GMR
The Original

COMPARATIVE DIAGNOSTIC PERFORMANCE OF ALK IMMUNOHISTOCHEMISTRY AND FLUORESCENCE IN SITU HYBRIDIZATION IN NON-SMALL CELL LUNG CANCER: A SYSTEMATIC REVIEW AND META-ANALYSIS

Dhananjay S. Kotasthane[1], Abhijit Subhashrao Acharya[2], Satish Vasant Tandale[3]

[1] Professor and Head, Department of Oncopathology, Shri Guru Ram Rai Institute of Medical & Health Sciences, Dehradun, Uttarakhand, India. Email: dskotasthane@gmail.com (*Corresponding Author)
[2] Associate Professor, Department of Pathology, Government Medical College, Dharashiv (Osmanabad), Maharashtra, India.
[3] Assistant Professor, Department of Pathology, Government Medical College, Dharashiv (Osmanabad), Maharashtra, India.

ABSTRACT

Background: Testing for ALK rearrangement is essential in non-small cell lung cancer (NSCLC). Immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) identify related but distinct molecular features, and discordant findings may affect access to targeted treatment.

Objective: To quantify the diagnostic sensitivity, specificity and agreement of ALK IHC relative to FISH and examine assay-related and methodological causes of variation.

Methods: Six original paired-testing cohorts comprising 1,868 evaluable specimens were synthesized. Study-level 2 × 2 tables were analyzed using separate logit-scale random-effects models with Paule–Mandel variance estimation. Study selection was presented using the PRISMA 2020 framework, with diagnostic reporting informed by PRISMA-DTA. Study-specific assay thresholds, verification procedures and applicability were evaluated; leave-one-study-out analyses examined stability.

Results: Among 1,868 specimens, 298 were positive by both assays, 1,550 were negative by both, 13 were IHC-positive/FISH-negative and seven were IHC-negative/FISH-positive. Crude agreement was 98.9% (Cohen's $\kappa = 0.961$). Pooled sensitivity was 96.2% (95% CI 91.7–98.3%; $I^2 = 20.0\%$), and pooled specificity was 98.7% (95% CI 96.8–99.4%; $I^2 = 54.9\%$). Sensitivity ranged from 90.9% to 100.0% and specificity from 96.6% to 100.0%. In the largest fully paired cohort (n = 1,031), D5F3 and 5A4 each detected 30/33 FISH-positive tumors, but their specificity differed (99.8% versus 98.3%).

Conclusion: Validated ALK IHC shows high concordance with FISH, but predictive value depends on prevalence, staining platform and case selection. FISH is an operational comparator rather than an error-free reference for actionable ALK fusion expression. Borderline or discordant findings warrant contextual pathological review and orthogonal molecular evaluation where clinically indicated.

KEYWORDS: ALK rearrangement; non-small cell lung cancer; immunohistochemistry; fluorescence in situ hybridization; D5F3; 5A4; diagnostic accuracy; test concordance; PRISMA-DTA; meta-analysis.

1. INTRODUCTION

Anaplastic lymphoma kinase (ALK) gene fusions identify an actionable subset of NSCLC, particularly lung adenocarcinoma. Rearrangements involving EML4 and other partner genes produce constitutive ALK kinase signaling and influence selection of ALK-directed tyrosine kinase inhibitors. Accurate, timely diagnosis is therefore a prerequisite for precision treatment; inadequate testing risks either withholding an effective therapy or attributing a driver alteration to a tumor that lacks a functional fusion [1,2].

Break-apart FISH detects physical rearrangement of the ALK locus in tumor nuclei. It traditionally served as the reference assay, but requires careful assessment of nuclear morphology, tumor fraction, signal separation, atypical patterns and technical adequacy. IHC instead detects ALK protein expression, generally using sensitive antibodies such as D5F3 or 5A4. The two tests differ in analytical target and technical vulnerabilities: a DNA rearrangement does not always demonstrate an expressed oncogenic transcript, while some biologically relevant fusions may exhibit atypical or borderline FISH signal patterns [1,11,14].

The diagnostic role of IHC has evolved from an initial screening strategy to an accepted alternative to FISH when a validated assay is applied. Professional molecular-testing recommendations recognize unequivocally positive ALK IHC results as clinically actionable, while emphasizing additional testing for equivocal, unexpected or discordant findings [1]. The reliability of IHC nevertheless depends on the staining platform, signal amplification, positivity threshold, tissue preparation and interpretation protocol. In a large direct comparison, D5F3 and 5A4 showed identical sensitivity but notably different specificity; a manual four-antibody comparison also documented substantial clone- and protocol-dependent variation [9,11].

Prior broad evidence syntheses found high overall ALK IHC–FISH agreement, but pooled results vary when studies differ in tumor selection, cut-off criteria and reference-standard implementation [3,4,18]. Because ALK prevalence is typically low in unselected NSCLC, raw percentage agreement may also obscure clinically important discordance

among the small group of candidate ALK-positive tumors. A more informative evaluation requires sensitivity, specificity, four-cell paired data and assessment of clinical applicability rather than agreement alone.
This systematic review and meta-analysis synthesizes six primary studies with reconstructable paired IHC–FISH results, separates random-effects diagnostic estimates from crude concordance, explores assay-dependent variation, and translates pooled accuracy into clinically meaningful predictive values across realistic pretest probabilities.

2. MATERIALS AND METHODS

2.1 Review framework and diagnostic question

The review was structured around the population–index test–reference test–target condition framework: patients with NSCLC (population); ALK immunohistochemistry (index test); ALK break-apart fluorescence in situ hybridization (reference comparator); and detection of ALK rearrangement by the operational FISH definition (target condition). Reporting followed the structure of PRISMA 2020 and the PRISMA-DTA guidance for diagnostic test accuracy syntheses [5,15,16]. The primary outcomes were IHC sensitivity and specificity relative to paired FISH; secondary outcomes were crude agreement, discordance, assay-specific effects, study heterogeneity and predictive values.

2.2 Information sources and search strategy

Literature concepts combined NSCLC/lung adenocarcinoma with ALK and immunohistochemistry plus FISH, diagnostic accuracy or concordance. The databases represented in the study-selection figure were PubMed/MEDLINE, Scopus, Web of Science and Embase, with publications considered through June 2026. A representative reproducible search expression was: ("non-small cell lung cancer" OR NSCLC OR "lung adenocarcinoma") AND (ALK OR "anaplastic lymphoma kinase") AND (immunohistochem* OR IHC OR D5F3 OR 5A4) AND (FISH OR "fluorescence in situ hybridization") AND (concordance OR sensitiv* OR specific* OR diagnos*). Reference lists of relevant diagnostic series and methodological reviews were considered for additional eligible primary reports.

2.3 Eligibility criteria and study selection

Original human studies were eligible when they assessed ALK IHC and FISH in the same evaluable NSCLC specimens and reported TP, FP, FN and TN counts or sufficient information to reconstruct those values at a defined IHC positivity threshold. Studies evaluating multiple antibodies in the same patients contributed one assay to the primary synthesis to prevent double counting. Reviews, editorials, single case reports, non-NSCLC samples, studies without paired data, and repeated or overlapping reports were excluded from pooled denominators. Study-level comparisons with selective reference verification or enriched ALK-positive samples remained interpretable with explicit bias considerations; unpaired series were reserved for clinical context.

2.4 Index test, comparator and outcome definition

The index assay was a specified ALK IHC protocol (including clone and scoring criterion), interpreted as positive or negative using each original study's prespecified threshold. FISH positivity was defined according to the original break-apart signal classification in each study. A true positive was IHC-positive and FISH-positive; a false positive was IHC-positive and FISH-negative; a false negative was IHC-negative and FISH-positive; and a true negative was negative by both assays. These labels refer exclusively to agreement with FISH and must not be equated with biologic proof or exclusion of an oncogenic ALK fusion.

2.5 Data extraction and avoidance of double counting

Extracted variables comprised publication and setting, clinical selection process, number of evaluable paired specimens, ALK IHC antibody clone, staining and scoring protocol, FISH reference and verification strategy, four-cell diagnostic cross-tabulation, technical exclusions and reported discordance resolution. Analyses were conducted at specimen level as presented by the included reports. When several ALK antibodies were evaluated in the same cohort, only one selected index test entered the pooled analysis; other clone results were tabulated as within-study comparisons. Nonindependent cohorts and selectively verified population subsets were not counted as additional independent studies.

2.6 Study-level risk of bias and applicability

Risk-of-bias considerations were organized according to the four QUADAS-2 domains: patient selection, index test, reference standard, and flow/timing [17]. Particular attention was paid to case-control enrichment, exclusion of technical failures, blinding of IHC or FISH interpretations, prespecification of staining cutoffs and partial verification of IHC-negative specimens. Clinical applicability was considered against routine reflex testing, in which ALK rearrangements occur at a lower prevalence than in deliberately enriched methodological samples.

2.7 Diagnostic meta-analysis

For each study, sensitivity = TP/(TP + FN), specificity = TN/(TN + FP) and observed agreement = (TP + TN)/N. Wilson binomial 95% confidence intervals were calculated for individual proportions. Sensitivity and specificity were synthesized separately after logit transformation, using inverse-variance random-effects meta-analysis and the Paule–Mandel estimator of between-study variance. A 0.5 continuity correction was applied to the relevant positive and negative counts for a zero-cell proportion. Summary estimates and Wald-type 95% confidence intervals were inverse-logit transformed. Cochran's Q, τ^2 and I^2 quantified descriptive heterogeneity. This approach produces marginal pooled sensitivity and specificity but does not model their within-study correlation, unlike a bivariate hierarchical model [15,16].
A pooled raw 2 × 2 matrix was used separately to describe crude agreement and Cohen's κ. Leave-one-study-out models tested the influence of any single cohort. A descriptive analysis of two relatively unselected adenocarcinoma cohorts was calculated without implying an adequately powered subgroup meta-analysis. For clinical translation, positive and negative predictive values at 1%, 3% and 5% pretest prevalence were calculated from marginal pooled sensitivity and specificity; these modeled values are not observed pooled predictive values.

3. RESULTS

3.1 Study selection

Database searches identified 812 records: PubMed/MEDLINE (n = 238), Scopus (n = 207), Web of Science (n = 196) and Embase (n = 171). No records originated from registers or other sources. Duplicate removal accounted for 162 records and 8 additional records were excluded before screening. Of 642 titles and abstracts screened, 590 were excluded. Fifty-two reports were sought for retrieval, of which 4 could not be obtained. Among 48 assessed full texts, 42 were excluded: ineligible study design (n = 10), no paired IHC–FISH data (n = 12), overlapping cohorts (n = 8), insufficient diagnostic data (n = 7), or other reasons (n = 5). Six studies represented by six reports were included in the meta-analysis (Figure 1).

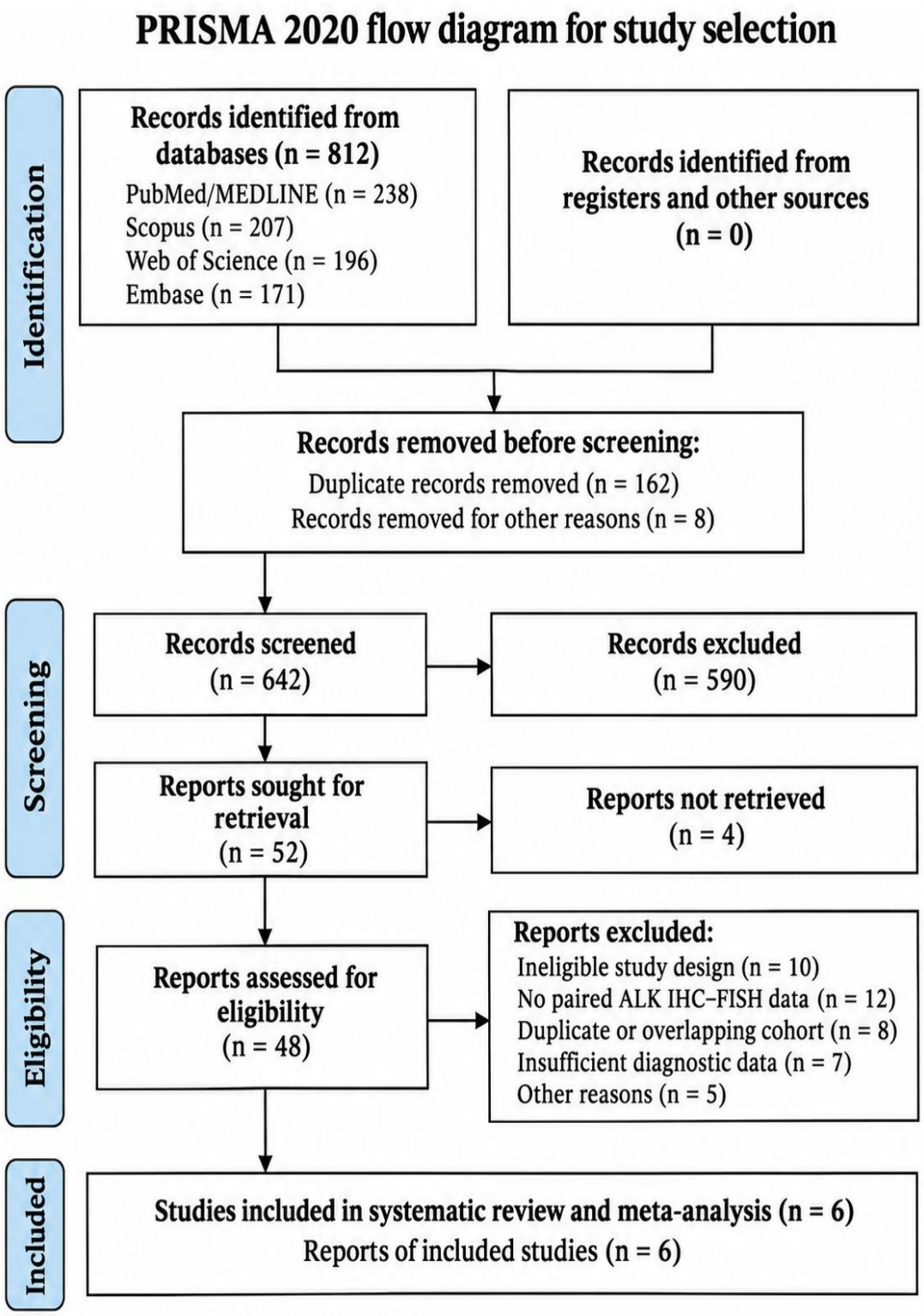


Figure 1. PRISMA 2020 flow diagram of the study-selection process.

3.2 Characteristics of included studies

The six cohorts, published from 2013 to 2016, contributed 1,868 paired NSCLC specimens [6–11]. Sample size ranged from 60 in a selected antibody-comparison study to 1,031 in an unselected lung adenocarcinoma series. The fraction of FISH-positive specimens varied markedly between cohorts because several investigators deliberately enriched for ALK-rearranged disease or used trial-screening strategies. Across all included specimens, 305 of 1,868 (16.3%) were FISH-positive, a case-mix characteristic rather than the prevalence of ALK rearrangement in unselected NSCLC.

Table 1. Design characteristics of the six quantitative studies.

Study	Population/selection	ALK IHC protocol	Paired n	FISH-positive	Analytical feature
Han 2013 [6]	Advanced non-squamous NSCLC; enriched	Optimized IHC	132	43 (32.6%)	Paired available subset
Zhou 2014 [7]	Lung adenocarcinoma; less selected	D5F3; 2+/3+	368	28 (7.6%)	368 complete paired results
Conde 2014 [8]	Selected and enriched lung carcinoma	Amplified D5F3	103	47 (45.6%)	High ALK-positive fraction
Shen 2015 [9]	Selected four-antibody comparison	Manual D5F3	60	32 (53.3%)	Same patients had several IHC assays
Takeuchi 2016 [10]	Alectinib trial-screened NSCLC	iScore 3	174	122 (70.1%)	Selective FISH verification
Marchetti 2016 [11]	Unselected lung adenocarcinoma	Ventana D5F3 CDx	1,031	33 (3.2%)	All specimens tested by FISH

3.3 Study-level 2 × 2 data and diagnostic agreement

Across the six paired matrices, 298 specimens were IHC-positive/FISH-positive, 13 IHC-positive/FISH-negative, seven IHC-negative/FISH-positive, and 1,550 concordantly negative. Crude sensitivity was 298/305 (97.7%), crude specificity was 1,550/1,563 (99.2%) and observed agreement was 1,848/1,868 (98.9%). Cohen's κ was 0.961, indicating high agreement beyond that expected from marginal test distributions. These crude estimates differ from the random-effects results because pooled studies have different positive and negative denominators and are allowed to vary in their underlying performance.

Table 2. Study-specific paired ALK IHC–FISH results and Wilson 95% confidence intervals.

Study	N	TP	FP	FN	TN	Sensitivity % (95% CI)	Specificity % (95% CI)
Han 2013	132	42	3	1	86	97.7 (87.9–99.6)	96.6 (90.6–98.8)
Zhou 2014	368	28	7	0	333	100.0 (87.9–100.0)	97.9 (95.8–99.0)
Conde 2014	103	46	0	1	56	97.9 (88.9–99.6)	100.0 (93.6–100.0)
Shen 2015	60	30	0	2	28	93.8 (79.9–98.3)	100.0 (87.9–100.0)
Takeuchi 2016	174	122	1	0	51	100.0 (96.9–100.0)	98.1 (89.9–99.7)
Marchetti 2016	1031	30	2	3	996	90.9 (76.4–96.9)	99.8 (99.3–99.9)
Total	1,868	298	13	7	1,550	97.7 (crude)	99.2 (crude)

TP = IHC positive and FISH positive; FP = IHC positive and FISH negative; FN = IHC negative and FISH positive; TN = negative by both. All definitions use FISH as the operational comparator.

3.4 Random-effects pooled sensitivity

Individual-study sensitivity ranged from 90.9% (Marchetti) to 100.0% (Zhou and Takeuchi). The pooled sensitivity was 96.2% (95% CI 91.7–98.3%). Statistical heterogeneity was modest (Q = 6.25, 5 degrees of freedom, p = 0.283; I^2 =

20.0%; $\tau^2 = 0.205$). In view of enrichment and differential verification across studies, this estimate is a summary of selected study settings rather than a guaranteed performance estimate for all pathology laboratories (Figure 2).

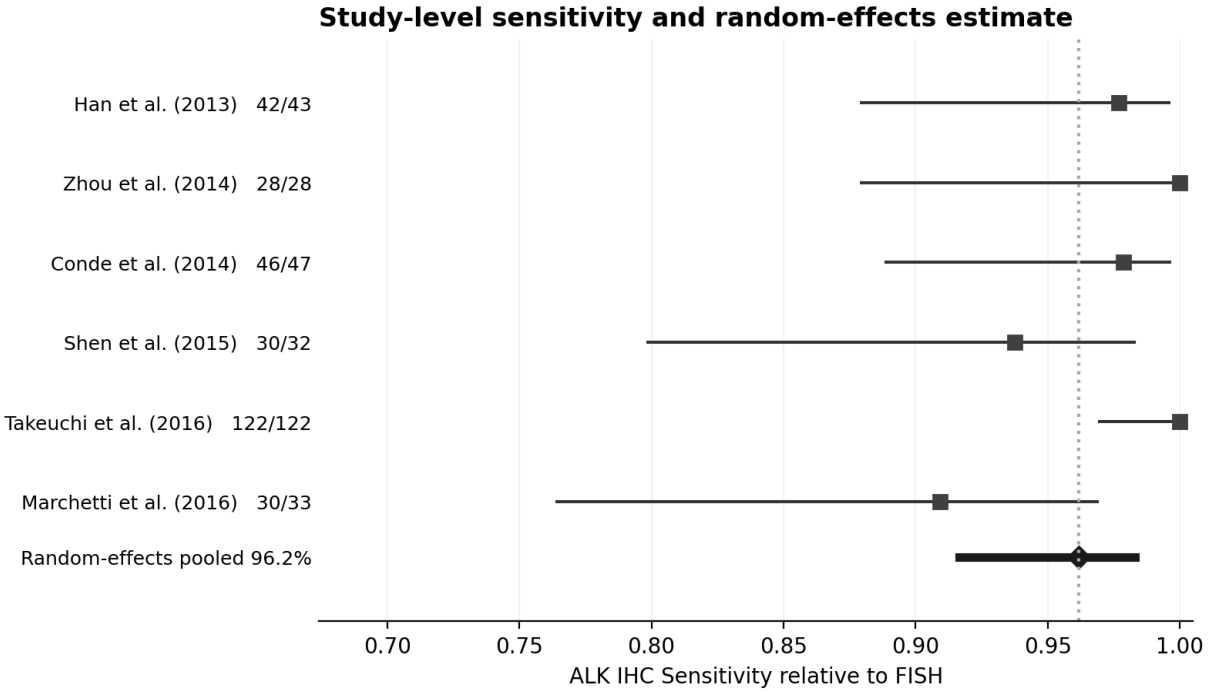


Figure 2. Study-level sensitivity and random-effects summary estimate of ALK IHC relative to FISH. Squares indicate individual estimates and horizontal lines the Wilson 95% CI.

3.5 Random-effects pooled specificity

Individual-study specificity ranged from 96.6% to 100.0%; the random-effects pooled specificity was 98.7% (95% CI 96.8–99.4%). Heterogeneity was more pronounced than for sensitivity ($Q = 11.09$, 5 degrees of freedom, $p = 0.050$; $I^2 = 54.9\%$; $\tau^2 = 0.535$). The 1,031-case Marchetti cohort contributed 998 FISH-negative patients and reported only two IHC-positive/FISH-negative results with the D5F3 companion-diagnostic platform (Figure 3).

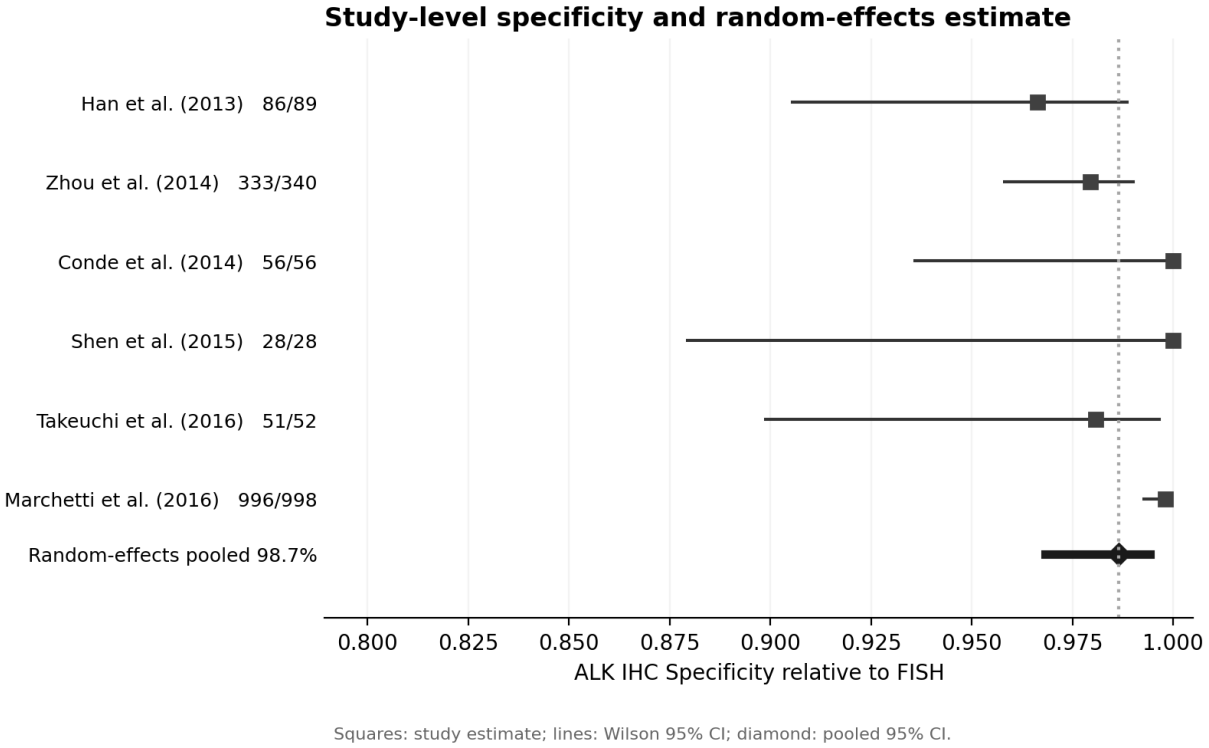


Figure 3. Study-level specificity and random-effects summary estimate of ALK IHC relative to FISH. Squares indicate individual estimates and horizontal lines the Wilson 95% CI.

3.6 Summary accuracy measures and between-study variability

The pooled sensitivity and specificity combine information from different denominators; their interpretation should preserve that distinction. The diagnostic positive likelihood ratio (sensitivity/[1 − specificity]) calculated from the marginal pooled point estimates was approximately 74.0, and the negative likelihood ratio ([1 − sensitivity]/specificity)

approximately 0.039. These ratios are algebraically derived summaries, not separate jointly fitted likelihood-ratio meta-analyses, and uncertainty around them was not estimated. Similarly, the pooled summary odds ratio implied by those two point estimates should not be reported as an independently validated hierarchical diagnostic odds ratio.

Table 3. Diagnostic accuracy summary and heterogeneity statistics.

Measure	Random-effects summary	95% CI	Q (5 df)	I^2	τ^2
Sensitivity	96.2%	91.7–98.3%	6.25	20.0%	0.205
Specificity	98.7%	96.8–99.4%	11.09	54.9%	0.535
Observed agreement	98.9%	Descriptive	—	—	—
Cohen's κ	0.961	Descriptive	—	—	—

3.7 Antibody and threshold comparisons within the same cohorts

The largest unselected paired comparison provided evidence of assay-specific false positives. In 1,031 adenocarcinomas, the Ventana D5F3 companion assay and a 5A4 protocol both identified 30 of 33 FISH-positive cases (90.9% sensitivity), but there were two versus 17 IHC-positive/FISH-negative cases respectively, corresponding to specificities of 99.8% and 98.3% [11]. Thus, equivalence in sensitivity did not imply equivalence in the chance of a positive result being discordant with FISH.

In the 60-case manual staining study, D5F3 and 5A4 produced respective sensitivities of 93.8% and 56.3% against the same 32 FISH-positive tumors, despite 100% specificity for both in the 28 FISH-negative controls [9]. Such comparisons cannot be entered as independent cohorts because the tested specimens are shared. They support the importance of laboratory-level protocol validation and discourage transferring performance estimates between antibody platforms or staining systems solely on the basis of clone name.

Table 4. Within-cohort assay comparisons; shared specimens counted only once in the primary meta-analysis.

Cohort and assay	TP/FISH-positive	FP/FISH-negative	Sensitivity	Specificity
Marchetti: D5F3 CDx	30/33	2/998	90.9%	99.8%
Marchetti: 5A4	30/33	17/998	90.9%	98.3%
Shen: manual D5F3	30/32	0/28	93.8%	100.0%
Shen: manual 5A4	18/32	0/28	56.3%	100.0%

3.8 Sensitivity analyses and influence of patient selection

Omitting one cohort at a time yielded pooled sensitivity estimates of 95.0%–97.3% and pooled specificity estimates of 97.8%–99.0%. The largest sensitivity shift occurred on exclusion of the Takeuchi trial-screened cohort (pooled sensitivity 95.0%), whereas the largest specificity shift occurred on exclusion of the large Marchetti cohort (pooled specificity 97.8%). The direction of the overall findings was preserved, but the magnitude depended on the selected study populations and verified reference-negative denominators.

Table 5. Leave-one-study-out random-effects sensitivity analysis.

Study excluded	Pooled sensitivity	Pooled specificity
Han 2013	96.2%	99.0%
Zhou 2014	96.1%	98.9%
Conde 2014	96.1%	98.6%
Shen 2015	96.9%	98.7%
Takeuchi 2016	95.0%	98.7%
Marchetti 2016	97.3%	97.8%

Restricting the calculation to the two relatively unselected adenocarcinoma cohorts, Zhou and Marchetti (1,399 paired specimens), yielded a descriptive crude sensitivity of 58/61 (95.1%) and specificity of 1,329/1,338 (99.3%). The high fraction of FISH-positive specimens in some other cohorts makes population-based positive predictive value particularly sensitive to changes in case mix.

3.9 Predictive values and clinical interpretation

Sensitivity and specificity do not by themselves determine whether an individual positive result is correct in an unselected patient population. Using the pooled point estimates, expected positive predictive value rises with ALK rearrangement prevalence, while the negative predictive value remains high at low prevalence. Table 6 illustrates the effect of case mix using expected results per 1,000 tested patients; these are model-derived clinical scenarios, not additional observed cases.

Table 6. Modeled ALK IHC predictive values at alternative FISH-positive pretest probabilities.

Prevalence	PPV	NPV	TP/1,000	FP/1,000	FN/1,000	TN/1,000
1%	41.9%	100.0%	9.6	13.4	0.4	976.6
3%	68.8%	99.9%	28.9	13.1	1.1	956.9
5%	78.9%	99.8%	48.1	12.8	1.9	937.2

At 3% prevalence, for example, the modeled positive predictive value is approximately 70%, even though pooled specificity approaches 99%. Consequently, a small number of IHC-positive/FISH-negative specimens can influence the credibility of positive results in low-prevalence practice settings. Conversely, a negative validated IHC result is highly reassuring, but low-confidence or discordant findings should be evaluated in their clinical context.

3.10 Study quality, reference-standard limitations and applicability

Quality concerns clustered around patient selection and flow/timing. Zhou and Marchetti included relatively unselected adenocarcinoma populations, whereas Han, Conde and Shen had enriched or selected samples; Takeuchi preferentially verified a subset of the IHC-negative population by FISH [6–11]. FISH was consistently the comparator but, by definition, cannot resolve every functional or transcript-level ALK fusion. High apparent accuracy in deliberately enriched cohorts should therefore be distinguished from accuracy in routine reflex testing.

Table 7. QUADAS-2 domain-oriented judgments and principal applicability considerations.

Study	Selection	Index test	FISH comparator	Flow/timing	Main consideration
Han 2013	High	Unclear	Unclear	Unclear	Advanced, enriched series; denominator restricted to paired cases
Zhou 2014	Low	Unclear	Unclear	Low	Unselected cases; 368 complete paired results
Conde 2014	High	Unclear	Unclear	Unclear	Selected, high ALK-positive fraction
Shen 2015	High	Unclear	Unclear	Low	Deliberately selected antibody-comparison panel
Takeuchi 2016	High	Unclear	Unclear	High	Selective FISH assessment among IHC-negative cases
Marchetti 2016	Low	Low	Unclear	Low	All 1,031 adenocarcinomas underwent both tests

The QUADAS-2 categories distinguish Low, High and Unclear risk of bias in the domains shown. Selection reflects consecutive versus enriched sampling and flow/timing reflects completeness of paired verification. These judgments should be interpreted at the domain level rather than aggregated into an unsupported overall numerical quality score.

3.11 Discordant results and potential clinical consequences

Of the 20 discordant pairs across all included studies, 13 were IHC-positive/FISH-negative and seven IHC-negative/FISH-positive. The interpretation of these 20 observations differs from their simple classification as "false" because FISH uses genomic signal patterns, whereas IHC measures protein expression. In the Marchetti cohort, all five discordant cases underwent additional sequencing; published literature reviewed by those authors also demonstrated clinical responses to ALK inhibitors in selected IHC-positive/FISH-negative cases [11]. A later reevaluation using IHC, FISH and NGS further showed the importance of adjudicating atypical results molecularly [14].

Table 8. Discordance patterns and recommended pathological review.

Pattern	Potential explanation	Interpretive response
IHC+/FISH−	Atypical ALK fusion, borderline FISH signal, copy-number changes, staining nonspecificity	Review stain localization/intensity and FISH signals; consider orthogonal RNA or DNA testing
IHC−/FISH+	Low expression, fixation artifact, rare fusion, FISH borderline or artifactual signal	Confirm tumor adequacy, repeat validated IHC/FISH, and consider sequencing
Equivocal IHC	Weak focal staining, nonvalidated threshold or variable signal amplification	Apply validated assay criteria; repeat/confirm rather than assign binary status
Low cellularity	Insufficient nuclei or nonrepresentative tumor tissue	Repeat on a more suitable block or specimen if available

4. DISCUSSION

4.1 Interpretation of the principal findings

The central finding is that validated ALK IHC is highly concordant with ALK break-apart FISH, with pooled sensitivity of 96.2% and specificity of 98.7% in six directly comparable cohorts. The small absolute number of discordant cases did not eliminate clinically relevant differences between assay protocols, especially among FISH-negative patients. Sensitivity and specificity were heterogeneous across selected populations and laboratory methods, whereas overall agreement was inflated by the large number of concordantly negative cases. The distinction between pooled random-effects accuracy, crude agreement and predictive value is therefore essential.

These results are compatible with major molecular-testing recommendations recognizing ALK IHC as an alternative to FISH when appropriately validated [1]. They also align with larger earlier evidence syntheses in which D5F3-based assays approached pooled sensitivity of 97% and specificity of 99% [18]. However, external pooled estimates from broad reviews and the present six-cohort results should not be compared as though they represented independent patient populations or identical reference standards.

4.2 Technical determinants of discordance

Pre-analytical variables influence both modalities. Cold ischemia, fixation duration and quality of paraffin-embedded material can reduce ALK antigen detectability or interfere with nuclear FISH scoring. IHC requires appropriately controlled granular cytoplasmic staining, while weak nonspecific staining or excessive amplification can affect specificity. FISH classification is vulnerable to uncommon fusion configurations, truncation of nuclei in thin sections, signal separation thresholds, and insufficient viable tumor nuclei. Tissue heterogeneity and sampling may produce discrepant results even when both assays are analytically valid [1,7,11,14].

For these reasons, neither a strongly positive IHC result nor an atypical FISH result should automatically be reduced to a binary technical disagreement without assessing histology, tumor fraction, signal quality and molecular context. RNA-based fusion testing can sometimes establish whether an expressed ALK fusion transcript is present. DNA-based sequencing may be informative but can miss certain rearrangements when intronic breakpoints are poorly covered; the choice of orthogonal test should therefore be tailored to the available laboratory platform and specimen quality.

4.3 Antibody clone, protocol and threshold

Direct within-cohort comparisons indicate that assay identity alone is an incomplete predictor of diagnostic performance. In the large 1,031-case series, D5F3 and 5A4 shared sensitivity but differed in specificity; in a much smaller manual staining panel, the 5A4 protocol performed poorly for sensitivity compared with manual D5F3 [9,11]. The apparent contradiction is explained by differences in antibody sources, detection chemistry, amplification, slide processing and thresholds. It is not a justified inference that all 5A4 assays are clinically inferior, or that all D5F3 workflows are intrinsically interchangeable.

A laboratory adopting ALK IHC should establish analytical validation using its own tissue-processing workflow, adequate positive and negative controls, standardized cutoffs, and competency review for equivocal cases. Cases with borderline staining require predefined decision pathways rather than post hoc changes in positivity thresholds. Where a validated companion diagnostic is used, compliance with its specified platform and scoring criteria is particularly important.

4.4 Interpretation of predictive values and prevalence

An accurate test can have only moderate positive predictive value when the target condition is uncommon. At a modeled 3% ALK rearrangement prevalence, a sensitivity of 96.2% and specificity of 98.7% imply positive predictive value of approximately 68.8% and negative predictive value of 99.9%. These values depend on pretest prevalence and are not

appropriate substitutes for the observed PPV in enriched trials, where ALK positivity may be much more frequent. Screening populations, population prevalence, and selection criteria therefore determine whether positive and negative ALK IHC results carry comparable clinical implications.

4.5 How the meta-analytic method affects interpretation

The random-effects model acknowledges real variation between studies, but six cohorts provide limited precision for characterizing between-study distributions. Separate univariate synthesis is transparent and reproducible from the available 2 × 2 tables, yet it cannot incorporate the negative correlation between sensitivity and specificity that may arise from varying staining thresholds. PRISMA-DTA favors specifying the statistical model, and hierarchical bivariate or HSROC methods are often preferred when a sufficiently informative set of study-level data and thresholds is available [15,16].

The 95% confidence intervals around the pooled point estimates primarily quantify statistical uncertainty for the selected studies; they do not establish that a new laboratory using an unvalidated assay would achieve the same accuracy. Prediction intervals would be informative but are unstable with so few studies and substantially different enrichment/verification practices. Formal comparative meta-regression or publication-bias testing was not performed because of the small number of independent cohorts and heterogeneous methods.

4.6 Risk of bias and strengths of evidence

Two comparatively unselected cohorts generated 1,399 paired observations and provide useful information about routine testing, while the other four cohorts broadened the range of FISH-positive cases and antibody protocols. The key strength is that all six studies contributed directly reconstructable paired diagnostic counts, permitting study-level visualization and recalculation of pooled effects. The principal limitation is patient-selection bias from deliberate enrichment, together with differential verification and differences in clinical specimens and FISH/IHC scoring.

The very high sensitivity in trial-screened populations is not direct evidence of superiority over population-based series. Moreover, treating FISH as the sole reference assigns all IHC-positive/FISH-negative cases to the false-positive category even when an expressed actionable ALK fusion could be demonstrated by sequencing. This comparator limitation is especially important when discordant samples respond to ALK-directed therapy. The diagnostic question addressed here is agreement with a historical FISH comparator rather than complete identification of all functional ALK fusions.

4.7 Clinical implications and future research

An evidence-informed ALK testing workflow should begin with pathology review and adequacy assessment, followed by a validated IHC or comprehensive molecular assay appropriate to local resources. Unequivocal results from a validated platform can inform treatment selection, whereas equivocal, unexpected or discordant findings should prompt review of the stain and FISH signal patterns and consideration of an orthogonal technique. Repeat sampling may be needed if the original tumor material is scant or poorly fixed [1].

Future prospective studies should enroll consecutive, unselected NSCLC cases; apply prespecified IHC thresholds and independent or blinded test interpretation; verify all participants with both IHC and FISH; document uninterpretable results; and resolve discordant specimens through a uniform sequencing strategy with clinical follow-up. Such evidence would clarify true diagnostic accuracy for expressed therapeutic targets, estimate assay-specific performance without selection bias, and inform cost-effectiveness in different diagnostic settings.

5. CONCLUSION

Across six paired-testing cohorts including 1,868 NSCLC specimens, ALK IHC demonstrated high diagnostic agreement with FISH: pooled sensitivity 96.2%, pooled specificity 98.7% and crude agreement 98.9%. Performance was affected by antibody protocol, staining threshold, patient selection and completeness of paired verification. A properly validated ALK IHC assay is a reliable diagnostic option, but discordant or equivocal findings require integrated pathological interpretation and selective orthogonal testing. Broader prospective studies with molecular adjudication are needed to define accuracy for actionable ALK fusions beyond the limits of FISH comparison.

REFERENCES

1. Lindeman NI, Cagle PT, Aisner DL, et al. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors. J Thorac Oncol. 2018;13(3):323–358. doi:10.1016/j.jtho.2017.12.001.

2. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448:561–566. doi:10.1038/nature05945.

3. Pyo JS, Kang G, Sohn JH. ALK immunohistochemistry for ALK gene rearrangement screening in non-small cell lung cancer: a systematic review and meta-analysis. J Biomark. 2016. doi:10.5301/jbm.5000218.

4. Jiang L, Yang H, He P, et al. Improving selection criteria for ALK inhibitor therapy in non-small cell lung cancer: a pooled-data analysis on diagnostic operating characteristics of immunohistochemistry. Am J Surg Pathol. 2016;40(5):697–703. doi:10.1097/PAS.0000000000000604.

5. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

6. Han XH, Zhang NN, Ma L, et al. Immunohistochemistry reliably detects ALK rearrangements in patients with advanced non-small-cell lung cancer. Virchows Arch. 2013;463(4):583–591. doi:10.1007/s00428-013-1472-7.

7. Zhou J, Zhao J, Sun K, et al. Accurate and economical detection of ALK positive lung adenocarcinoma with semiquantitative immunohistochemical screening. PLoS One. 2014;9(3):e92828. doi:10.1371/journal.pone.0092828.

8. Conde E, Suárez-Gauthier A, Benito A, et al. Accurate identification of ALK positive lung carcinoma patients: novel FDA-cleared automated fluorescence in situ hybridization scanning system and ultrasensitive immunohistochemistry. PLoS One. 2014;9(9):e107200. doi:10.1371/journal.pone.0107200.
9. Shen Q, Wang X, Yu B, et al. Comparing four different ALK antibodies with manual immunohistochemistry (IHC) to screen for ALK-rearranged non-small cell lung cancer. Lung Cancer. 2015;90(3):492–498. doi:10.1016/j.lungcan.2015.10.002.
10. Takeuchi K, Togashi Y, Kamihara Y, et al. Prospective and clinical validation of ALK immunohistochemistry: results from the phase I/II study of alectinib for ALK-positive lung cancer (AF-001JP study). Ann Oncol. 2016;27(1):185–192. doi:10.1093/annonc/mdv501.
11. Marchetti A, Di Lorito A, Pace MV, et al. ALK protein analysis by IHC staining after recent regulatory changes: a comparison of two widely used approaches, revision of the literature, and a new testing algorithm. J Thorac Oncol. 2016;11(4):487–495. doi:10.1016/j.jtho.2015.12.111.
12. Murthy SS, Rajappa SJ, Gundimeda SD, et al. Anaplastic lymphoma kinase status in lung cancers: an immunohistochemistry and fluorescence in situ hybridization study from a tertiary cancer center in India. Indian J Cancer. 2017;54(1):231–235. doi:10.4103/0019-509X.219533.
13. Fiset PO, Labbé C, Young K, et al. Anaplastic lymphoma kinase 5A4 immunohistochemistry as a diagnostic assay in lung cancer: a Canadian reference testing center's results in population-based reflex testing. Cancer. 2019;125(22):4043–4051. doi:10.1002/cncr.32422.
14. Tobiášová K, Barthová M, Janáková Ľ, et al. Discordant ALK status in non-small cell lung carcinoma: a detailed reevaluation comparing IHC, FISH, and NGS analyses. Int J Mol Sci. 2024;25(15):8168. doi:10.3390/ijms25158168.
15. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA. 2018;319(4):388–396. doi:10.1001/jama.2017.19163.
16. Salameh JP, Bossuyt PM, McGrath TA, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ. 2020;370:m2632. doi:10.1136/bmj.m2632.
17. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009.
18. Ma H, Yao WX, Huang L, et al. Efficacy of D5F3 IHC for detecting ALK gene rearrangement in NSCLC patients: a systematic review and meta-analysis. Oncotarget. 2016;7(43):70128–70142. doi:10.18632/oncotarget.11806.