Supplementary Material

1. Appendix A – Search Strategy

Final search strategy: #6

#1 AND #2 AND #3 = #5

#5 NOT #4 = #6

Inclusion date: 1990- 2024

01/01/1990- 31/12/2024: 8685 papers

Filter:

Clinical Study, Clinical Trial, Clinical Trial Protocol, Clinical Trial, Phase I, Clinical Trial, Phase II, Clinical Trial, Phase III, Clinical Trial, Phase IV, Comparative Study, Evaluation Study, Meta-Analysis, Multicenter Study, Observational Study, Pragmatic Clinical Trial, Randomized Controlled Trial, Review, Systematic Review, Validation Study, English, Humans, MEDLINE

'Cancer' Keywords

Search #1

("cancer" OR "neoplasms" OR "malignancy" OR "tumor" OR "carcinoma" OR "oncology" OR "malignant disease" OR "breast cancer" OR "breast carcinoma" OR "ductal carcinoma" OR "lobular carcinoma" OR "mammary neoplasm" OR "prostate cancer" OR "prostatic carcinoma" OR "adenocarcinoma of the prostate" OR "prostate neoplasm" OR "colorectal cancer" OR "colon cancer" OR "rectal cancer" OR "colorectal carcinoma" OR "colon adenocarcinoma" OR "colorectal neoplasm" OR "large bowel cancer" OR "lung cancer" OR "pulmonary carcinoma" OR "non-small cell lung cancer" OR "small cell lung cancer" OR "bronchogenic carcinoma" OR "lung neoplasm")

‘Quality Indicators’ Keywords

Search #2

("quality indicator*" OR "quality measure*" OR "performance indicator*" OR "performance measure*" OR "benchmark*" OR "process measure*" OR "outcome measure*" OR "quality standard*" OR "clinical indicator*" OR "healthcare quality" OR "quality assessment" OR "quality assurance" OR "quality control" OR "service quality" OR "quality improvement" OR "quality metric*" OR "quality evaluation" OR "healthcare indicator*")

‘Anatomic Pathology, Molecular Pathology & Cytopathology’ Keywords

Search #3

("pathology" OR "histopathology" OR "anatomical pathology" OR "diagnostic pathology" OR "tissue pathology" OR "clinical pathology" OR "surgical pathology" OR "molecular pathology" OR "genetic testing" OR "DNA analysis" OR "RNA sequencing" OR "genomics" OR "epigenetics" OR "PCR" OR "next-generation sequencing" OR "molecular diagnostics"OR "cytopathology" OR "fine needle aspiration" OR "Pap smear" OR "cytology" OR "fluid cytology" OR "cell block" OR "exfoliative cytology" OR "cytopathologic evaluation")

Expanded Search String for Exclusion

Search #4

("chemopathology" OR "chemical pathology" OR "clinical chemistry" OR "biochemistry" OR "toxicology" OR "metabolic pathology" OR "immunopathology" OR "immune response" OR "autoimmunity" OR "immunology" OR "autoimmune diseases" OR "antibody testing" OR "immune markers" OR "hypersensitivity" OR "haematopathology" OR "hematopathology" OR "blood pathology" OR "leukaemia" OR "lymphoma" OR

"bone marrow pathology" OR "hematologic diseases" OR "hemoglobinopathies" OR "anaemia OR "forensic pathology" OR "forensics" OR "cause of death" OR "medicolegal pathology" OR "autopsy findings" OR "postmortem examination" OR "toxicology reports")

2. Appendix B – Supplementary Table 1

Supplementary Table 1. SummaryTtable of All Included Studies

Reference Year	Country	Author	Study Title	Funding	Methodology	Research Design	Sample Size	Type of Cancer	Domain	Phase	Part	Aim	Outcome	Purpose	Quality Indicators
2008	USA	Hecht et al.	Validation of tissue microarray technology in ovarian cancer: results from the Nurses' Health Study	National Cancer Institute	Quantitative	Prospective Observational	174	Gynaecological	Histopathology	Analytic	Staining and preparation, Diagnostic accuracy	To evaluate the effect of tissue characteristics (i.e., age of sample, histology, borderline, and invasion) on staining, to estimate the number of core biopsies required to ensure adequate representation of the entire tumor, and to calculate inter-rater reliability for	The interclass correlation coefficient and the mean Spearman correlation coefficients among 3 cores were ≥0.91 and 0.87, respectively. Tissue age and tumor histology were not predictive of an	Completeness, Accuracy	Number of core biopsies required for adequate tumour representation Inter-rater reliability for stain interpretation

stain interpretation	inability to evaluate stains, but borderline tumors had a 2 to 4-fold increase in the risk of having uninterpretable cores over invasive tumors. There was moderate to substantial concordance between the two pathologists for estrogen receptor-α.

												Overall, quality was better in countries that have operated organised programmes for a longer time, plausibly as a result of long-lasting monitoring and quality assurance activities. There is a clear need to standardise the cytological and histological classificati		Referral rate for repeat cytology Referral rate for colposcopy Attendance rate of colposcopy after referral Detection rate of CIN2
											To report screening process performance and making comparisons within countries to help identify impact of screening activities in terms of patient outcome				
										Diagnostic accuracy Patient outcome and follow up				
			Process performance of cervical screening programmes in Europe			Retrospective Observational			Analytic, Post-analytic					
2009	italy	Ronco et al.		European Network for Information on Cancer	Quantitative		Gynaecological	Cytopathology					Accuracy	

													ons used in screening, as well as data registration systems across Europe. With the scrape sampling technique, control smears showed abundant epithelial cells and a cleaner background and processing was faster than with tissue block sampling.			
2006	Sweden	Lexander et al.	Evaluation of two sample preparation methods for prostate proteome analysis	NA	Quantitative	Non-randomised Comparative	11	Prostate	Molecular Pathology	Analytic	Turnaround time, Staining and preparation	To compare a scraping technique for harvesting tissue from radical prostatectomy specimens for proteome analysis and compare it with the standard technique	For proteomic analysis, the	Timeliness, Completeness	Time taken to process proteomic analysis Quality of proteomic analysis

													scrape sample technique has several advantages over the tissue block method. We found a poor concordance between referral and hospital cervical cytology reports, and a better concordance of hospital cytology reports with biopsy findings than of referral		
											To assess the degree of concordance between cervical cytology reports from laboratories outside our hospital with hospital cytology reports, and of both, with biopsy reports.			Interlaboratories concordance rate between cytology smear results	
			A comparison between referral and hospital cervical cytology laboratory reports			Non-randomised Comparative									Concordance rate between cytology smear and biopsy results
1995	Israel	Korach et al.		NA	Quantitative		95	Gynaecological	Cytopathology	Analytic	Diagnostic accuracy			Accuracy		

													cytology reports. There was a tendency for overdiagnosis in both the referral and the hospital cytology reports.		
			Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study								To quantify the accuracy and reproducibility of pathologists' diagnoses of melanocytic skin lesions	Diagnoses spanning moderately dysplastic nevi to early stage invasive melanoma were neither reproducible nor accurate in this large study of		Inter-rater and intra-rater concordance rate
2017	USA	Elmore et al		National Cancer Institute	Quantitative	Non-randomised Comparative	187	Skin	Histopathology	Analytic	Diagnostic accuracy			Accuracy	Diagnostic accuracy and reproducibility

pathologists in the USA. Efforts to improve clinical practice should include using a standardized classification system, acknowledging uncertainty in pathology reports, and developing tools such as molecular markers to support pathologists' visual assessments.

2016	UK	Stanczuk et al	Clinical validation of hrHPV testing on vaginal and urine self-samples in primary cervical screening (cross-sectional results from the Papilloma virus Dumfries and Galloway-PaVDaG study)	Chief Scientist Office, Scotland	Quantitative	Retrospective Observational	5318	Gynaecological Cytopathology	Pre-analytic, Analytic	Sample Collection, Diagnostic accuracy	To assess the sensitivity and specificity for CIN2+ through biospecimens and interlaboratory and intralaboratory performance of the hrHPV assay in self-collected samples	The sensitivity of self-collected vaginal samples for the detection of CIN2+ was similar to that of cervical samples and justifies consideration of this sample for primary screening. The intralaboratory and interlaboratory agreement for hrHPV positivity in	Accuracy	Specimen quality between self-collected and clinician-taken samples Inter- and intralaboratory concordance rate

self-samples was high (κ values 0.98 (0.96 to 0.99) and 0.94 (0.92 to 0.97) for vaginal samples and 0.95 (0.93 to 0.98) and 0.90 (0.87 to 0.94) for urine samples).

2004	USA	Diaz et al	The use of TMA for interlaboratory validation of FISH testing for detection of HER2 gene amplification in breast cancer	Quantitative	Non-randomised Comparative	41	Breast	Molecular Pathology	Analytic	Diagnostic accuracy	To assess the usefulness of TMA technology for validation of HER2 FISH testing.	For positive and negative results, the concordance rate between laboratories was 100%. Using kappa statistical analysis to determine interobserver agreement, HER2 to chromosome 17 gene copy ratios showed strong agreement between laboratori	Accuracy	Concordance rate between immunohistochemistry and fluorescence in situ hybridisation Inter laboratory concordance rate Interobserver agreement

es with kappa = 0.85 (perfect agreement = 1.0). Good concordance was observed between HER2 IHC and in situ hybridization testing. TMA is a robust and effective method for the technical validation of HER2 FISH testing and should be considered for use by quality

													assessment programs. Management with only cytological follow-up would result in 29 residual cases of recurrent CIN3+ by 10 years and would cost £358 222 (€440 426; $574 910) (discounted) per 1000 women treated.		
2012	UK	Legood et al	Cost-effectiveness of human papillomavirus testing after treatment for cervical intraepithelial neoplasia	NHS Cancer Screening Programme	Quantitative	Other	NA	Gynaecological	Cytopathology	Post-analytic	Patient Outcomes and Follow up, Cost-effectiveness	To evaluate the cost effectiveness of human papillomavirus testing after treatment for cervical intraepithelial neoplasia (CIN).	Implementation of human papilloma virus test of cure in	Accuracy, Cost-effectiveness	Cost per treatment Number of cancer cases averted

cytologically negative women according to the sentinel sites protocol would avert an additional 8.4 cases of CIN 3+ and reduce costs by £9388 per 1000 women treated.

2018	Italy	Stabile et al	Not All Multiparametric Magnetic Resonance Imaging-targeted Biopsies Are Equal: The Impact of the Type of Approach and Operator Expertise on the Detection of Clinically Significant Prostate Cancer	NA	Quantitative	Non-randomised Comparative	244	Prostate	Histopathology	Pre-analytic	Sample collection	To compare the rate of detection of clinically significant prostate cancer (csPCa) of different mpMRI targeted approaches and to assess the role of operator expertise in the detection of csPCa	The overall csPCa detection rate was 58% for FB and 45% for CB (p = 0.07). A significantly higher rate of csPCa detection in targeted samples was observed for FB compared to CB (57% vs 36%; p = 0.002). On MVA, FB and operator expertise were significantly	Accuracy	Experience of operator

associated with a higher probability of csPCa detection in targeted samples (odds ratio [OR] 2.4 and 1.7, respectively; both $p \leq 0.03$). When the same analyses were repeated for patients undergoing FB, operator expertise remained an independent predictor of csPCa

													detection (OR 1.9; p = 0.004).			
2008	Sweden, France, Spain, Germany, UK, Denmark	Dillner et al	Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: joint European cohort study	European Union Biomed 5 contract HPV-based cervical cancer screening, Digene, Spanish Ministry of Health, Swedish Cancer Society, Europe against Cancer, ARC (Association de Recherche contre le Cancer), the Region Champagne-Ardenne and	Quantitative	Prospective Observational	24295	Gynaecological	Cytopathology	Post-analytic	Patient outcomes and follow up	To obtain large scale and generalisable data on the long term predictive value of cytology and human papillomavirus (HPV) testing for development of cervical intraepithelial neoplasia grade 3 or cancer (CIN3+).	The cumulative incidence rate among women with negative cytology results who were positive for HPV increased continuously over time, reaching 10% at six years, whereas the rate among women with positive cytology results	Accuracy	Long term predictive value of cancer

				Cancéropôle Grand Est									who were negative for HPV remained below 3%. Of the 160 cases in both surveys, 111 (69%) achieved 90% consensus (graded). All 43 graded cores scored as IHC-positive were fluorescence in situ hybridization-positive, whereas all but 3 of the 68 IHC-negative graded		
2006	USA	Fitzgibbons et al	Interlaboratory comparison of immunohistochemical testing for HER2: results of the 2004 and 2005 College of American Pathologists HER2 Immunohistochemistry Tissue Microarray Survey	NA	Quantitative	Retrospective Observational	160	Breast	Molecular Pathology	Analytic	Diagnostic accuracy	To review the results of the 2004 and 2005 College of American Pathologists HER2 Immunohistochemistry Tissue Microarray Survey.		Accuracy	Concordance between laboratories and inter-diagnostic test agreement

cores were fluorescence in situ hybridization-negative. Ninety-seven (95%) of 102 laboratories in 2004 and 129 (91%) of 141 laboratories in 2005 correctly scored at least 90% of the graded cores.

2010	The Netherlands	Gök et al	HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study	Delphi Bioscience, Scherpenzeel, Zeist, Netherlands, Noord-Holland and Flevoland, IKA (Comprehensive Cancer Center Amsterdam)	Quantitative	Retrospective Observational	28073	Gynaecological	Cytopathology	Pre-analytic	Sample collection	To determine whether offering self sampling of cervicovaginal material for high risk human papillomavirus (HPV) testing is an effective screening method for women who do not attend regular cervical screening programmes.	The compliance rate in the self sampling group was significantly higher than in the control group (crude 26.6% v 16.4%, P<0.001; adjusted 27.5% v 16.6%, P<0.001). Offering self sampling by sending a device for collecting cervicovaginal specimens for high risk HPV	Completeness	Compliance rate of self sampler

testing to women who did not attend regular screening is a feasible and effective method of increasing coverage in a screening programme.

														At least 70% of programmes could provide data for most indicators. Overall, 3.2% of smears were classified as unsatisfactory. At a national level 6.6% of women was advised to repeat cytology and 62.2% of them actually did. However 13/71 programmes		Re-referral rate Compliance to repeat cytology Compliance with colposcopy
2006	Italy	Ronco et al	Activity level and process indicators of organised programmes for cervical cancer screening in Italy	The Italian National Cancer Screening Programme	Quantitative	Retrospective Observational	8,698,480	Gynaecological	Cytopathology	Analytic, Post-analytic	Diagnostic accuracy Patient outcome and follow up	To calculate national statistics, to make comparisons between programmes and to monitor screening quality		Completeness, Accuracy	Cancer detection rate

recommended repeat cytology to >10% of screened women. The raw detection rate of biopsy-proven CIN2+ was 2.7 per 1,000 screened women (2.8 per 1,000 when standardised on the Italian population).

1997	Italy	Anania et al	Percutaneous large core needle biopsy versus surgical biopsy in the diagnosis of breast lesions	NA	Quantitative	Retrospective Observational	196	Breast	Histopathology	Pre-analytic, Analytic	Sample collection, Diagnostic accuracy	To value LCNB accuracy in the determination of morphobiological parameters and as an alternative to the open SB diagnostic procedure of breast lesions.	Percutaneous LCNB has high diagnostic accuracy for histological classification. LCNB has the same accuracy as SB for morphobiological parameters. The cost of LCNB is markedly lower than SB.	Accuracy, Cost-effectiveness	Rate of inadequate sampling Cost of sampling Diagnostic accuracy of biopsy

2015	France	Long et al	Why and how immunohistochemistry should now be used to screen for the BRAFV600E status in metastatic melanoma? The experience of a single institution (LCEP, Nice, France)	NA	Quantitative	Non-randomised Comparative	188	Skin	Molecular Pathology	Analytic	Turnaround time, Diagnostic accuracy	To assess and compare the specificity, sensibility, cost-effectiveness and turnaround time (TAT) of immunohistochemistry (IHC) and molecular biology for detection of the BRAFV600E mutation in metastatic melanoma patients.	The BRAFV600E mutation was detected by pyrosequencing in 91/188 (48%) patients. IHC was strongly positive (3+) in all of these 91 cases. IHC was strongly positive in 9/188 (5%) cases in which the molecular testing failed due to non-amplifiable DNA. The average	Timeliness, Accuracy	Turnaround time of pathology test Concordance rate between IHC and molecular biology

2004	UK	Harrison et al	Analysis of inadequate cervical smears using Shewhart control charts	NA	Quantitative	Retrospective Observational	100	Gynaecological	Cytopathology	Pre-analytic	Sample collection	To assess variation in cervical smear quality among general practices and laboratories	TAT was 48 h vs. 96 h, for IHC vs. molecular biology testing, respectively. There was wide variation in the proportion of inadequate smears with 23% of practices showing evidence of special cause variation and four of the six laboratories showing evidence of special cause	Accuracy	Proportion of inadequate smears between general practices and laboratories

variation.

There was no evidence of a clinically important association between high rates of inadequate smears and better detection of dyskaryosis (R2 = 0.082).

2007	Australia	Youl et al	Diagnosing skin cancer in primary care: how do mainstream general practitioners compare with primary care skin cancer clinic doctors?	NA	Quantitative	Non-randomised Comparative	11116	Skin	Histopathology	Pre-analytic	Sample collection	To measure and compare the casemix and diagnostic accuracy of excised or biopsied skin lesions managed by mainstream general practitioners and doctors within primary care skin cancer clinics.	Overall, sensitivity for diagnosing any skin cancer was similar for skin cancer clinic doctors (0.94) and GPs (0.91), although higher for skin cancer clinic doctors for BCC (0.89 v 0.79; P < 0.01) and melanoma (0.60 v 0.29; P < 0.01). The overall NNE was similar for skin	Accuracy	Expertise of operator

														cancer clinic doctors (1.9; 95% CI, 1.8%-2.1 %) and GPs (2.1; 95% CI, 1.9%-2.3 %). This did not change after adjusting for years of clinical experienc e.		
1998	India	Kottke et al	The Quality of Pap Test Specimens Collected by Nurses in a Breast and Cervical Cancer Screening Clinic	NA	Quantitative	Non-randomised Comparative	232 18	Gynaecological	Cytopathology	Pre-analytic	Sample collection	To determine whether nurses can collect high-quality Pap tests.	The proportion of specimens that were wholly satisfactory was 79.8% (95% CI = 75.9–83.7) for the nurses,	Accuracy	Expertise of operator

														65.3% (95% CI = 56.7–73.9) for other Rosebud providers, and 81.7% (95% CI = 81.2–82.2) for non-Rosebud providers.		
			Next-generation sequencing of cytologic preparations: An analysis of quality metrics			Quality Improvement					Staining and preparation	To evaluate the quality of DNA extracted from cytology smears and liquid-based preparations (smears/LBPs) for next-generation sequencing (NGS).	Cytology smears and LBPs are viable and often superior alternatives to formalin-fixed samples for NGS in cancer diagnostics. They provide high-quality DNA suitable		Mean target coverage Median insert size Percentage of usable bases on target Duplication rate
2017	USA	Hwang et al		NA	Qualitative		26.87, 30	Non-specific	Cytopathology	Analytic		Researchers compared these DNA		Accuracy	Post-shearing

												samples with those obtained from core biopsies and cell blocks to assess their suitability for molecular testing.	for comprehensive molecular testing, especially when traditional biopsy materials are inadequate or unavailable.		fragment size
2012	Canada	McFadyen et al	Physician level reporting of surgical and pathology performance indicators: a regional study to assess feasibility and impact on quality	Cancer Care Ontario	Qualitative	Quality Improvement	437	Prostate Colorectal	Histopathology	Post-analytic	Report quality	The aim of the study was to evaluate the feasibility and impact of reporting surgical and pathology performance indicators at the individual physician level	The study demonstrated that physician-level reporting of surgical and pathology performance indicators is feasible in a regional setting and can positively	Completeness	Surgical margin status Lymph node yield

													influence clinical quality improvement without negative consequences.		
												To validate and standardize	Digital image analysis of Ki-67 expression in breast carcinoma is a reliable, reproducible, and standardized method that can enhance the accuracy of routine pathology assessments and improve	Accuracy	Accuracy
			Laboratory validation studies in Ki-67 digital image analysis of breast carcinoma: a pathway to routine quality assurance									digital image analysis of Ki-67 expression in breast carcinoma as a reliable tool for routine quality assurance in histopathology.			Reproducibility
															Inter-laboratory consistency
															Comparative performance of digital image analysis tools
2019	Australia	Wang et al		NA	Quantitative	Quality Improvement	unsure	Breast	Molecular Pathology	Analytic	Diagnostic accuracy				

													quality assurance practices The study found significant variability in cervical cytology practices across Estonia, with disparities in laboratory performance, improved specimen adequacy over time, and a need for enhanced quality assurance measures to improve		Specimen quality Cytology-histology correlation Inter-laboratory consistency Accuracy of cervical cytology
2018	Estonia	Orumaa et al	Quality Assessment of Cervical Cytology Practices in Estonia From 2007 to 2018	Estonian Research Council	Qualitative	Retrospective Observational	138	Gynaecological	Cytopathology	Pre-analytic, Analytic	Sample collection, diagnostic accuracy	To evaluate the quality of cervical cytology specimen collection, evaluation, and reporting in Estonia		Accuracy	

												screening accuracy. DNA image analysis techniques showed varying levels of reproducibility across different instruments and laboratories, highlighting the need for standardized protocols to ensure consistent and reliable results.		
2010	The Netherlands	Thunissen et al	Quality assurance in DNA image analysis on diploid cells	EEC	Quantitative	Quality Improvement	NA	Non-specific	Molecular Pathology	Analytic	Diagnostic accuracy, Staining and preparation	Evaluate the performance, accuracy, and reproducibility of DNA image analysis techniques across different laboratories to ensure reliable measurements of DNA in diploid cells.		Completeness, Accuracy	Inter-instrument reproducibility Accuracy of DNA measurements Reliability of DNA image analysis techniques

2013	UK	Deans et al	Improvement in the quality of molecular analysis of EGFR in non-small-cell lung cancer detected by three rounds of external quality assessment	NA	Quantitative	Prospective Observational	143	Lung	Molecular Pathology	Analytic	Diagnostic accuracy	To evaluate and improve the accuracy and reliability of EGFR mutation testing across multiple laboratories through three rounds of external quality assessment	The study found significant improvement in the accuracy and reliability of EGFR mutation testing across laboratories after three rounds of external quality assessment, with enhanced performance in interpreting and reporting results.	Accuracy	Accuracy of EGFR mutation detection Reliability and consistency in result interpretation

2010	Italy	Risio et al	Virtual microscopy for histology quality assurance of screen-detected polyps	Regione Piemonte	Qualitative	Retrospective Observational	457	Colorectal	Histopathology	Analytic	Diagnostic accuracy	To assess the accuracy of virtual microscopy in terms of reliability and diagnostic reproducibility in colorectal cancer screening programmes.	Intra- and inter-observer agreements were substantially unmodified by the use of the virtual microscopy approach compared with traditional optical microscopy; moreover, for some histological features critical for patient management in colorectal cancer screening programm	Accuracy	Intra- and inter-observer agreement

			es (such as the presence of a villous component within the adenoma), virtual microscopy increased interobserver agreement (κ statistics 0.66 versus 0.52).

2017	UK	Ellison et al	A comparative study of germline BRCA1 and BRCA2 mutation screening methods in use in 20 European clinical diagnostic laboratories	NA	Quantitative	Non-randomised Comparative	20	Breast	Molecular Pathology	Analytic	Diagnostic accuracy	We evaluated the performance of different BRCA1/2 gene testing practices in routine diagnostic use in 20 European laboratories, with a focus on next-generation sequencing-based strategies	A small number of errors in the detection of pathogenic and significant variants were identified (2.6% diagnostic error rate). There was a high degree of concordance (>97%) across all laboratories for all variants detected. No systematic technical flaw was	Accuracy	Inter-laboratory agreement

													identified in the strategies employed across the participating laboratories.		
			Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union--an inter-laboratory study								To assess the variability of oestrogen receptor (ER) testing using immunocytochemistry, centrally stained and unstained slides from breast cancers were circulated to the members of the European Working Group for	There was almost complete concordance among readers (kappa=0.95) in ER-negative tumours on the stained slide and excellent concordance among readers (kappa=0.82) on the		
2004	UK	Wells et al		European Commission	Quantitative	Non-randomised Comparative	90	Breast	Histopathology	Analytic	Staining and preparation, Diagnostic accuracy			Completeness, Accuracy	Immunocytochemical staining protocol Inter-rater reliability for stain interpretation

Breast Screening Pathology, who were asked to report on both slides.	slides stained in each individual laboratory. Tumours showing strong positivity were reasonably well assessed (kappa=0.57 and 0.4, respectively), but there was less concordance in tumours with moderate and low levels of ER, especially when these were heterogen

													eous in their staining. More than 80% of the laboratories demonstrated ER positivity in the medium- and high-expressing tissue spots, but only about 43% succeeded with tissue spots with low expression. Poor interlaboratory agreement was based on		
2002	Germany	Von Wasielewski et al	Tissue array technology for testing interlaboratory and interobserver reproducibility of immunohistochemical estrogen receptor analysis in a large multicenter trial	NA	Mixed Method	Non-randomised Comparative	172	Breast	Histopathology	Pre-analytic, Analytic	Staining and preparation, Diagnostic accuracy	To compare immunohistochemical analysis in Germany across multiple centres. For interlaboratory testing, a new tissue microarray (TMA) system was chosen, enabling each of the 172 participating laboratories to analyze 30 tissue spots from the same specimen		Accuracy	Immunohistochemical staining protocol Interlaboratory and interobserver agreement

														insufficient retrieval efficacy as shown by additional tests using autoclave pretreatment		
												To assess the degree of interobserver variation between local pathologists and review pathology for pathological stage, including seminal vesicle invasion and extraprostatic extension, and for	There was a high concordance between review pathology and local pathologists existed for seminal vesicle invasion (94%, κ=0.83), agreement was much less for extraprost		Inter-rater concordance rate
			Impact of pathology review of stage and margin status of radical prostatectomy specimens (EORTC trial 22911)								Diagnostic accuracy, Patient outcomes and follow					Biochemical progression-free survival depending on surgical margin status
2006	Canada; Belgium	van der Kwast et al		EORTC	Quantitative	Prospective Observational	566	Prostate	Histopathology	Analytic, Post-analytic				Accuracy		

												surgical margin status	atic extension (57.5%, κ =0.33) and for surgical margin status There was substantial interlaboratory agreement in the interpretation of HER2/neu IHC			
2014	Canada	Sheffield et al	HER2/neu testing in gastric cancer by immunohistochemistry: assessment of interlaboratory variation	Canadian Institutes for Health Research	Qualitative	Non-randomised Comparative	45	Upper Gastrointestinal	Histopathology	Analytic	Staining and preparation, Diagnostic accuracy	To assess interlaboratory variation in IHC staining and interpretation across multiple laboratories	despite variability in protocols, including differences in IHC methodologies, interpretive errors, and intratumor	Completeness, Accuracy	Immunohistochemical staining protocol Inter-laboratory agreement Diagnostic accuracy

													use a multitumor tissue microarray to compare immunohistochemical staining results in our laboratory with published data and assess interlaboratory variability of immunohistochemical staining	al heterogeneity multiple-tumor microarrays are an efficient method for assessing the sensitivity and specificity of staining with any antibody used diagnostically. As a tool for quality assurance, they offer the advantage of taking into account local differences in tissue		
			Tissue microarrays are an effective quality assurance tool for diagnostic immunohistochemistry			Retrospective Observational					Diagnostic accuracy				Accuracy, Cost-effectiveness	Interlaboratory variability
2002	USA	Hsu et al		NA	Qualitative		351	Non-specific	Histopathology	Analytic						Cost analysis

													fixation, processing, and staining. They also allow cost-effective assessment of interlaboratory variability in immunohistochemical staining			
2020	Canada	Decoste et al	Comparing Pathology Report Quality Indicators in 2 Distinct Whipple Resection Specimen Protocols	Dalhousie University	Quantitative	Retrospective Observational	132	Upper Gastrointestinal	Histopathology	Post-analytic	Report quality	To determine if the introduction of axial sectioning resulted in more complete pathology reporting of pancreaticoduodenectomy specimens	There was higher-quality pathology reports since the introduction of axial sectioning. additional contributors include College of American	Completeness	Completeness of report using synoptic reporting protocol

													Pathologists protocols, increasing subspecialty practice, and updates to the American Joint Committee on Cancer staging criteria. There was a high level of agreement between pathologists for inadequate smears.			
1996	Italy	Ciatto et al	Interlaboratory reproducibility in reporting inadequate cervical smears--a multicentre multinational study	NA	Quantitative	Prospective Observational	300	Gynaecological	Cytopathology	Pre-analytic, Analytic	Sample Collection, Diagnostic accuracy	To interpret vaginal smears initially reported as inadequate and suboptimal	Suboptimal smears showed fair to good agreemen	Accuracy	Specimen quality Inter-rater agreement	

													t with the majority report only in five instances, whereas agreement was poor for eight cytologists		
2000	Canada	Scurry et al	Thin layer compared to direct smear in thyroid fine needle aspiration	NA	Quantitative	Retrospective Observational	728	Thyroid	Cytopathology	Pre-analytic, Analytic	Specimen handling and transport, Diagnostic accuracy	To compare the efficacy of thin layer versus direct smear methods in thyroid FNAC	Thin layer preparation showed a trend towards a higher proportion of true benign diagnoses (31% vs. 24%), a lower proportion of inadequate specimens (41% vs. 50%),	Completeness, Accuracy	Specimen handling protocols Diagnostic rate

													and a lower false negative rate (3% vs. 9%) compared to direct smear.		
			Evaluation of a standardized protocol for processing adrenal tumor samples: preparation for a European adrenal tumor bank								To evaluate standardized protocols for processing adrenal tumor samples to prepare for a European adrenal tumor bank	The study identified optimal protocols and endpoints for tissue handling that can be defined for a proposed European adrenal tumor biobank to ensure the quality of molecular analyses.		Specimen handling protocols Interlaboratory reproducibility and comparability of molecular endpoints
2010	Germany	Johnsen et al		NA	Quantitative	Non-randomised Comparative	NA	Adrenal	Molecular Pathology	Pre-analytic	Specimen handling and transport			Completeness	

2020	Taiwan	Wang et al	Good staining quality ensuring the reproducibility of Ki67 assessment	NA	Quantitative	Non-randomised Comparative	NA	Breast	Molecular Pathology	Analytic	Staining and preparation	To evaluate the agreement of Ki67 labeling index (LI) in clinical practice	Good Ki67 staining quality can minimize staining variations among different laboratories, which is essential for the reproducibility of Ki67 LI	Completeness	Immunohistochemical staining quality Reproducibility of Ki67 labeling index

2002	Canada	Parker et al	Assessment of Interlaboratory Variation in the Immunohistochemical Determination of Estrogen Receptor Status Using a Breast Cancer Tissue Microarray	NA	Quantitative	Retrospective Observational	NA	Breast	Molecular Pathology	Analytic	Diagnostic accuracy	To assess interlaboratory variation in the immunohistochemical determination of estrogen receptor status using a breast cancer tissue microarray	The study identified significant interlaboratory variation in estrogen receptor immunohistochemistry results, highlighting the need for standardized protocols and quality assurance measures to ensure consistent and accurate testing across laboratories	Accuracy	Interlaboratory concordance rates Reproducibility of immunohistochemical staining Adherence to standardized protocols

2002	Austria	nig et al	Quality assurance for detection of estrogen and progesterone receptors by immunohistochemistry in Austrian pathology laboratories	NA	Quantitative	Retrospective Observational	1000	Breast	Molecular Pathology	Analytic	Staining and preparation, Diagnostic accuary	To assess the quality of the immunohistochemical assays, and to assess the interlaboratory and interobserver variability performed by different laboratories in Austria.	The study identified significant variability in receptor detection accuracy among laboratories, highlighting the need for standardized protocols and quality assurance measures	Completeness, Accuracy	Immunohistochemical staining quality Inter-laboratory concordance rate

2011	UK	Deans et al	Molecular Genetics External Quality Assessment Pilot Scheme for KRAS Analysis in Metastatic Colorectal Cancer	NA	Quantitative	Audit	NA	Colorectal	Molecular Pathology	Analytic, Post-analytic	Diagnostic accuracy, Report quality	To evaluate the performance of laboratories in KRAS mutation testing for metastatic colorectal cancer and assess the need for standardized external quality assessment	The pilot EQA scheme revealed few genotyping errors among participating laboratories. However, there was considerable variation in the reporting nomenclature of genotyping results, highlighting the need for standardized reporting practices.	Accuracy	Accuracy of genotyping results Consistency in reporting nomenclature Inter- and intra-laboratory concordance

2011	USA	Shuch et al	Quality of pathological reporting for renal cell cancer: implications for systemic therapy, prognostication and surveillance	NA	Quantitative	Retrospective Observational	NA	Kidney	Histopathology	Post-analytic	Report quality	To evaluate whether current nephrectomy pathology reports are sufficient to allow clinicians to use prognostic nomograms, tailor surveillance, enroll patients into adjuvant trials, and select systemic therapy for renal cell carcinoma	The study found that many pathology reports lacked essential information, such as tumor size, grade, and necrosis, which are critical for prognostication and treatment planning. This deficiency hampers the ability to apply prognostic models and make informed clinical decisions	Completeness	Report completeness-inclusion of key pathological parameters (e.g., tumor size, grade, necrosis)

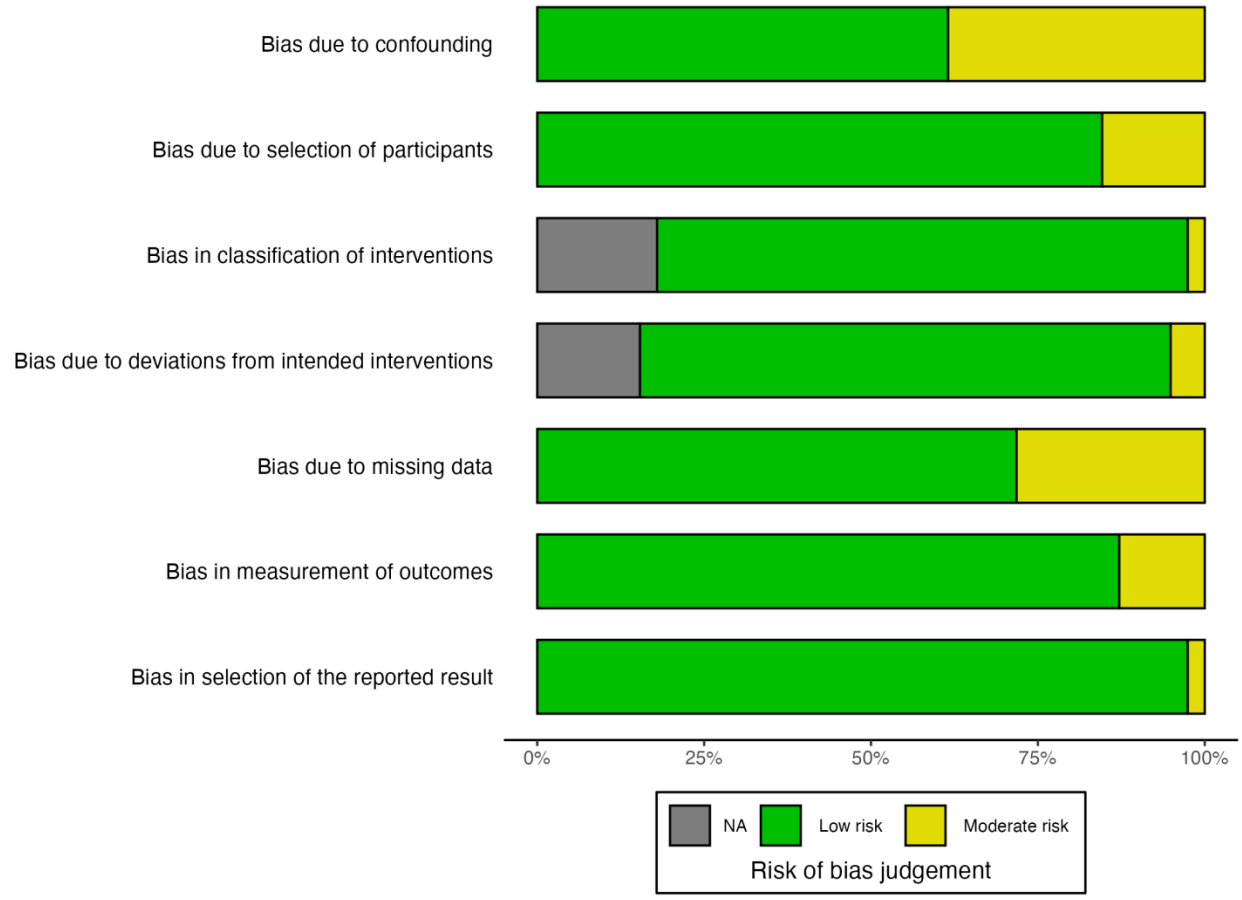
3. Appendix C – Supplementary Table 2

Supplementary Table 2. Risk of Bias of Studies using ROBIN-I. D1: Bias due to Confounding; D2: Bias due to Selection of Participants; D3: Bias in Classification of Intervention; D4: Bias due to Deviations from Intended Interventions; D5: Bias due to Missing Data; D6: Bias in Measurement of Outcomes; D7: Bias in Selection of Reported Results

Study	D1	D2	D3	D4	D5	D6	D7	Overall
Hecht et al.(12)	Low	Low	NA	NA	Low	Low	Low	**Low**
Ronco et al (13)	Low	Low	NA	NA	Moderate	Low	Low	**Moderate**
Lexander et al (14)	Low	Low	Low	Low	Low	Low	Low	**Low**
Korach et al (15)	Moderate	Low	NA	NA	Moderate	Moderate	Low	**Moderate**
Stanczuk et al (16)	Low	Low	NA	NA	Low	Low	Low	**Low**
Diaz et al (17)	Low	Low	NA	NA	Low	Low	Low	**Low**
Stabile et al (18)	Moderate	Moderate	Low	Low	Low	Moderate	Low	**Moderate**
Dillner et al (19)	Moderate	Moderate	Low	Low	Moderate	Low	Low	**Moderate**
Fitzgibbons et al (20)	Low	Moderate	Low	Low	Low	Low	Low	**Low**
Gök et al (21)	Moderate	Moderate	Low	Low	Low	Low	Low	**Moderate**

Ronco et al (22)	Moderate	Low	Low	Low	Moderate	Low	Low	**Moderate**
Anania et al(23)	Moderate	Low	Low	Low	Moderate	Low	Low	**Moderate**
Long et al(24)	Low	Low	Low	Low	Low	Low	Low	**Low**
Harrison et al(25)	Low	Low	Low	Low	Moderate	Low	Low	**Moderate**
Youl et al (26)	Low	Low	Low	Low	Moderate	Low	Low	**Moderate**
Kottke et al (27)	Low	Low	Low	Low	Moderate	Low	Low	**Moderate**
Orumaa et al (28)	Low	Low	Low	Low	Moderate	Low	Low	**Moderate**
Deans et al (29)	Low	Low	Low	Low	Moderate	Low	Low	**Moderate**
Risio et al (30)	Low	Low	Low	Low	Low	Low	Low	**Low**
Ellison et al (31)	Low	Low	Low	Low	Low	Low	Low	**Low**
Wells et al (32)	Low	Low	Low	Low	Low	Low	Low	**Low**
Von Wasielewski et al (33)	Low	Low	Low	Low	Low	Low	Low	**Low**
van der Kwast et al (34)	Moderate	Low	Low	Low	Low	Low	Low	**Moderate**
Sheffield et al (35)	Low	Low	Low	Low	Low	Low	Low	**Low**

Hsu et al (36)	Low	Low	Low	Low	Low	Low	Low	**Low**
Ciatto et al (37)	Low	Low	Low	Low	Low	Low	Low	**Low**
Decoste et al(38)	Moderate	Low	Low	Low	Low	Low	Low	**Moderate**
Scurry et al (39)	Moderate	Low	Low	Low	Low	Low	Low	**Moderate**
Johnsen et al (40)	Low	Moderate	Low	Low	Moderate	Low	Low	**Moderate**
Wang et al(41)	Moderate	Low	Low	Low	Low	Low	Low	**Moderate**
Parker et al (42)	Low	Low	Low	Moderate	Low	Low	Low	**Moderate**
Regitnig et al (43)	Moderate	Low	Moderate	Moderate	Low	Moderate	Low	**Moderate**
Shuch et al (44)	Low	Low	NA	NA	Low	Low	Low	**Low**
Elmore et al (45)	Low	Low	NA	Low	Low	Moderate	Low	**Moderate**
Hwang et al (46)	Moderate	Moderate	Low	Low	Low	Low	Moderate	**Moderate**
McFadyen et al (47)	Moderate	Low	Low	Low	Low	Low	Low	**Moderate**
Wang et al (48)	Moderate	Low	Low	Low	Low	Low	Low	**Moderate**
Thunnissen et al (49)	Moderate	Low	Low	Low	Low	Moderate	Low	**Moderate**
Deans et al (50)	Low	Low	Low	Low	Low	Low	Low	Low

4. Appendix D – Supplementary Figure 1

Supplementary Figure 1. Bar Chart of the Risk of Bias Assessment by Domain