy was not assessed, were acceptance criteria reasonable for the other test performance characteristics that were evaluated in the study?

2. **Inconsistency of results.** Inconsistency refers to unexplained heterogeneity of results for the same test comparisons in different published papers. When reviewing the literature, explanations for heterogeneity should be considered. If a plausible explanation cannot be identified, the quality of evidence should be downgraded. Whether it is downgraded by one or two levels will depend on the magnitude of the inconsistency in the results. Examples of factors

that may explain inconsistency in results could be differences in population or differences in study methods.

3. **Indirectness of evidence.** Direct evidence consists of research that directly uses the test of interest, in the population of interest, and measures the outcomes important to patients. For example, studies of diagnostic accuracy are indirect evidence regarding that test's relationship to a patient outcome, because it must be inferred that greater diagnostic accuracy will result in improved patient outcomes. Direct evidence would be provided by a study design that measured the patient outcomes related to the use of the test, rather than only measuring test accuracy.
    a. Only clinical response represents direct evidence of biomarker validity for specific use.
    b. If the study evaluated how well their accuracy compares with the reference standard (a specific gold standard test developed in the clinical trial or recognized otherwise as equivalent of such), then diagnostic sensitivity and diagnostic specificity need to be determined against the reference standard for each specific use of the biomarker. Please note that it is not possible to evaluate diagnostic accuracy if a proper reference standard test for a specific drug and specific clinical setting is not designated or used in the study. For example, Intra-Class Correlation (ICC) Coefficient may indicate that the tests are similar but does not reflect diagnostic accuracy of the assessed test, which is the most basic and most relevant parameter of diagnostic test performance. Correlations and other statistical tools used to demonstrate test similarity are only indirect evidence about the potential of the test to achieve the same diagnostic accuracy as the reference standard. **IMPORTANT**: many studies used "agreement with positive" and "agreement with negative" or "concordance for positive results" and "concordance for negative results" - if the study used any designated "gold standard" (e.g. evaluating if pharmDx 22C3 can replace pharmDx 28-8 for specific purpose [but not in general], the pharmDx 28-8 would be a gold standard in that study), the above terms would be synonymous to "dx sensitivity" and "dx specificity". However, if pharmDx 28-8 was evaluated for if it could replace pharmDx 22C3, now pharmDx 22C3 becomes the gold standard and diagnostic accuracy of 28-8 is calculated against 22C3 results. If none are designated as "gold standard" or "reference test", diagnostic accuracy is not assessed, but rather the similarity of the tests, which are not relevant for patient stratification for targeted therapy. However, look for "hidden" information/results/data, which may be available in the study, but is not specifically stated.
    c. Another type of indirectness can arise from indirect comparisons of two or more alternative tests. Tests should ideally be compared within one study, using the same set of patients.

4. **Imprecision.** In general, results are imprecise when studies include relatively few patients and few events and thus have a wide confidence interval around the estimate of the effect. In this case, one may judge the quality of the evidence lower than it otherwise would be considered because of resulting uncertainty about the results. Any study that includes less than 20 positive

and 20 negative cases as identified by the designated reference standard using a specific cutoff point identified in the clinical trial, may be unpowered. Some studies may require larger number of cases/patients depending on the questions asked. Any conclusions of such studies are highly uncertain. In addition, the total number of samples could be reasonable, but various different types of samples are included (different tissue types, different tumors, different pre-analytical conditions and tissue processing, etc.). When this is the case and it is known that the difference in type may affect IHC results (e.g. alcohol-fixed cytology samples were grouped with FFPE histology samples), the study may still be under-powered for each type of the sample, which should be evaluated separately.

Table 1A. Quality of Evidence Scoring (see above for detailed explanations)

	Problem	Score
Bias/study limitations		
a. Purpose of IHC test[1]	Not defined or identified	-2
b. Reference standard for accuracy	Not defined or identified	-2
c. Readout	Operators were not blinded	-1
d. Sample selection	Samples are not fully representative of condition evaluated for the purpose of the test, and/or samples are not consecutive, but preselected for certain characteristics (e.g. enriched for positive or negative results)	-1 for low risk of bias -2 for high risk of bias
e. Acceptance criteria	Acceptance criteria too low	-1
Inconsistency of results	Differences between published studies cannot be explained by different populations or different methods used (e.g. the results of this study are different than that previously published, but there is no explanation for the cause of the difference)	-1
Indirectness of evidence		
a. Clinical response	Not addressed/evaluated	-2
b. Diagnostic accuracy	Not addressed/evaluated	-2
c. Study population/samples	Study population or samples are different from one study to another (e.g. study you evaluated used different types of tumours or patients with significantly different characteristics than that for which the test was originally developed)	-2
Imprecision	Less than 20 positive and 20 negative cases are included for given cutoff relevant to the purpose of the test	-1

1 – Purpose is defined using so-called "3D approach" (**D**iagnostic test is used for defined **D**isease, for specific **D**rug/therapy)

Table 2A. Quality of Evidence for Predictive PD-L1 IHC Assay

Quality of Evidence	Level of Confidence	Final Score
High Quality	We are confident that the test makes an important contribution to the determination of outcome (predictive strength)	10
Moderate Quality	We are somewhat confident that the test makes an important contribution to the determination of the outcome. The estimate of the observed predictive strength or diagnostic accuracy is likely close to the true effect, but there is a possibility that it is substantially different.	6 - 9
Low Quality	We have little confidence in the predictive estimate of the test. The true predictive strength and/or diagnostic accuracy could be substantially different from the estimate of test validity.	<6

Table 1B. Quality of 3D Evidence Scoring (see above for detailed explanations)

	Problem	Score
Study limitations		
a. Role of IHC test	Not defined or identified	-2
b. PD-L1 IHC protocol	Same IHC protocol was used for all patients that were tested for PD-L1	-2
c. Readout	Readout was not stated	-3
d. Sample selection	Not all samples are representative of disease	-3
e. Acceptance criteria	Acceptance criteria for analytical performance did not follow assay specification	-1
Indirectness of evidence		
d. Clinical response	Not addressed/evaluated	-2
e. Clinical response was correlated with test results for specific readout parameters (e.g. cutoff points)*	Not addressed/evaluated	-2

* - Irrespective of the outcome

Table 2B. Quality of 3D Evidence for Predictive PD-L1 IHC Assay

Quality of Evidence	Level of Confidence	Final Score
High Quality	We are confident that the published paper demonstrated clinical evidence of an association (or absence of an association) between a specific diagnostic assay, specific drug, and specific disease population.	9-10
Moderate Quality	We are somewhat confident that the published paper demonstrated clinical evidence of an association (or absence of an association) between a specific diagnostic assay, specific drug, and specific disease population.	6 - 8
Low Quality	We have little confidence that the published paper demonstrated clinical evidence of an association (or absence of an association) between a specific diagnostic assay, specific drug, and specific disease population.	<6

Sources

1. Don-Wauchope, A.C. and P.L. Santaguida, Grading Evidence for Laboratory Test Studies Beyond Diagnostic Accuracy: Application to Prognostic Testing. EJIFCC, 2015. 26(3): p. 168-82.
2. Brozek, J.L., et al., Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy, 2009. 64(8): p. 1109-16.
3. Schunemann, H.J., et al., GRADE: assessing the quality of evidence for diagnostic recommendations. ACP J Club, 2008. 149(6): p.2.
4. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.

Appendix C

CAP-ACP PD-L1 Interchangeability Evidence Systematic Review

1. Paper INTERCHANGE number designation

* 2. First author: Last name, initials (e.g. Smith DA)

* 3. Month/year of publication (e.g. 08/2016)

* 4. Journal name

* 5. What type of affiliation do the authors have? Please note that this is not a question about Conflict of Interest (e.g. authors declaring membership in advisory boards or similar), but affiliation only.

- ◯ Industry affiliation (some or all authors are employees of company selling tests (e.g. Dako, Ventana/Roche) or selling drugs (Merck, BMS, etc.)
- ◯ Academic affiliation and other (but none of the authors have industry affiliation)
- ◯ Other ONLY (but none of the authors have industry affiliations)
- ◯ Academic affiliation ONLY
- ◯ Other (please specify)

* 6. Is this a report of interchangeability from EQA Proficiency Testing for the PD-L1 IHC Assay?

- ◯ No
- ◯ YES (Please specify EQA program name and country)

* 7. Which primary Ab clones were compared? (select all that apply)

- [ ] 22C3
- [ ] E1L3N
- [ ] 28-8
- [ ] CAL10
- [ ] SP142
- [ ] ZR3
- [ ] SP263
- [ ] Other (please specify)

* 8. 22C3 clone

- [ ] 22C3 was not used
- [ ] 22C3 concentrate from Dako, Agilent
- [ ] 22C3 was used as part of pharm Dx 22C3, Dako/Agilent
- [ ] 22C3 from other source (please specify)

* 9. 28-8 clone

- [ ] 28-8 was not used
- [ ] 28-8 from Abcam
- [ ] 28-8 was used as part of pharmDx 28-8 Dako/Agilent
- [ ] 28-8 from other source (please specify)

* 10. SP142

- [ ] SP142 was not used
- [ ] SP142 concentrate from Spring Biosciences
- [ ] SP142 was used as part of VENTANA PD-L1 (SP142) Assay
- [ ] SP142 from other source (please specify)

* 11. SP263

- [ ] SP263 was not used
- [ ] SP263 was used as part of VENTANA PD-L1 (SP263) Assay
- [ ] SP263 from other source (please specify)

* 12. E1L3N

- [ ] E1L3N was not used
- [ ] E1L3N from Cell Signaling (CST)
- [ ] E1L3N from other source (please specify)

* 13. What kind of samples were used in the study? (select all that apply)

- [ ] Whole tissue sections of FFPE tumours
- [ ] TMA with benign tissues only
- [ ] Whole tissue sections of FFPE benign tissues
- [ ] Whole sections of cytology cell blocks
- [ ] TMA with tumours only
- [ ] TMA of cytology cell blocks only
- [ ] TMA with tumours and benign tissues
- [ ] Cell lines
- [ ] Other (please specify)

* 14. This publication... (select all that apply)

- [ ] Makes a clear distinction between the PD-L1 IHC assay in general and the PD-L1 IHC predictive assay with a specific purpose (specific drug(s) and specific clinical setting). If assays are compared, they are compared to a reference standard for a specific purpose that is determined in the relevant clinical trial.
- [ ] Makes a clear distinction between the PD-L1 IHC assay and the PD-L1 antibody clone (as a single IHC reagent). The conclusions derived from the results obtained by IHC protocols are not "transferred" to the "performance of the primary Ab clone".
- [ ] Makes a clear distinction between clinical validation (qualification of the biomarker), diagnostic validation, and technical validation of the PD-L1 IHC Assay.
- [ ] Makes no clear distinction between "antibody" and "assay/test" or "antibody" and "IHC protocol"
- [ ] Makes no clear distinction between detection of PD-L1 molecule by IHC test and detection of PD-L1 for specific purpose (drug, disease, disease stage, etc.). The test(s) were not evaluated/compared for specific purpose. No gold standard for specific purpose was used in the study.
- [ ] Makes no clear distinction between different spheres of validation (clinical vs. diagnostic vs. technical)

* 15. How many samples of each sample type were included in the study?

FFPE tumours N =	
FFPE benign N =	
Cell block samples N =	
Cell lines N =	
Other (please specify sample type) N =	
Other (please specify sample type) N =	

* 16. Statistical analysis included (select all that apply)

- [ ] Pearson Correlation
- [ ] Spearman Correlation
- [ ] Cohen's kappa coefficient
- [ ] Intra-Class Correlation (ICC)
- [ ] Concordance Correlation Coefficient (CCC)
- [ ] Bland–Altman plot (Difference plot)
- [ ] Positive percentage agreement (Dx Sensitivity)
- [ ] Negative percentage agreement (Dx Specificity)
- [ ] Overall rate of agreement

Other (please specify)

* 17. Which test performance characteristics were evaluated? (select all that apply)

- [ ] Analytical Sensitivity
- [ ] Analytical Specificity
- [ ] Diagnostic Sensitivity
- [ ] Diagnostic Specificity
- [ ] Clinical Sensitivity (against clinical responses/outcomes)
- [ ] Clinical Specificity (against clinical responses/outcomes)
- [ ] Analytical Reproducibility
- [ ] Reportable Range
- [ ] Diagnostic Accuracy (both dx sensitivity and specificity)
- [ ] Readout Accuracy (for categorical assessment, with cutoff points)
- [ ] Readout Precision (for readout as continuous variable as in TPS)
- [ ] Readout Concordance
- [ ] Other (please specify)

* 18. Which IHC assay was used as a designated gold standard (reference test)?

- [ ] None; no IHC assay was designated as a "gold standard" or "reference" test
- [ ] pharmDx 22C3
- [ ] pharmDx 28-8
- [ ] Other (please specify)
- [ ] VENTANA PD-L1 (SP263)
- [ ] VENTANA PD-L1 (SP142)

* 19. Which type of readout was performed? (select all that apply)

- [ ] Continous (0 to100%) tumor percentage score (TPS)
- [ ] TPS categorical with cutoff points (e.g. 1%, 50% or other)
- [ ] Other (please specify)
- [ ] Categorical for immune cells
- [ ] Combined positive score

* 20. Which criteria were used to claim that the compared tests were "equal" or "interchangeable" (e.g. Intra-Class Correlation coefficient, diagnostic accuracy, etc.)?

1

2

3

4

5

* 21. State study conclusion(s)

Conclusion 1:

Conclusion 2:

Conclusion 3:

22. Risk of bias in patient selection (could the selection of patients/cases have introduced bias?):

- Low risk
- High risk
- Unclear risk

23. Risk of bias for index test (index test is any test that is being compared to reference test/other test); could the conduct or readout of the index test have introduced bias)?

- Low risk
- High risk
- Unclear risk

24. Risk of bias for reference test (reference test is any test that has already been validated for specific purpose, e.g. pharmDx 22C3); could the reference standard, its conduct, or its readout have introduced bias?

- Low risk
- High risk
- Unclear risk

25. Are there concerns that the included patients/cases do not match the study question and that the results may not be applicable to intended application(s)?

- ◯ Low risk
- ◯ High risk
- ◯ Unclear risk

26. Are there concerns that the index test, its conduct, or its readout differ from the main study question (e.g. the study purpose was to compare prim Ab clones, but suboptimal IHC protocols were used for index test)?

- ◯ Low risk
- ◯ High risk
- ◯ Unclear risk

27. Are there any concerns that the target condition as defined by the reference standard does not match the study question (e.g. the study is using the results of pharmDx 22C3, but does not assess dx accuracy for relevant cutoff points)?

- ◯ Low risk
- ◯ High risk
- ◯ Unclear risk

* 28. The overall grade for evidence (use modified GRADE criteria as per Instructions for Reviewers):

- ◯ High
- ◯ Low
- ◯ Moderate

Comment:

* 29. Your name

Appendix D

CAP-ACP PD-L1 Interchangeability STATISTICAL Evidence Systematic Review

1. Paper STATISTICS INTERCHANGE number designation

* 2. First author: Last name, initials (e.g. Smith DA)

* 3. Month/year of publication (e.g. 08/2016)

* 4. Journal name

5. In this comparison of different tests and their performance, was any test or other target (e.g. clinical outcome) designated as "gold standard"?

- ◯ Yes, one test was designated as "gold standard"
- ◯ Clinical outcome was designated as "gold standard"
- ◯ No, no test was designated as "gold standard", but recognized "gold standard" was included
- ◯ No, no "gold standard" was designated and no recognized "gold standard" was included
- ◯ Yes, but the performance of the tests was evaluated against multiple designated "gold standards"
- ◯ Other (please specify)

6. If "gold standard" was designated or recognized "gold standard" was included, what was it?

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

* 7. Statistical analysis included (select all that apply)

- [ ] Pearson Correlation
- [ ] Spearman Correlation
- [ ] Cohen's kappa coefficient
- [ ] Intra-Class Correlation (ICC)
- [ ] Concordance Correlation Coefficient (CCC)
- [ ] Bland–Altman plot (Difference plot)
- [ ] Positive percentage agreement (Dx Sensitivity)
- [ ] Negative percentage agreement (Dx Specificity)
- [ ] Overall rate of agreement
- [ ] Other

Other (please specify)

* 8. Which type of DATA was created by a readout? (select all that apply)

- [ ] Continous (0 to100%) tumor percentage score (TPS)
- [ ] TPS categorical with cutoff points (e.g. 1%, 50% or other)
- [ ] Other (please specify)
- [ ] Categorical for immune cells
- [ ] Combined positive score

9. Was selection of statistical methods appropriate for analysis of pathologist's readout (scoring)?

- ◯ Yes
- ◯ No
- ◯ Other (please specify)

10. Was the number of cases included in the study adequate considering statistical power?

- ◯ Yes
- ◯ No
- ◯ Other (please specify)
- ◯ It is not possible to determine based on available information in methods

* 11. Which criteria were used to claim that the compared tests were "equal" or "interchangeable" (e.g. Intra-Class Correlation coefficient, diagnostic accuracy, etc.)?

1
2
3
4
5

12. Which tests were validated in this study for which specific use (to be able to replace which gold standard) if applicable?

1.
2.
3.
4.
5.
6.
7.
8.

* 13. The overall grade for STATISTICAL evidence (use modified GRADE criteria as per Instructions for Reviewers):

- ◯ High
- ◯ Moderate
- ◯ Low

Comment:

* 14. Your name

Appendix E

CAP-ACP PD-L1 Cytology Evidence Systematic Review

1. Paper number designation (e.g. C1)

* 2. First author: Last name, initials (e.g. Smith DA)

* 3. Month/year of publication (e.g. 08/2016)

* 4. Journal name

* 5. What type of affiliation do the authors have? Please note that this is not a question about Conflict of Interest (e.g. authors declaring membership in advisory boards or similar), but affiliation only.

- ◯ Industry affiliation (some or all authors are employees of company selling tests (e.g. Dako, Ventana/Roche) or selling drugs (Merck, BMS, etc.)
- ◯ Academic affiliation ONLY
- ◯ Other (please specify)
- ◯ Academic affiliation and other (but none of the authors have industry affiliation)
- ◯ Other ONLY (but none of the authors have industry affiliations)

* 6. Is this a report of interchangeability from EQA Proficiency Testing for the PD-L1 IHC Assay?

- ◯ No
- ◯ YES (Please specify EQA program name and country)

* 7. Which primary Ab clones were evaluated? (select all that apply)

- [ ] 22C3
- [ ] E1L3N
- [ ] 28-8
- [ ] CAL10
- [ ] SP142
- [ ] ZR3
- [ ] SP263
- [ ] Other (please specify)

* 8. 22C3 clone

- [ ] 22C3 was not used
- [ ] 22C3 concentrate from Dako, Agilent
- [ ] 22C3 was used as part of pharm Dx 22C3, Dako/Agilent
- [ ] 22C3 from other source (please specify)

* 9. 28-8 clone

- [ ] 28-8 was not used
- [ ] 28-8 from Abcam
- [ ] 28-8 was used as part of pharmDx 28-8 Dako/Agilent
- [ ] 28-8 from other source (please specify)

* 10. SP142

- [ ] SP142 was not used
- [ ] SP142 concentrate from Spring Biosciences
- [ ] SP142 was used as part of VENTANA PD-L1 (SP142) Assay
- [ ] SP142 from other source (please specify)

* 11. SP263

- [ ] SP263 was not used
- [ ] SP263 was used as part of VENTANA PD-L1 (SP263) Assay
- [ ] SP263 from other source (please specify)

* 12. E1L3N

- [ ] E1L3N was not used
- [ ] E1L3N from Cell Signaling (CST)
- [ ] E1L3N from other source (please specify)

* 13. What kind of samples were used in the study? (select all that apply)

- [ ] Whole tissue sections of FFPE tumours
- [ ] Whole tissue sections of FFPE benign tissues
- [ ] TMA with tumours only
- [ ] TMA with tumours and benign tissues
- [ ] TMA with benign tissues only
- [ ] Whole sections of cytology cell blocks
- [ ] TMA of cytology cell blocks only
- [ ] Cell lines
- [ ] Cytology smears
- [ ] If cytology smears were used, what fixative was used and what time of fixation was allowed?

* 14. This publication... (select all that apply)

- [ ] Makes a clear distinction between the PD-L1 IHC assay in general and the PD-L1 IHC predictive assay with a specific purpose (specific drug(s) and specific clinical setting). If assays are compared, they are compared to a reference standard for a specific purpose that is determined in the relevant clinical trial.
- [ ] Makes a clear distinction between the PD-L1 IHC assay and the PD-L1 antibody clone (as a single IHC reagent). The conclusions derived from the results obtained by IHC protocols are not "transferred" to the "performance of the primary Ab clone".
- [ ] Makes a clear distinction between clinical validation (qualification of the biomarker), diagnostic validation, and technical validation of the PD-L1 IHC Assay.
- [ ] Makes no clear distinction between "antibody" and "assay/test" or "antibody" and "IHC protocol"
- [ ] Makes no clear distinction between detection of PD-L1 molecule by IHC test and detection of PD-L1 for specific purpose (drug, disease, disease stage, etc.). The test(s) were not evaluated/compared for specific purpose. No gold standard for specific purpose was used in the study.
- [ ] Makes no clear distinction between different spheres of validation (clinical vs. diagnostic vs. technical)

* 15. How many samples of each sample type were included in the study?

FFPE tumours N =	
FFPE benign N =	
Cell block samples N =	
Cell lines N =	
Other (please specify sample type) N =	
Other (please specify sample type) N =	

* 16. Statistical analysis included (select all that apply)

- [ ] Pearson Correlation
- [ ] Spearman Correlation
- [ ] Cohen's kappa coefficient
- [ ] Intra-Class Correlation (ICC)
- [ ] Concordance Correlation Coefficient (CCC)
- [ ] Bland–Altman plot (Difference plot)
- [ ] Positive percentage agreement (Dx Sensitivity)
- [ ] Negative percentage agreement (Dx Specificity)
- [ ] Overall rate of agreement
- [ ] Other

Other (please specify)

* 17. Which test performance characteristics were evaluated? (select all that apply)

- [ ] Analytical Sensitivity
- [ ] Analytical Specificity
- [ ] Diagnostic Sensitivity
- [ ] Diagnostic Specificity
- [ ] Clinical Sensitivity (against clinical responses/outcomes)
- [ ] Clinical Specificity (against clinical responses/outcomes)
- [ ] Analytical Reproducibility
- [ ] Reportable Range
- [ ] Diagnostic Accuracy (both dx sensitivity and specificity)
- [ ] Readout Accuracy (for categorical assessment, with cutoff points)
- [ ] Readout Precision (for readout as continuous variable as in TPS)
- [ ] Readout Concordance
- [ ] Other (please specify)

* 18. Which IHC assay was used as a designated gold standard (reference test)?

- [ ] None; no IHC assay was designated as a "gold standard" or "reference" test
- [ ] pharmDx 22C3
- [ ] pharmDx 28-8
- [ ] Other (please specify)
- [ ] VENTANA PD-L1 (SP263)
- [ ] VENTANA PD-L1 (SP142)

* 19. Which type of readout was performed? (select all that apply)

- [ ] Continous (0 to100%) tumor percentage score (TPS)
- [ ] TPS categorical with cutoff points (e.g. 1%, 50% or other)
- [ ] Other (please specify)
- [ ] Categorical for immune cells
- [ ] Combined positive score

* 20. Which criteria were used to claim that the compared tests were "equal" or "interchangeable" (e.g. Intra-Class Correlation coefficient, diagnostic accuracy, etc.)?

1

2

3

4

5

* 21. State study conclusion(s)

Conclusion 1:

Conclusion 2:

Conclusion 3:

22. Risk of bias in patient selection (could the selection of patients/cases have introduced bias?):

- Low risk
- High risk
- Unclear risk

23. Risk of bias for index test (index test is any test that is being compared to reference test/other test); could the conduct or readout of the index test have introduced bias)?

- Low risk
- High risk
- Unclear risk

24. Risk of bias for reference test (reference test is any test that has already been validated for specific purpose, e.g. pharmDx 22C3); could the reference standard, its conduct, or its readout have introduced bias?

- Low risk
- High risk
- Unclear risk

25. Are there concerns that the included patients/cases do not match the study question and that the results may not be applicable to intended application(s)?

- ◯ Low risk
- ◯ High risk
- ◯ Unclear risk

26. Are there concerns that the index test, its conduct, or its readout differ from the main study question (e.g. the study purpose was to compare prim Ab clones, but suboptimal IHC protocols were used for index test)?

- ◯ Low risk
- ◯ High risk
- ◯ Unclear risk

27. Are there any concerns that the target condition as defined by the reference standard does not match the study question (e.g. the study is using the results of pharmDx 22C3, but does not assess dx accuracy for relevant cutoff points)?

- ◯ Low risk
- ◯ High risk
- ◯ Unclear risk

* 28. The overall grade for evidence (use modified GRADE criteria as per Instructions for Reviewers):

- ◯ High
- ◯ Low
- ◯ Moderate

Comment:

* 29. Your name

Appendix F

CAP-ACP PD-L1 3D Evidence Systematic Review

1. Paper 3D number designation

2. Your name (Last, First)

* 3. First author: Last name, initials (e.g. Smith DA)

* 4. Month/year of publication (e.g. 08/2016)

* 5. Journal name

* 6. What type of affiliation do the authors have? Please note that this is not a question about Conflict of Interest (e.g. authors declaring membership in advisory boards or similar), but affiliation only.

- ◯ Industry affiliation (some or all authors are employees of company selling tests (e.g. Dako, Ventana/Roche) or selling drugs (Merck, BMS, etc.)
- ◯ Academic affiliation ONLY
- ◯ Other (please specify)
- ◯ Academic affiliation and other (but none of the authors have industry affiliation)
- ◯ Other ONLY (but none of the authors have industry affiliations)

7. PD-L1 status was determined by an IHC assay in this clinical trial.

- ◯ Yes
- ◯ No

* 8. Was PD-L1 testing performed for specific Drug(s)?

- ◯ Yes
- ◯ No

If "Yes", please specify

* 9. How many PD-1/PD-L1 inhibitors were being evaluated in this clinical trial (please state the number below)?

* 10. Was PD-L1 testing performed for specific Disease indications (tumor types specifically identified)?

- ◯ Yes
- ◯ No

If "Yes", please specify

* 11. Which primary Ab clones were employed? (select all that apply)

- ☐ 22C3
- ☐ 28-8
- ☐ SP142
- ☐ SP263
- ☐ E1L3N
- ☐ CAL10
- ☐ ZR3
- ☐ Other (please specify)

* 12. 22C3 clone

- ☐ 22C3 was not used
- ☐ 22C3 was used as part of pharm Dx 22C3, Dako/Agilent
- ☐ 22C3 concentrate (from any source) in an assay that was specifically designed for this trial
- ☐ Assay protocol with 22C3 Ab (from any source) was not specifically design for this protocol nor defined in the published trial results

* 13. 28-8 clone

- ☐ 28-8 was not used
- ☐ 28-8 was used as part of PD-L1 IHC pharmDx 28-8, Dako/Agilent
- ☐ 28-8 concentrate (from any source) in an assay that was specifically designed for this trial
- ☐ Assay protocol with 28-8 Ab (from any source) was not specifically design for this protocol nor defined in the published trial results

* 14. SP142 clone

- ☐ SP142 was not used
- ☐ SP142 was used as part of VENTANA PD-L1 (SP142) assay
- ☐ SP142 concentrate (from any source) in an assay that was specifically designed for this trial
- ☐ Assay protocol with SP142 Ab (from any source) was not specifically design for this protocol nor defined in the published trial results

* 15. SP263 clone

- [ ] SP263 was not used
- [ ] SP263 was used as part of VENTANA PD-L1 (SP263) assay
- [ ] SP263 concentrate (from any source) in an assay that was specifically designed for this trial
- [ ] Assay protocol with SP263 Ab (from any source) was not specifically design for this protocol nor defined in the published trial results

* 16. E1L3N clone

- [ ] E1L3N was not used
- [ ] E1L3N concentrate (from any source) in an assay that was specifically designed for this trial
- [ ] Assay protocol with E1L3N Ab (from any source) was not specifically design for this protocol nor defined in the published trial results

* 17. Any other anti-PD-L1 clone

- [ ] Other anti-PD-L1 Ab was not used
- [ ] Other anti-PD-L1 Ab concentrate (from any source) in an assay that was specifically designed for this trial
- [ ] State name of the other anti-PD-L1 Ab used in this trial
- [ ] Assay protocol with Other anti-PD-L1 Ab (from any source) was not specifically design for this protocol nor defined in the published trial results

* 18. Was the pathologist's readout (interpretation) specifically defined for the IHC assay (e.g. cell type and cut-off point)?

- ( ) Yes
- ( ) No

* 19. Was PD-L1 IHC assay repurposed from different clinical trial? (select all that apply)

- [ ] No
- [ ] Yes, the protocol was repurposed and the readout was unchanged
- [ ] Yes, the protocol was repurposed, but the readout was changed

* 20. State trial conclusion(s) regarding the clinical validity of the PD-L1 IHC biomarker (association with clinical outcomes such as PFS, OS, ORR), etc.

- [ ] PD-L1 IHC results were informative regarding clinical outcomes by using specific cut-off with specific drug used
- [ ] PD-L1 IHC results were informative regarding clinical outcomes irrespective of cut-off points used in the study with specific drug used
- [ ] PD-L1 IHC results were not informative regarding clinical outcomes with specific drug used

* 21. The overall grade for 3D Evidence (use modified GRADE criteria developed for this purpose as per Instructions for Reviewers):

- ◯ High
- ◯ Low
- ◯ Moderate

22. Other comments about the paper:

Appendix G

CAP-ACP PD-L1 Sample Selection Evidence Systematic Review

1. Paper Sample Selection number designation

2. Your name (Last, First)

* 3. First author: Last name, initials (e.g. Smith DA)

* 4. Month/year of publication (e.g. 08/2016)

* 5. Journal name

* 6. What type of affiliation do the authors have? Please note that this is not a question about Conflict of Interest (e.g. authors declaring membership in advisory boards or similar), but affiliation only.

- ◯ Industry affiliation (some or all authors are employees of company selling tests (e.g. Dako, Ventana/Roche) or selling drugs (Merck, BMS, etc.)
- ◯ Academic affiliation ONLY
- ◯ Other (please specify)
- ◯ Academic affiliation and other (but none of the authors have industry affiliation)
- ◯ Other ONLY (but none of the authors have industry affiliations)

7. In this study, what was the purpose for determining the PD-L1 status of the tumour?

- ☐ PD-L1 was used as predictive biomarker
- ☐ PD-L1 was used as prognostic biomarker
- ☐ PD-L1 was used as diagnostic biomarker
- ☐ Expression frequency/prevalence study
- ☐ Other (please specify)

* 8. What percentage of cases showed tissue heterogeneity for PD-L1 expression in this study?

- [ ] 0%
- [ ] <10%
- [ ] 10 - 50%
- [ ] > 50%
- [ ] If "Yes", please specify

* 9. Tissue heterogeneity was demonstrated between:

- [ ] Different areas in single biopsy/resection sample
- [ ] Different sections of the same tumour
- [ ] Different biopsies/cytology samples from the same patient where samples were obtained at the same time or within very short interval (< 1 month)
- [ ] Different biopsies/cytology samples from the same patient where samples were obtained at different time (>1 month)

* 10. In this study, tissue heterogeneity effects were linked to clinical outcomes.

- ◯ Yes
- ◯ No

If "Yes", please specify

11. Other comments about the paper:

Appendix H

Grading of Recommendations

Designation	Recommendation	Rationale
Strong recommendation	Recommend for or against a particular PD-L1 testing practice (and include must or should)	Strength of evidence is convincing based on consistent, generalizable, good quality evidence; further studies are unlikely to change the conclusions
Recommendation	Recommend for or against a particular PD-L1 testing practice (and include should or may)	Strength of evidence is adequate based on limitations in the quality of evidence; further studies may change the conclusions
Expert opinion	Recommend for or against a particular PD-L1 testing practice (and include should or may)	Important testing element to address but strength of evidence is inadequate; gaps in knowledge may require further studies

Adapted/modified from GRADE* and College of American Pathologists**

*Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schünemann HJ. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013 Jul;66(7):719-25.

**Fitzgibbons PL, Bradley LA, Fatheree LA, Alsabeh R, Fulton RS, Goldsmith JD, Haas TS, Karabakhtsian RG, Loykasek PA, Marolt MJ, Shen SS, Smith AT, Swanson PE. College of American Pathologists Pathology and Laboratory Quality Center. Principles of analytic validation of immunohistochemical assays: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med. 2014 Nov;138(11):1432-43.