Legal Medicine 84 (2026) 102894

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Toward methodological standardization in forensic immunohistochemistry: a critical appraisal of 144 post-mortem studies and a proposed evaluative framework

Luca Tomassini [a,*], Martina Onofri [b], Cristiana Gambelunghe [c], Piergiorgio Fedeli [a], Roberto Scendoni [d], Niccolò Pini [e], Massimo Lancia [c]

[a] School of Law, Legal Medicine, University of Camerino, Camerino, Italy
[b] Section of Legal Medicine, Department of Medicine and Surgery, University of Perugia, Terni, Italy
[c] Forensic Medicine, Forensic Science and Sports Medicine Section, Department of Medicine and Surgery, University of Perugia, Perugia, Italy
[d] Department of Law, Institute of Legal Medicine, University of Macerata, Macerata, Italy
[e] Section of Legal Medicine and Insurance Medicine, Department of Biomedical Sciences for Health, University of Milan, Milan, Italy

ARTICLE INFO

Keywords:
Immunohistochemistry
Forensic pathology
Post-mortem examination
Methodological standardization
Evidentiary reliability
Medicolegal autopsy

ABSTRACT

Immunohistochemistry (IHC) is widely applied in post-mortem forensic investigations. Its evidentiary value depends critically on methodological standardization. Following PRISMA guidelines and PROSPERO registration (CRD420251063965), we systematically searched PubMed, Scopus, and Web of Science for original IHC studies on human autopsy material. A total of 144 studies were included and evaluated against six predefined methodological criteria (M1–M6) addressing slide interpretation metrics, multi-reader assessment, observer blinding, concordance reporting, quality control, and predefined positivity thresholds.

IHC was applied across diverse forensic domains, most frequently in traumatic brain injury (16%), sudden cardiac death (14%), COVID-19-related death (13%), and wound vitality/wound dating (11%). Quantitative or semi-quantitative slide reading (M1) was adopted in 75.7% of studies, and quality control measures (M5) in 52.8%. By contrast, blinded reading (M3) was implemented in only 40% of applicable studies, inter-observer concordance (M4) in 10%, and predefined positivity thresholds (M6) in only 4.2% of studies. Only 38.9% of studies met three or more methodological criteria, and a single study satisfied all six criteria.

Forensic IHC demonstrates broad application and considerable inferential ambition, but its methodological foundations remain inconsistently consolidated. We propose that future forensic IHC studies place greater emphasis on analytical validation, standardized interpretation criteria, assessment of observer agreement, and systematic reporting of relevant pre-analytical variables in order to improve reproducibility and comparability.

1. Introduction

Immunohistochemistry (IHC) has represented a well-established tool in anatomical pathology and biomedical research for several decades, owing to its ability to integrate conventional morphological analysis with molecular and functional information obtained directly within the tissue context [1–4]. The ability to localize specific antigens within the histological architecture has established IHC as a reference technique in numerous clinical fields, particularly in oncology, inflammatory and neurodegenerative diseases, and in the investigation of cellular mechanisms of injury and stress responses [3,5,6].

In the medicolegal setting, IHC has been progressively introduced as a supportive tool in forensic autopsy practice, particularly in cases in which macroscopic examination and conventional histology prove insufficient to provide unequivocal answers [7,8]. In this context, IHC has been employed to address forensic questions of high interpretative complexity, including the assessment of lesion vitality, the reconstruction of perimortem events, the distinction between vital and post-mortem injuries, the interpretation of asphyxial and traumatic mechanisms, and, more recently, the support of the diagnosis of specific causes of death, including cardiovascular and metabolic conditions [9–13]. It is acknowledged that the frequency of IHC application in forensic autopsy

* Corresponding author at: School of Law, Legal Medicine, University of Camerino, Via del Bastione 1, 62032 Camerino (MC), Italy.
E-mail addresses: luca.tomassini@unicam.it (L. Tomassini), cristiana.gambelunghe@unipg.it (C. Gambelunghe), piergiorgio.fedeli@unicam.it (P. Fedeli), r.scendoni@unimc.it (R. Scendoni), niccolo.pini@unimi.it (N. Pini), massimo.lancia@unipg.it (M. Lancia).

https://doi.org/10.1016/j.legalmed.2026.102894
Received 28 May 2026; Received in revised form 16 June 2026; Accepted 16 June 2026
Available online 18 June 2026

practice varies considerably across institutions and jurisdictions; in many settings, histological examination is not routinely performed and immunohistochemical analysis remains an exceptional rather than standard investigative tool [14–16]. The present review does not characterize casework frequency but evaluates the methodological quality of the published research on which any such use is scientifically grounded.

Despite the quantitative expansion of the literature on the use of IHC in the post-mortem forensic setting, substantial uncertainty persists regarding the actual scientific and evidentiary value of these applications. Although results are often presented as promising, many studies exhibit marked heterogeneity in study design, case and control selection, management of pre- and post-mortem variables, and, most notably, in the methods used for reading and interpreting immunohistochemical preparations [15–17]. Such methodological fragmentation has repeatedly been identified as one of the main limitations to the reproducibility of findings and to their transferability across different forensic contexts [18,19].

In the medicolegal field, the value of a technique cannot be assessed solely on the basis of the biological plausibility of the marker or the statistical significance achieved within a single study. Rather, it depends critically on the presence of standardized protocols, systematic control of confounding variables, transparency of interpretative procedures, and the ability to replicate findings in independent contexts [15,18,19]. In the absence of these requirements, there is a risk of attributing forensic significance to immunohistochemical signals that reflect nonspecific phenomena, post-mortem changes, or observational bias, rather than biological processes genuinely related to the medicolegal question [20,21].

In light of these critical issues, the present study was conceived to address a predominantly qualitative gap in the existing literature. The aim was not to identify individual immunohistochemical markers as "valid" or "invalid" in absolute terms, nor to propose a hierarchy of forensic biomarkers, but rather to systematically assess the methodological robustness of published studies and their ability to provide reliable forensic evidence. In particular, the analysis was focused on evaluating the degree of standardization of immunohistochemical techniques, the methods used for slide reading and interpretation, the management of pre-analytical and post-mortem variables, and the adoption of measures aimed at reducing observational bias and ensuring reproducibility of results. A review of this nature is broad in scope by design: restricting the analysis to a single stain, tissue, or forensic scenario would yield local quality metrics while missing the systemic patterns in reporting practice that constitute the principal finding of interest. The six methodological criteria (M1–M6) apply transversally across stain types and forensic contexts, making the breadth of inclusion an epistemic requirement rather than a limitation of the analysis.

On this basis, the present work seeks to answer the following question: *to what extent are immunohistochemical techniques applied to human post-mortem tissues, as reported in the recent forensic literature, methodologically standardized and capable of providing reliable, reproducible, and defensible evidence in the medicolegal context?*

Legal frameworks for the admissibility of scientific evidence—including the Daubert [22–26]—require testable methodology, known error rates, and demonstrated reproducibility. These requirements apply to forensic IHC research as a body of scientific evidence: what determines evidentiary weight is the quality of the research, not the name of the marker. A stain with a well-established clinical profile does not automatically inherit that validation when applied to post-mortem forensic tissue, where autolysis, variable fixation times, and cause-of-death heterogeneity introduce sources of variability that require explicit methodological control and documentation. This is precisely the dimension that M1–M6 are designed to assess.

Through a systematic review oriented toward methodological quality rather than biological outcomes per se, this study aims to provide a critical overview of the current state of forensic immunohistochemistry, identifying its main strengths, recurring structural limitations, and the areas in which further efforts toward standardization are required to enable a truly probative use of the technique.

2. Materials and methods

The present study was conducted as a systematic review with qualitative synthesis and methodological appraisal of the included studies. This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines and has been registered with PROSPERO (https://www.crd.york.ac.uk/PROSPERO/view/CRD420251063965). A systematic literature search was conducted using PubMed, Scopus, and Web of Science to retrieve studies published in English on immunohistochemical techniques applied to forensic autopsy cases. The aim of this review was to assess the actual scientific value of autopsy-based studies in supporting the use of immunohistochemistry in forensic settings through a targeted evaluation of each study's focus, the distribution of the employed markers, the context of application, and the methodological robustness of the articles—that is, their capacity to substantiate or refute the use of immunohistochemistry in specific forensic scenarios.

Studies published within the time frame from May 20, 2015, to May 20, 2025, were considered eligible for inclusion.

To account for the heterogeneity of terminology used in the literature, a broad free-text search strategy was adopted using the following terms:

- **Pubmed**: ((((((((((((((cadaver[MeSH Terms]) OR (cadaver)) OR (postmortem)) OR (autopsy[MeSH Terms])) OR (autopsy)) OR (dead bodies)) OR (deceased)) OR (death[MeSH Terms])) OR (death)) AND (forensic sciences[MeSH Terms])) OR (forensic)) OR (forensics)) OR (medicolegal)) AND (immunohistochemistry[MeSH Terms])) OR (immunohistochemistry)) OR (immunohistochemical)
- **Scopus**: (TITLE-ABS KEY(cadaver OR postmortem OR autopsy OR "dead bodies" OR deceased OR death) AND TITLE-ABS-KEY ("forensic science" OR forensic OR forensics OR medicolegal) AND TITLE-ABS-KEY(immunohistochemistry OR immunohistochemical))
- **Web of Science**: ((((((((((TS = (cadaver)) OR TS = (postmortem)) OR TS = (autopsy)) OR TS = (dead bodies)) OR TS = (deceased)) OR TS = (death)) AND TS = (forensic science)) OR TS = (forensic)) OR TS = (forensics)) OR TS = (medicolegal)) AND TS = (immunohistochemistry)) OR TS = (immunohistochemical)))

The selection process followed the PRISMA guidelines. In the initial screening phase, titles and abstracts were evaluated in Rayyan, a free web and mobile application that facilitates semi-automated screening (http://rayyan.qcri.org) [27,28]. Two independent reviewers assessed each record according to predefined inclusion and exclusion criteria, and any discrepancies were resolved through discussion and consensus. Inter-reviewer agreement at the full-text eligibility stage was quantified using Cohen's kappa (κ) statistic with 95% confidence intervals. The articles that passed this stage underwent a second, more detailed screening. In this phase, titles, abstracts, methods, and keywords were re-evaluated to confirm relevance. Simultaneously, the search was complemented by a manual review of the reference lists to identify additional eligible studies. Finally, the selection of studies based on full-text evaluation was carried out by two additional independent reviewers.

The publications finally selected for analysis had to meet the following inclusion and exclusion criteria:

Inclusion criteria:

- Post-mortem investigations;
- Studies with forensic relevance;
- Studies on human tissues;
- Studies involving immunohistochemistry;
- Original research articles.

Exclusion criteria:

- Studies conducted on living subjects;
- Case reports and case series (low level of evidence);
- Studies on diagnoses not relevant to forensic interest (e.g., post-mortem diagnosis of neoplasia);
- Systematic reviews and meta-analyses;
- Letters or other non-research article types.

For each article included in the analysis, the following data were extracted and recorded:

General information

Author and year of publication; DOI/PMID/WOS identifier; Country of affiliation; Type of study; Field of application; Specific macroarea; IHC use.

Study design

Research question; Number of cases; Number of controls (if present); Inclusion and exclusion criteria specified for cases (yes/no); Inclusion of decomposed cadavers (yes/no).

Subject Information and postmortem variables.

Cause of death reported for cases (yes/no); Cause of death reported for controls (yes/no); Comorbidities reported for cases (yes/no); Comorbidities reported for controls (yes/no); Time of death clarified for cases (yes/no); Time of death clarified for controls (yes/no); Postmortem interval (time between death and examination) reported for cases (yes/no); Postmortem interval reported for controls (yes/no); Body preservation conditions reported for cases (yes/no); Body preservation conditions reported for controls (yes/no); Confounding factors clarified for cases (yes/no); Confounding factors clarified for controls (yes/no).

Technical and Methodological Aspects.

Immunohistochemical markers analyzed; Type of primary antibodies used; Method of antibody detection (direct/indirect) Slide interpretation criteria; Tissues examined.

Regarding the thematic macro-areas, the included studies were categorized according to the main field of forensic application of immunohistochemistry. The following areas were identified:

- Acute Kidney Injury (AKI)
- Acute Respiratory Distress Syndrome (ARDS)
- Asphyxial Death
- COVID-19–related death
- Drowning
- Hypothermia/Hyperthermia
- Post-Mortem Interval (PMI) Estimation
- Sudden Cardiac Death (SCD)
- Sudden Infant Death Syndrome (SIDS)/Sudden Unexpected Infant Death (SUID)
- Substance-Related Death
- Suicide
- Traumatic Brain Injury (TBI)
- Vitality / Wound Dating
- Other

Studies classified under the “Other” category included all cases of death that did not fit into any of the aforementioned categories.

Given the heterogeneity of study designs, analytical methods, and outcome measures among the included articles, a quantitative meta-analysis was not feasible. Therefore, a descriptive statistical analysis was conducted on the selected variables from the 144 papers included herein. Results were summarized using qualitative and descriptive approaches, and frequency distributions were used to describe the occurrence of methodological characteristics, types of markers, and forensic application areas.

To assess the replicability and reproducibility of the HIC protocols employed in the studies, data were stratified according to two key categorical dimensions: field of application and cause of death. Data visualization techniques were employed to illustrate distribution patterns and infer methodological trends.

To enhance interpretability and methodological rigor, six classification parameters (M1–M6) were defined and systematically applied across all included studies. These criteria assess research quality independently of the specific marker or forensic application domain: the established clinical use of an antibody in a diagnostic setting does not substitute for rigorous documentation of its performance in post-mortem forensic tissue, where autolysis, fixation variability, and cause-of-death heterogeneity require the same methodological controls regardless of whether the stain is novel or widely adopted.

M1 – Slide Reading Criteria: Slides were read according to an explicitly declared quantitative or semi-quantitative scoring system (e.g., H-score, Allred score, IRS, 0–3/0–4 scales, +/++/+++, percentage-positive cells, counts/HPF/ROI, or digital image-analysis metrics). M1 was intentionally conceived as a binary criterion assessing the presence or absence of a predefined slide-reading framework rather than the degree of objectivity of the adopted scoring method. The specific type of scoring system employed was recorded separately and analyzed descriptively, as different approaches may substantially differ in reproducibility despite all providing an explicit interpretative framework.

M2 – Diagnostic Confirmation: Slide interpretation was conducted by $\geq$2 independent readers or via digital analysis tools, with declared parameters.

M3 – Reader Blinding: Studies were assessed for the implementation of blinded or masked reading protocols to minimize observer bias.

M4 – Reported Concordance: The presence of inter- or intra-observer agreement metrics was documented, including statistical measures such as Cohen's kappa (κ), intraclass correlation coefficient (ICC), or qualitative descriptors of concordance.

M5 – Quality Assurance and Control Measures: Explicit use of internal controls, including positive and negative control samples, implementation of quality control protocols, and procedures for managing technical artefacts and batch variability.

M6 – Defined Positivity Thresholds: Clear specification of pre-established cut-off values used to determine marker positivity, ensuring reproducibility and consistency in immunohistochemical interpretation This criterion was considered applicable only to studies that operationalized a dichotomous positivity or diagnostic decision. Studies reporting continuous, semi-quantitative, or purely descriptive expression patterns without a binary positivity call were considered not applicable for M6. Accordingly, non-fulfilment of M6 was interpreted as the absence of a pre-specified interpretative threshold where such a threshold was required, and not as evidence of poor methodological quality per se.

When applicable, trends were identified through comparative assessment of study features (e.g., presence of controls, statistical significance, and evidence of improved discriminatory power), allowing evaluation of methodological transparency and reproducibility across the body of evidence.

Data recorded for the selected papers and the corresponding M1–M6 assessments are available in Supplementary Table S1, which provides the complete study-level dataset for all 144 included studies.

To complement the M1–M6 methodological framework, a secondary quality appraisal was performed using the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies-2) as described by Whiting et al. (2011) [29]. Unlike the M1–M6 criteria, which were specifically developed to evaluate reporting transparency, interpretative reproducibility, and methodological standardization of forensic immunohistochemical investigations, QUADAS-2 is a validated instrument designed to assess risk of bias in diagnostic research [29].

The two approaches were therefore considered complementary rather than overlapping. While M1–M6 focus on how immunohistochemical evidence is generated, interpreted, and reported, QUADAS-2 evaluates potential sources of systematic bias related to study design,

case selection, reference standards, and analytical procedures.

Each included study was preliminarily classified according to its methodological structure. Studies presenting a diagnostic-comparative design were assessed using QUADAS-2 domains adapted to the forensic setting. Studies lacking a diagnostic-comparative structure were considered not suitable for formal QUADAS-2 scoring. Exploratory studies incorporating a diagnostic-comparative design were assessed when the QUADAS-2 domains were deemed applicable. The complete QUADAS-2 assessment is reported in Supplementary Table S2.

3. Results

Overall, 32591 records were identified through database searching. After removal of 897 duplicate records, 31694 records underwent title and abstract screening, of which 31,471 were excluded as not relevant to the aims of the review. A total of 223 reports were assessed for eligibility in full text. Following full-text evaluation, 80 reports were excluded for not meeting the inclusion criteria, resulting in 143 studies identified through database screening. One additional eligible study was identified through manual screening of reference lists and fulfilled the inclusion criteria. Consequently, a total of 144 studies were included in the final systematic review. The study selection process is summarized in Fig. 1.

At the full-text eligibility stage ($n = 223$ reports), the two reviewers achieved an overall agreement of 86.5%, corresponding to substantial inter-reviewer agreement (Cohen's $\kappa = 0.71$, 95% CI 0.62–0.81). Thirty discordant reports were resolved through consensus discussion.

The geographical distribution of the included studies, based on the corresponding author's institutional affiliation, was as follows: 38 studied from Italy, 17 from Germany, 16 from Japan, 13 studies from both China and the USA, 8 studies from Romania, 4 studies from Australia, 3 studies each from India and Poland, 2 studies from Denmark, England, Finland, France, and Russia, 1 study each for Austria, Brazil, Croatia, Greece, Iraq, Mexico, Serbia, Slovakia, Spain, Switzerland, Taiwan, Tunisia, and the UK. Lastly, 6 studies were the result of international collaborations.

With respect to study design, 82 were case series (56.9%), 34 experimental/translational studies (23.6%), 20 cohort studies (13.9%), 5 case–control studies (3.5%), and 3 cross-sectional studies (2.1%). All of the studies explicitly stated the research question.

Regarding the antibody detection methodology, all studies (100%) employed an indirect immunohistochemical technique. With respect to the type of primary antibody used, the vast majority of studies employed mouse monoclonal antibodies (93 studies, 64.58%). Rabbit polyclonal antibodies are the second most represented category (77 studies, 53.47%), while rabbit monoclonal antibodies were used in 27 studies (18.75%). 6 studies reported using goat-derived polyclonal antibodies (4.17%); sheep polyclonal antibodies were described in 2 studies (1.39%). Chicken monoclonal, guinea pig polyclonal, hamster monoclonal, and mouse polyclonal antibodies were employed in 1 study each. Lastly, 14 studies (9.72%) did not provide any detailed information on the type of primary antibody used. It should be noted that these categories are not mutually exclusive, as individual studies may include multiple types of antibodies (e.g., mouse monoclonal and rabbit polyclonal Abs). Consequently, the sum of the percentages exceeds 100%. The reported values therefore, represent frequencies of mention, not exclusive proportions. In fact, 60 studies (41.67%) used more than one type of antibody. Only the 20.83% of studies ($n = 30$) used antibodies of the same clonality and derived from the same host species aimed at different markers. Conversely, the 27.78% of studies ($n = 40$) used a single antibody to detect a single target.

Regarding the IHC markers investigated, the most frequently reported markers were CD68 (21 studies, 14.58%), GFAP (15, 10.41%), CD3 (12, 8.33%), and βAPP (11, 7.64%). IL-6 and p-tau were both targeted in 8 studies each (5.56%), while CD45, SARS-CoV-2 spike (S) protein, and tryptase were investigated in 7 studies each (4.86%). CD20, CD34, CD61, and HSP70 appeared in 6 studies (4.17%), followed by CD4, CD8, CD15, HIF-1- α, IBA1, IL-15, and SARS-CoV-2 nucleocapsid (N) protein (each in 5 studies, 3.47%). Cx43, FN, IL-1β, MB, MBP, TMEM119, and VIM (each in 4 studies, 2.78%).

Markers appearing in three studies (2.08%) included AQP3, C5b-9, HSP27, IL-8, MCP-1, MPO, NeuN, S100, TH, TNF-α. αSMA, and iNOS. Those cited in two studies (1.39%) were 8-OHdG, ACE2, ACT, AE1/AE3, AQP1, AQP4, AQP5, CASP3, CCR2, CD117, CD138, CD163, CD1a, CD31, CMV, Casp9, DES, DMD, FOXO3, fibrinogen, GPA, IL-10, KI67, MHCII, NOX2, NSE, NT, Ng, Nrf2, P-selectin, SIRT1, SP-A, TTF1, TUNEL, Ub, α7-nAChR, p40, and p53. The remaining markers were reported in only one study each (0.69%), encompassing a highly diverse array of targets: 4-HNE, 6-MAM, AANF, Aβ42, ADFP, ALB, α-syn, ANF, AQP2, AR, AT-8, ATTR, AVP, BCG, BDNF, β2-nAChR, BMP4, BNP, C4d, C9, Casp3, CCC9, CD2, CD19, CD45-RO, CD56, CD62P, CD206, c-fos, CgA, CIRBP, CK, CK7, CKAE1AE3, CML, COL1, COL3, COL4, COX5B, CRMP4, CT, cTN, CTSD, CV, DCTN1, DENV, DNAH9, EBV, eNOS, ERG, FVIII, Fb, FcεRIα, ferritin, fsTnl, GAP-43, GCG, GDNF, Gephyrin, GLUT-1, GPC, GPX1, H1N1, HO-1, HPS, HSP60, HSP90, HSV-1, HSV-2, Iba1, IgE, ICAM-1, IL-18, IL-33, insulin, ISN, JunB, KCNAB1, KCNJ3, KIF5B, KL1, Leptospira spp. antigen, LFB, LOX-1, MAP2c, MCT, MENK, MMP2, MMP9, MRP-8, MYH6, *N. meningitidis* antigen, NF, NF-H, NOS2, Nsp3, Olig-2, P. jirovecii antigen, PCNA, PDGFR-α, PDGFR-β, PG, PSA-NCAM, perforin, PG, protamine, PSA-NCAM, pT181, pT205, pT217, pT231, RAGE, RBM3, renin, Rickettsia spp. antigen, RSV antigen, *S. mansoni* antigen, SARS-CoV glycoprotein and SARS-CoV-2 S2 subunit of the Spike protein, SF3B3, σ-1R, SOCS-1, SOD2, SP, T. gondii antigen, TG, TLR4, TLR9, TMPRSS2, Tn, TNC, TNNT2, TPH, TPH2, TTR, V2R, VCAM-1, VCP, VDR, VP1, YFV, and ZO-1. A summary of the 20 most frequently utilized markers and their recurrence across studies is presented in Fig. 2, illustrating the predominance of inflammatory, neuronal, and vascular targets in the immunohistochemical landscape of forensic research.

The analysis of the anatomical distribution of sampled tissues across the 144 included studies revealed a marked predominance of neural and cardiac specimens. Specifically, nervous tissue was investigated in 59 studies (40.97%), followed by cardiac tissue in 40 (27.78%), pulmonary tissue in 26 (18.06%), cutaneous tissue in 19 (13.19%), renal tissue in 15 (10.42%), hepatic tissue in 10 (6.94%), and splenic tissue and pancreatic tissue in 3 each (2.08%). Bone marrow, cerebrospinal fluid (CSF), gingival tissue, placenta, and more broadly-defined soft tissues each appeared in 2 studies (1.39%).

Additional matrices—each reported in a single study (0.7%)—included adipose tissue, antemortem thrombi and postmortem clots, aorta and coronary arteries, blood, intestine and, more specifically, small intestine, larynx, mandibular salivary glands, olfactory and respiratory mucosae, pericardial fluid, skeletal muscle, thymus, tracheal glands, the umbilical cord, urine, and vascular smooth muscle cells. It should be noted that these categories are not mutually exclusive, as individual studies may have investigated multiple matrices; therefore, the reported values represent frequencies of mention, not exclusive proportions. In fact, 26 studies (18.06%) conducted analyses across two or more tissues.

In all studies analyzed (100%), the formalin fixation and paraffin embedding (Formalin-Fixed Paraffin-Embedded, FFPE) method was used for tissue preservation.

Case and control characteristics

Across the 144 studies, the number of analyzed cases per study showed a mean of 47.3, a standard deviation of 64.25, and a median of 29, with a minimum of 4 and a maximum of 622 cases.

For the subset of studies that included control groups, the number of controls per study had a mean of 24.24, a standard deviation of 25.21, and a median of 16.5, ranging from 1 to 162 controls.

Out of the 144 included studies, 138 (95.83%) explicitly reported the adoption of inclusion criteria for the examined cases, while 6 (4.16%) did not. Regarding exclusion criteria, 90 studies (62.5%) specified them, whereas 54 (37.5%) made no mention of exclusion parameters.

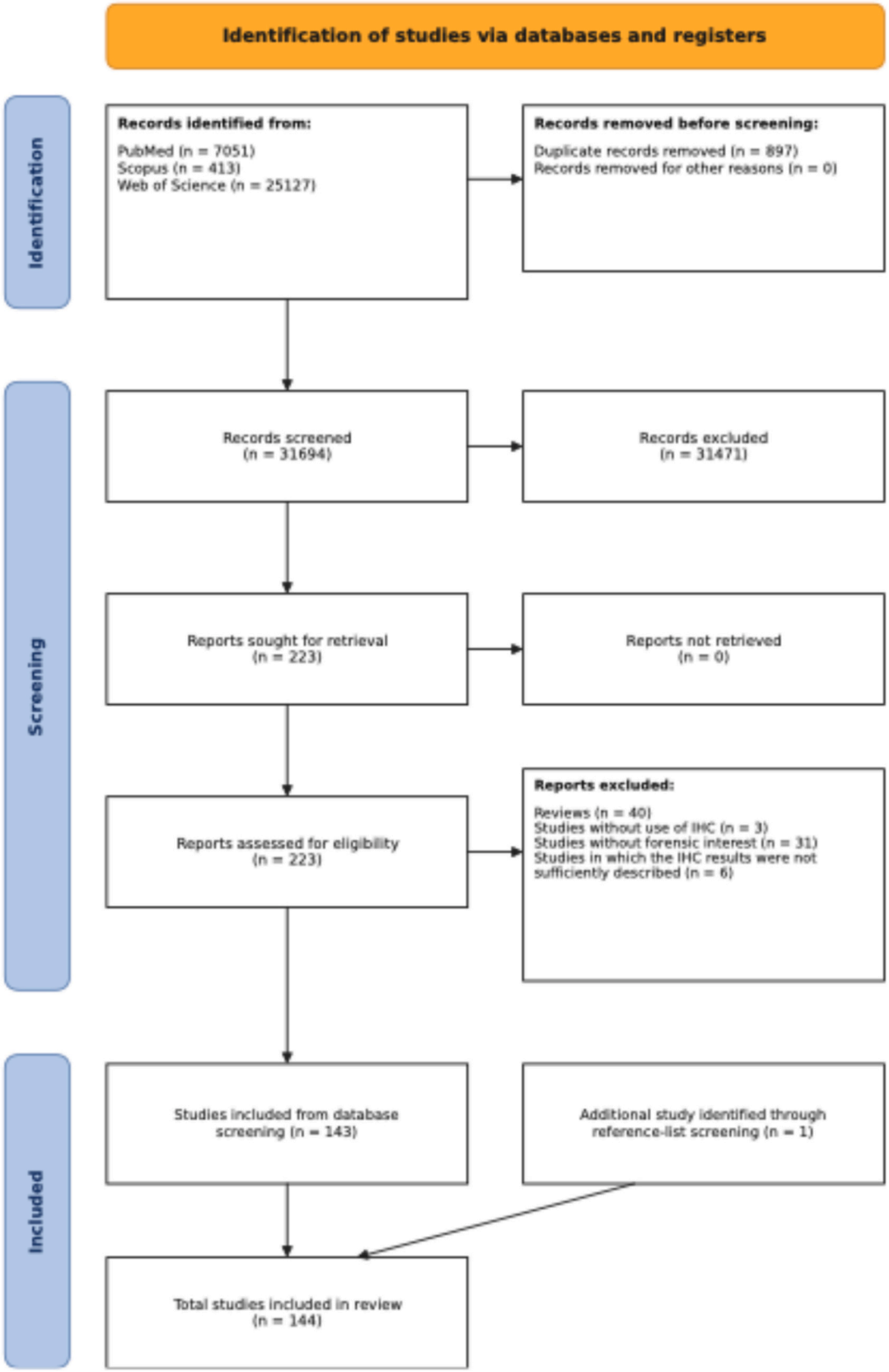

Fig. 1. PRISMA 2020 flow diagram for new systematic reviews, which included searches of databases and registers only. Source: Page MJ, et al. BMJ 2021;372:n71. doi:https://doi.org/10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

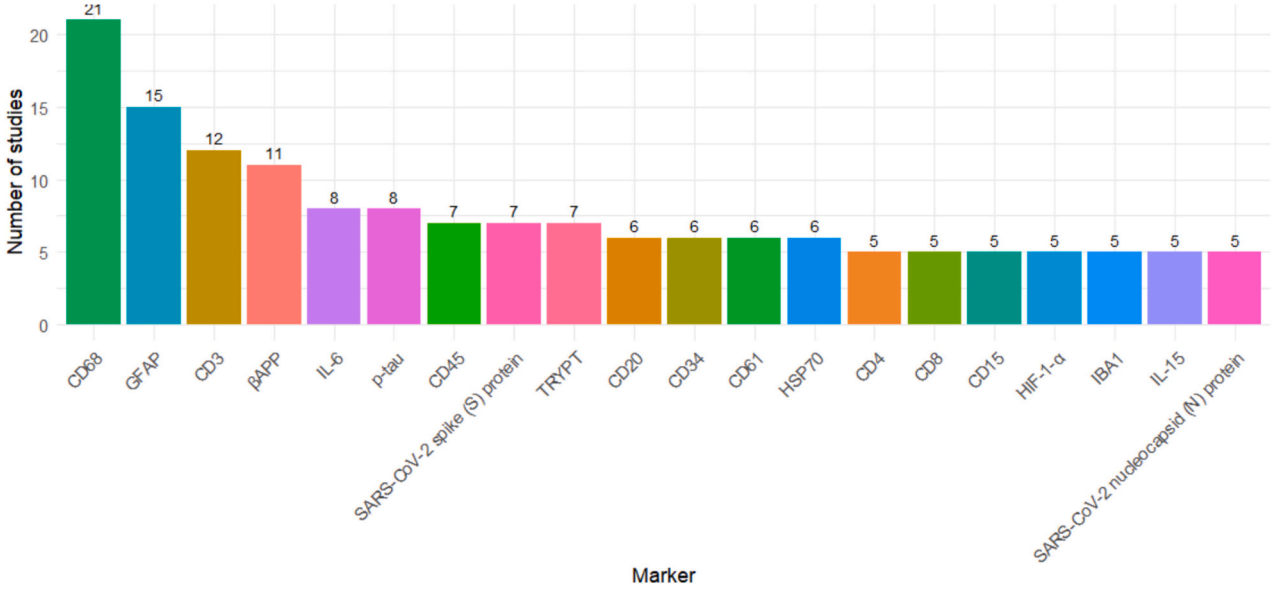

Fig. 2. The 20 most investigated markers in the selected studies and their relative frequency of mention.

With respect to the condition of the examined populations, 14 studies (9.72%) explicitly included cadavers showing signs of putrefaction, while the remaining 130 (90.28%) did not list putrefaction among their exclusion criteria.

Although studies conducted on living subjects were excluded from the review, 140 out of 144 studies explicitly stated the exclusion of living individuals from their study populations. In the few instances where the presence of living subjects was reported, these exclusively referred to control samples, not to the main study cohorts.

Information on the cause of death of the analyzed cases was reported and explicitly specified in 129 studies (89.58%), while it was not mentioned in the remaining 15 studies (10.42%). Regarding controls, the cause of death was provided in 87 studies (60.41%), not specified in 14 studies (9.72%), and not applicable in 43 studies (29.86%) due to the absence of control groups.

With respect to comorbidities, these were reported for the case populations in 76 studies (52.78%) and not specified in 68 studies (47.22%). For the control populations, comorbidities were described in 43 studies (29.86%) and omitted in 59 studies (40.97%), while 43 studies (29.86%) had no controls.

The determination of the time of death in case populations was classified as follows:

- Certain in 28 studies (19.44%)
- Estimated in 53 studies (36.81%)
- Not reported in 61 studies (42.36%)
- Two studies presented mixed reporting patterns:
- Morita S. (2015): partially certain and partially estimated;
- Reggiani Bonetti L. (2015): partially estimated and partially unreported.

For control populations, the time of death was:

- Certain in 15 studies (10.41%)
- Estimated in 33 studies (22.91%)
- Not reported in 53 studies (36.81%)
- Not applicable in 43 studies (29.86%)

A certain time of death was defined as one confirmed in a hospital setting or supported by unequivocal circumstantial evidence. An estimated time of death referred to determinations based on inferential or instrumental parameters (e.g., body temperature or other postmortem indicators).

The postmortem interval (PMI) was reported for case samples in 101 studies (70.13%) and not reported in 43 studies (29.86%). For control samples, the PMI was specified in 67 studies (46.53%), unreported in 34 studies (23.61%), and not applicable in the remaining 43 studies (29.86%) without controls.

Regarding body preservation conditions, these were described for case samples in 49 studies (34.03%), whereas 95 studies (65.97%) did not provide this information. Among the control samples, preservation conditions were reported in 30 studies (20.83%), omitted in 71 studies (49.31%), and not applicable in 43 studies (29.86%) due to the absence of controls.

Finally, the analysis of potential confounding factors was addressed in 128 studies (88.89%), which explicitly described the identification, management, or statistical control of confounding variables such as age, sex, postmortem interval, fixation time, tissue type, and comorbidities. The remaining 16 studies (11.11%) either did not account for confounding factors or did so only partially—for instance, by acknowledging their existence without detailing the corrective measures or adjustments employed in the analytical process.

Thematic classification

Regarding the thematic macro-areas of forensic application, studies were distributed as follows: 1 study (0.69%) was classified under “Acute Kidney Injury (AKI)”, 2 studies (1.39%) were classified as “Acute Respiratory Distress Syndrome (ARDS), 7 studies (4.86%) under ‘Asphyxia’, 19 studies (13.19%) were categorized as ‘COVID-19-related death’, 4 papers (2.78%) were classified as ‘Drowning’, 3 (2.08%) as ‘Hypothermia/Hyperthermia’, 20 studies (13.89%) as ‘Sudden Cardiac Death (SCD)’,8 studies (5.56%) as ‘Sudden Infant Death Syndrome/Sudden Unexpected Infant Death (SIDS/SUID)’, 7 papers (4.86%) investigated ‘Substance-related death’, 3 studies (2.08%) were classified in the ‘Suicide’ category”, and 23 studies (15.97%) investigated “Traumatic Brain Injury”. 5 studies (3.47%) focused on the use of IHC for “Postmortem Interval” estimation, while 16 papers (11.11%) aimed at studying “Vitality/wound dating”. Lastly, 26 studies (18.06%) were classified as “Other”.

Results interpretation and methodological rigor

The primary aim of this systematic review was to critically evaluate the methodological quality and forensic validity of IHC techniques applied to human post-mortem tissues. In particular, the analysis focused on assessing the methodological rigor of the included studies and on documenting the presence of standardized approaches, shared

protocols, or reproducible interpretative frameworks supporting the forensic use of IHC.

As partially outlined in the Materials and Methods section, specific attention was paid to how immunohistochemical slides were interpreted and standardized across studies. In this context, the "Slide reading" column reported in the Supplementary Table S1 represented a key analytical element, as it allowed a structured appraisal of the interpretative strategies adopted in each study.

For the purposes of the present results analysis, slide-reading methodologies were examined using six predefined methodological parameters (M1–M6), reflecting different levels of standardization, transparency, and control in IHC interpretation, including the use of predefined scoring metrics, multi-reader or digital assessment, blinding procedures, reported concordance, quality control measures, and pre-specified positivity thresholds.

The combined analysis of the six core methodological parameters (M1–M6) applied to the 144 included studies allowed a systematic quantification of the degree of standardization and interpretative robustness of immunohistochemical techniques used in forensic settings.

Although slide evaluation was performed in 100% of the studies, only 122 studies (84.7%) provided some information on the slide reading criteria. The adoption of either a quantitative or semi-quantitative approach in(M1) emerged as the most frequently satisfied criterion (109 of 144 studies, 75.69%). Specifically, 50 studies (34.72%) adopted a quantitative approach based on percentages of positive staining or counts per field (HPF/ROI), 59 studies (40.97%) adopted a semi-quantitative scoring systems (e.g., 0–3/0–4, +/++/+++, or composite scores), while 3 studies (2.08%) provided mixed, quantitative/semi-quantitative information. 10 studies (6.9%) performed strictly qualitative assessments, while, as above mentioned, 22 studies (15.28) did not provide any information. In several cases, qualitative and quantitative approaches coexisted within the same study. The overall prevalence of each methodological criterion across the entire dataset is summarized in Fig. 3.

It should be noted that M1 was designed to capture the presence of an explicit slide-reading framework rather than the degree of objectivity of the adopted scoring system. Consequently, studies employing simple ordinal scales (e.g., +/++/+++) and those using more structured approaches such as H-score, IRS, Allred score, or digital image analysis were all considered M1-positive. The distribution of scoring methodologies was therefore analyzed separately to provide additional information regarding the heterogeneity of interpretative approaches.

Quality assessment using QUADAS-2

To provide an external validation of the methodological findings obtained through the M1–M6 framework, a complementary quality appraisal was performed using QUADAS-2.

Of the 144 included studies, 110 (76.4%) were considered amenable to formal QUADAS-2 assessment, whereas 34 studies (23.6%) were classified as descriptive, mechanistic, or otherwise unsuitable for diagnostic risk-of-bias evaluation.

Among the studies eligible for scoring, overall risk of bias was judged as high in 102 studies (92.7%) and unclear in the remaining 8 studies (7.3%), while no study fulfilled all criteria required for a low overall risk-of-bias judgement.

The most frequent source of bias emerged from the patient-selection domain, reflecting the widespread use of a two-gate sampling structure in which cases and controls were recruited from separate, highly selected forensic autopsy populations. Additional concerns were commonly related to insufficient reporting of reader blinding, absence of pre-specified positivity thresholds, and limitations of the reference standards used in several forensic scenarios.

Importantly, the QUADAS-2 assessment and the M1–M6 framework explored different methodological dimensions. Whereas M1–M6 evaluated reporting transparency and reproducibility safeguards, QUADAS-2 focused on structural sources of systematic bias. Consequently, the two approaches should be regarded as complementary components of methodological quality assessment rather than alternative instruments.

The second most frequently satisfied criterion concerned the robustness of slide interpretation (M2), met by 65 of 144 studies (45.14%). Of these, 51 studies (35.42%) reported independent assessment by two or more human readers, 13 (9.03%) used software-based or digital image analysis with predefined parameters, and 1 (0.69%) combined multi-reader and software-assisted evaluation. M2 was not satisfied in 79 studies (54.86%): 67 (46.53%) provided no information on the number or means of slide assessment, 10 (6.94%) relied on a single reader, and 2 (1.39%) reported a reading modality that did not fulfil the M2 definition as applied.

Reader blinding (M3) was reported in 52 studies (36.11%), highlighting that explicit mitigation of observational bias is not yet a consolidated practice in most forensic IHC applications. Even less frequent was the reporting of inter- or intra-observer concordance (M4), documented in only 13 studies (9.03%) through statistical coefficients (κ, ICC) or qualitative descriptions of agreement. The sparse and discontinuous distribution of these criteria across individual studies is clearly visualized in Fig. 3, where large vertical bands of absence are evident for M4 and, to a lesser extent, M3.

The explicit declaration of quality control procedures (M5), including the use of positive and negative controls and the management

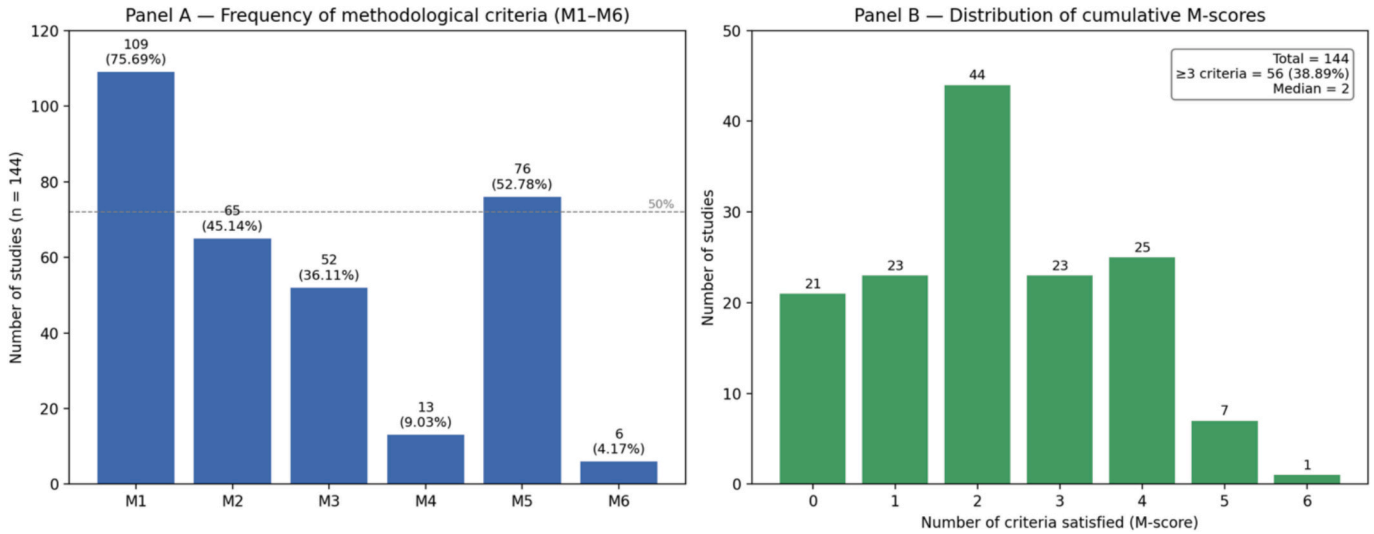

Fig. 3. Graphical representation of the M1 – M6 criteria met by each study.

of technical artefacts, was reported in only 76 studies (52.78%).

Similarly, predefined cut-off values or positivity thresholds (M6) were specified in only 6 studies (4.17%). However, this finding should be interpreted with caution, as a substantial proportion of the included studies were exploratory, descriptive, or mechanistic investigations that did not aim to generate a binary positivity or diagnostic classification. In these contexts, a predefined positivity threshold may not be required. Therefore, the absence of M6 should not be interpreted as an indicator of poor methodological quality, but rather as the lack of a pre-specified interpretative threshold in studies where such a threshold may be relevant for reproducibility and inter-study comparability.

When the six original criteria (M1–M6) were considered jointly, only 56 studies out of 144 (38.89%) satisfied at least three criteria, 33 studies (22.91%) satisfied at least four criteria, and 8 studies (5.56%) fulfilled five or more criteria, corresponding to a high level of methodological standardization. Only one study satisfied all six criteria. The distribution of the total number of criteria met per study is shown in Fig. 3, which demonstrates a clear clustering around two to three satisfied criteria and the presence of a non-negligible subset of studies with zero or one reported criterion.

Overall, the distribution of methodological criteria reveals a recurring pattern: high standardization for M1, moderate standardization for M2, predominantly through independent human readers, M3, and M5, primarily through the inclusion of positive and negative controls during IHC staining, and low standardization for criteria related to inter-rater reliability and decision thresholds (M4 and M6).

Finally, the analysis of methodological continuity indicates that cross-citation of previous studies aimed at explicit replication of

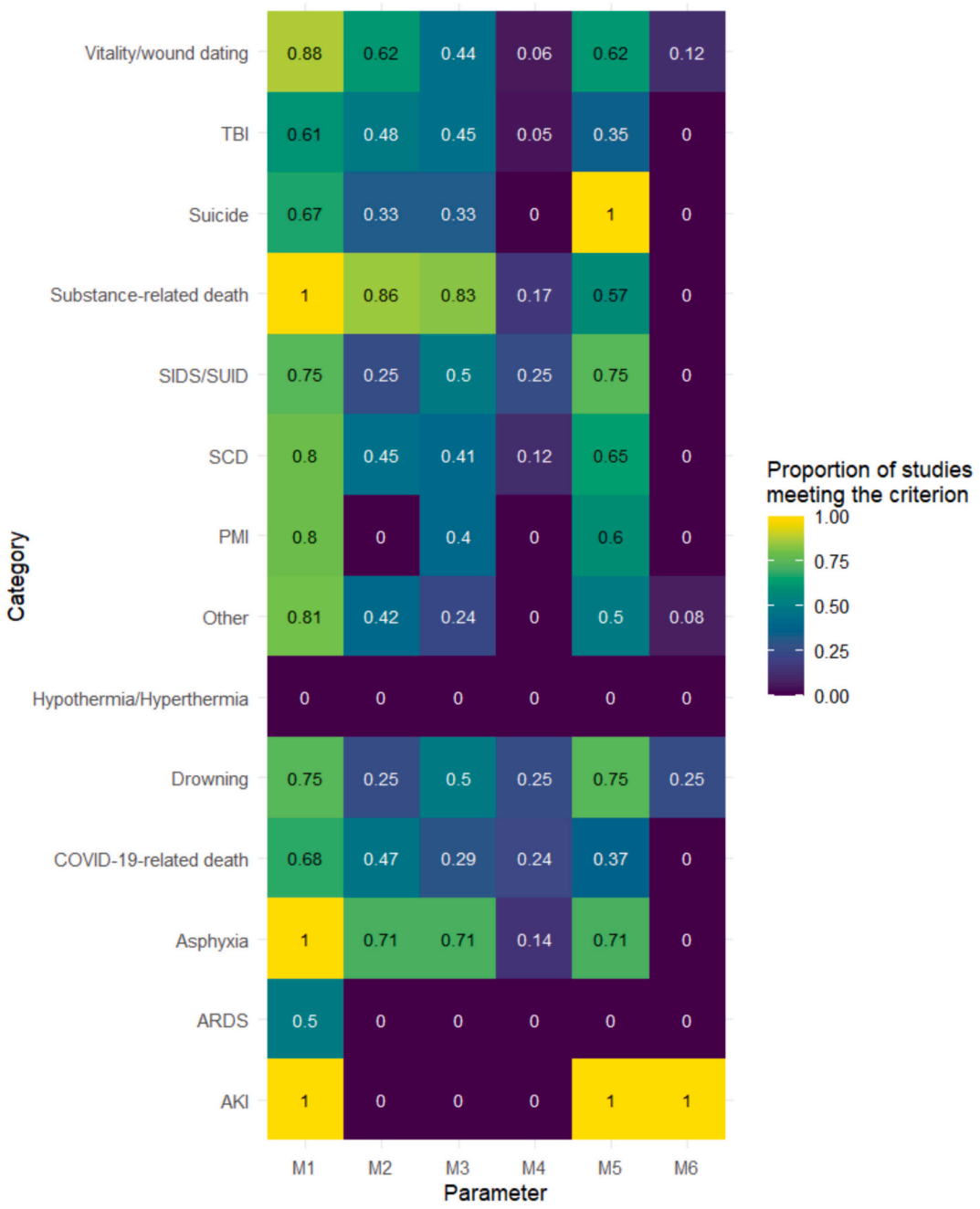

Fig. 4. Heatmap showing the percentage of studies within each forensic macro-area satisfying the M1–M6 methodological criteria. Results are presented for descriptive purposes only. Categories represented by fewer than five studies (AKI, ARDS, and hypothermia/hyperthermia) should be interpreted with caution and are not intended for comparative inference.

immunohistochemical setups is largely unsystematic. In most cases, studies share individual markers or similar biological targets but differ substantially in reading metrics, regions of interest, and interpretative criteria, contributing to overall methodological fragmentation.

When the six methodological criteria (M1–M6) are examined by forensic macro-area, the distribution and combination of standardization elements vary across categories. These patterns are presented as descriptive and exploratory observations, as several macro-areas are represented by a limited number of studies and no statistical comparison between categories was performed (Fig. 4).

The drowning category, represented by 4 studies, exhibits the highest methodological completeness, with all criteria satisfied in at least 25% of the studies. Although not generalizable due to the limited sample size, this case illustrates that a fully structured immunohistochemical workflow is feasible in forensic applications when methodological rigor is explicitly prioritized. Vitality/wound dating studies ($n = 16$) exhibit the second-highest proportion of studies meeting each criterion, despite less than 10% of the studies reporting the evaluation of inter-rater concordance and cut-off criteria (M4 and M6, respectively). Asphyxia, Covid-19-related death, SIDS/SUID, SCD, TBI, and substance-related death categories exhibit methodological completeness in a greater proportion of studies for each criterion, except for pre-defined cut-off values (M6), which no single study adopted. For each category, the reviewed studies display a higher adherence to the adoption of a semi-quantitative or quantitative assessment and the inclusion of quality controls, while the inter-rater concordance and diagnostic confirmation parameters appear to be inconsistently applied.

Suicide and PMI estimation studies show moderate adoption of predefined metrics and quality controls but limited use of blinding and organized reading.

For AKI (n = 1), ARDS ($n = 2$), and hypothermia/hyperthermia ($n = 3$), the very limited number of available studies precludes any meaningful characterization of methodological completeness. The corresponding values are reported for transparency purposes only and should not be interpreted comparatively.

The heterogeneous "Other" category shows marked variability, reflecting conceptual diversity rather than a unified methodological approach. The two ARDS studies displayed different M1–M6 profiles; however, no inference can be drawn from this category because of the extremely limited number of available studies.

Overall, the macro-area analysis is consistent with the global findings and provides an exploratory overview of how methodological practices are distributed across the currently available forensic IHC literature. These patterns help explain the fragmented methodological landscape of forensic immunohistochemistry despite the increasing use of multi-marker panels.

In addition to methodological heterogeneity, the included studies differed markedly in the conceptual role assigned to IHC within the forensic investigation. Based on the classification reported in the study table, IHC applications were stratified into three mutually exclusive categories: standard morphological evidence, mixed use, and non-morphological (functional–forensic) use.

When interpreted together, these findings indicate that slightly more than half of the reviewed studies employ immunohistochemistry in a strictly functional–forensic sense, while a substantial proportion remains confined to descriptive or hybrid applications. This conceptual heterogeneity parallels the methodological fragmentation highlighted in the preceding sections and provides an additional explanation for the limited inter-study comparability and evidentiary consolidation observed in forensic immunohistochemistry, despite the increasing use of multi-marker panels.

Combined or overlapping categories (e.g., SIDS/SUID, SCD, TBI combined with vitality/wound dating) were treated as distinct analytical entities in the descriptive classification but overlap conceptually with their parent macro-areas; therefore, counts should not be interpreted as mutually exclusive.

4. Discussion

4.1. General overview and interpretation of results

Analysis of 144 forensic IHC studies confirms the scientific vitality and methodological unevenness of the field. Rather than evaluating the biological validity of specific markers, this review systematically assessed methodological standardization—the dimension that conditions reproducibility, transferability, and evidentiary weight of published findings across all forensic application domains.

Overall, the results confirm that IHC has been extensively investigated across several key forensic domains, including traumatic brain pathology, ischemic and arrhythmic cardiac pathology, assessment of cutaneous wound vitality, asphyxial scenarios, and estimation of the post-mortem interval [9,17,30–32]. However, based on the findings of the present study, this widespread application does not appear to be accompanied by a corresponding consolidation of methodological standards. The analysis of the 144 included studies reveals marked heterogeneity in study design, sampling strategies, selection of control groups, management of confounding variables, and, most notably, in the reading and interpretation of immunohistochemical preparations.

A first relevant finding concerns the predominantly retrospective nature of the analyzed studies. The prevalence of retrospective and descriptive designs reflects an operational reality that is understandable in the forensic setting, where access to cases and biological samples is often constrained by judicial contexts and non-programmable case series. Nevertheless, this approach inherently limits the ability to systematically control crucial variables such as the post-mortem interval at the time of sampling, body preservation conditions, the actual pathophysiology of death, and the presence of relevant comorbidities. As a result, the scientific output, although rich in observations, often remains fragile from an inferential standpoint.

A second key issue emerging from the analysis concerns the management of pre- and post-mortem variables, which represent one of the main critical factors in the forensic application of IHC. Although the post-mortem interval is reported in the majority of studies with respect to case samples, it is frequently absent or incompletely documented in control groups, when these are included. Even more rarely are body preservation conditions, refrigeration procedures, tissue sampling and fixation times, and potential differences in handling between cases and controls described in sufficient detail.

In addition, variability in fixation procedures may represent a further source of methodological heterogeneity. Beyond fixation time itself, differences in the type, concentration, buffering, and preparation of fixatives may substantially influence antigen preservation, staining intensity, and antibody performance. The limited reporting of these pre-analytical parameters in many studies further complicates inter-study comparability and may contribute to variability in immunohistochemical findings.

A further limitation emerging from the present review concerns the scarce reporting of fixation-related variables, despite their recognized influence on immunohistochemical reproducibility. Evidence from pathology literature indicates that fixation delay, fixative composition, buffering, fixation duration, and tissue processing may all affect antigen preservation and staining performance [33,34]. Notably, Sato et al. demonstrated that formalin pH and fixation duration directly influence both Ki-67 immunoreactivity and nucleic-acid amplification, recommending fixation in 10% neutral buffered formalin for no longer than one week as the most reproducible protocol under controlled laboratory conditions [35,36]. However, these findings derive primarily from surgical pathology specimens and cannot be directly extrapolated to forensic autopsy tissues, where post-mortem interval, autolysis, refrigeration, and variable fixation delays introduce additional sources of variability. The limited reporting of fixation protocols in forensic IHC studies therefore remains a significant obstacle to reproducibility and inter-study comparability.

This lack of information carries significant methodological weight, as many immunohistochemical markers used in forensic settings are well known to be sensitive to autolysis, protein degradation, and post-mortem changes.

In particular, in tissues with high post-mortem vulnerability such as the brain, the myocardium, and lung, the absence of accurate documentation of pre-analytical conditions makes it extremely difficult to distinguish biological signals genuinely related to vital or perimortem processes from alterations secondary to post-mortem evolution. This limitation is especially evident in studies that attribute temporal or causal significance to the expression of markers of hypoxia, inflammation, or cellular stress without a uniform control of the variables that may influence their expression in a nonspecific manner.

A further critical aspect concerns the selection and characterization of control groups. Although a substantial proportion of studies include controls, their definition is often heterogeneous and, in some cases, conceptually problematic. In many instances, control groups consist of individuals who died from different causes, sometimes with systemic diseases or prolonged agonal states that may influence the expression of the same markers under investigation. Moreover, documentation of the cause of death, comorbidities, and post-mortem interval in control subjects is frequently incomplete or absent, thereby compromising the validity of comparisons and introducing potential systematic bias.

These structural limitations are particularly relevant in light of the explicit aim, stated in many studies, of using IHC as a supportive tool for determining the cause of death or reconstructing perimortem events. In the medicolegal setting, the interpretation of an immunohistochemical signal cannot be dissociated from a rigorous comparison with truly comparable baseline conditions, as failure to do so entails the risk of attributing forensic significance to nonspecific biological variations.

A central methodological feature of the present study is the systematic application of six predefined criteria (M1–M6) to assess interpretive standardization across 144 forensic IHC studies. Rather than restricting the evaluation to biological outcomes, this approach provides an objectifiable measure of methodological quality grounded in established standards for evidence-based laboratory practice. The results demonstrate that most studies satisfy only a limited number of criteria, with emphasis concentrated on the adoption of a slide-reading metric (M1), while critical safeguards—blinding, inter-observer concordance, and predefined positivity thresholds—remain systematically underimplemented. This pattern reflects a disconnect between the inferential ambition of forensic IHC and its operational methodological rigor.

A further consideration concerns the interpretation of M1. This criterion was designed to assess the presence of an explicit slide-reading framework rather than to rank the objectivity of different scoring methodologies. Consequently, simple ordinal systems (e.g., +/++/+++) and more structured approaches such as H-score, IRS, Allred score, or digital image analysis were coded equivalently for the purposes of M1. While this binary approach maximizes transparency and reproducibility of the assessment process, it does not capture the substantial differences in objectivity that may exist among scoring systems. M1 should therefore be interpreted as an indicator of methodological formalization rather than as a measure of interpretative rigor. The descriptive stratification reported in the Results provides a more nuanced representation of this methodological heterogeneity and identifies the progressive adoption of structured and digital scoring approaches as a potential direction for future standardization. We deliberately avoided post-hoc subdivision of M1 into graded levels of objectivity. Although sample ordinal systems (e.g., +/++/+++) and more structured approaches differ substantially in reproducibility, introducing an additional ranking would have required an independent validation process and could have reduced the transparency and inter-rater reproducibility of the framework itself.

To complement this framework, a secondary methodological appraisal was performed using the validated QUADAS-2 instrument [29]. While the M1–M6 criteria were specifically designed to evaluate reporting transparency, interpretative reproducibility, and procedural standardization in forensic immunohistochemistry, QUADAS-2 assesses the risk of systematic bias arising from study design, case selection, reference standards, and analytical procedures. The two approaches therefore evaluate distinct but complementary methodological dimensions.

The QUADAS-2 analysis broadly supported the conclusions derived from the M1–M6 assessment. Among studies amenable to formal appraisal, the majority were judged at high overall risk of bias, predominantly due to two-gate patient selection, with cases and controls recruited from separate highly selected forensic populations, together with limitations affecting reference standards and study architecture. Importantly, the principal weaknesses identified through QUADAS-2 differed from those highlighted by the M1–M6 framework. Whereas QUADAS-2 primarily captured structural sources of bias, the M1–M6 criteria identified deficiencies in reporting transparency and reproducibility safeguards, particularly regarding blinding procedures, concordance assessment, and pre-specified interpretative thresholds. Taken together, these findings suggest that the current limitations of forensic immunohistochemistry extend beyond biomarker performance and involve both the methodological design of studies and the standardization of interpretative procedures.

This uneven distribution of methodological criteria reflects a conception of IHC that remains strongly rooted in descriptive histopathological tradition rather than in a fully quantitative and standardized approach. Although the adoption of semi-quantitative or quantitative systems represents an advance beyond purely qualitative assessment, the absence of procedures to control observational bias and the lack of shared decision-making criteria substantially limit inter-study reproducibility and the transferability of results across different forensic contexts.

A specific consideration is warranted regarding the interpretation of M6 (predefined positivity thresholds). The presence of a predefined positivity threshold should not be regarded as a universal indicator of methodological quality across all forensic immunohistochemical studies. A substantial proportion of the forensic IHC literature is exploratory, descriptive, or mechanistic in nature and reports continuous or semi-quantitative expression patterns without attempting to generate a binary positive/negative classification. In such contexts, the absence of a predefined threshold is not inherently problematic and should not be interpreted as a methodological deficiency. Conversely, when immunohistochemical findings are used to support a dichotomous diagnostic or classificatory conclusion, a pre-specified threshold becomes an important reproducibility safeguard [27]. In these circumstances, reliance on data-derived or post-hoc thresholds may reduce inter-study comparability and increase the risk of optimistic interpretation, a concern that is also reflected in established methodological assessment frameworks such as QUADAS-2.

Immunohistochemistry and forensic macro-areas: interpretative potential and structural limitations.

Analysis by forensic macro-areas of application allows clarification of how the observed methodological heterogeneity is not uniform, but rather differentially distributed according to the specific medicolegal question addressed. Given the uneven distribution of studies across categories and the very small number of studies available for some macro-areas, these observations should be interpreted as descriptive and hypothesis-generating rather than as evidence of confirmed differences between forensic domains. This finding is of particular interest, as it suggests that the limitations of forensic IHC do not depend solely on general technical shortcomings, but also reflect different interpretative traditions and varying levels of inferential ambition across application domains.

In the context of cerebral pathology, particularly in studies focused on hypoxic–ischemic damage and traumatic brain injury, immunohistochemistry is frequently employed with a functional purpose, serving as a tool to document dynamic biological processes such as neuronal

suffering, axonal damage, and glial response. Markers such as beta-APP, GFAP, phosphorylated tau, neurofilaments, and cellular stress proteins are interpreted as temporal and pathogenetic surrogates, often with the aim of distinguishing between vital, perimortem, and post-mortem events [12,37–39]. Several studies have adopted more structured methodological approaches, including explicit scoring criteria and standardized immunohistochemical protocols, although these practices remain inconsistently implemented across the field [30,40,41]. The challenge lies in extending these practices from individual exemplary studies to the field as a whole: the present analysis shows that even within the TBI category, the majority of studies do not implement blinding, concordance assessment, or predefined cut-off values (M3–M6). Closing this gap represents one of the most important opportunities for methodological consolidation in forensic IHC.

A structurally analogous pattern, albeit with distinct implications, is observed in cardiac pathology [42–44]. Here, immunohistochemistry has been explored mainly in the context of myocardial inflammation and cellular stress responses, including myocarditis, virus-associated myocardial injury, and heat-shock protein expression. While these studies demonstrate the ability of IHC to provide additional information beyond routine histology, they also reveal persistent methodological limitations, particularly regarding case selection, control-group definition, and the interpretation of quantitative findings. Consequently, immunohistochemical results should be interpreted as complementary rather than independently diagnostic evidence and integrated with conventional histopathological, clinical, and circumstantial data. Further progress will likely depend on greater methodological standardization and more consistent reporting of interpretative criteria.

A relevant aspect emerging from the present analysis concerns the heterogeneous conceptual role attributed to immunohistochemistry across studies. In a non-negligible proportion of cases, IHC was employed exclusively as standard morphological evidence, serving to characterize cellular populations, tissue components, or antigen localization without interpreting marker expression as a biological response to injury, vitality, or post-mortem timing. This descriptive use included immunophenotypic markers of inflammatory or immune cells (e.g., CD3, CD4, CD8, CD20, CD68, IBA1, TMEM119) as well as structural or vascular markers (e.g., CD34). In these contexts, IHC functioned primarily as an ancillary histopathological tool rather than as a source of dynamic or inferential information.

In the field of cutaneous wound vitality, immunohistochemistry remains one of the ancillary approaches most frequently explored to support the distinction between vital and post-mortem lesions. A wide range of inflammatory, vascular, and cellular stress or activation markers has been investigated for this purpose [45–48]. Several studies have adopted explicit semi-quantitative or quantitative interpretation criteria and, in some cases, predefined diagnostic thresholds. Nevertheless, substantial heterogeneity persists in marker selection, scoring systems, and study design, limiting comparability across investigations. Furthermore, formal assessments of inter-observer reproducibility remain uncommon, and approaches to interpretation and reporting remain heterogeneous across studies. Variability related to survival time, post-mortem interval, and individual biological response further complicates the interpretation of immunohistochemical findings and underscores the need for broader validation studies.

Studies focusing on pulmonary tissue and hypoxia-related deaths illustrate both the potential and the limitations of forensic immunohistochemistry. In these investigations, markers associated with hypoxic stress, inflammatory responses, and alveolar alterations— including surfactant protein A, HIF-1α, mast-cell markers, and other markers associated with hypoxic stress—have been explored as ancillary indicators of asphyxial or hypoxic mechanisms of death [49–51]. At the same time, the available evidence highlights important methodological constraints, including variability in study design, limited validation across independent cohorts, and inconsistent approaches to marker quantification. Notably, marker performance may be influenced by post-mortem tissue alterations, as illustrated by the differing behaviour of HIF-1α and SP-A in decomposed cases [13,49–51]. Collectively, these studies suggest that immunohistochemical findings may provide useful complementary information when interpreted alongside conventional pathological evidence, but further validation, standardization, and reproducibility studies remain necessary before such markers can be regarded as robust forensic diagnostic tools.

4.2. Functional use of immunohistochemistry and evidentiary value

A relevant finding of the present review is the widespread use of immunohistochemistry for inferential rather than purely descriptive purposes. Across multiple forensic domains, IHC was frequently employed to support interpretations regarding wound vitality, survival time, tissue responses, and mechanisms of death. While such applications are biologically plausible and increasingly common, their evidentiary value depends on the reproducibility and transparency of the underlying methodology.

The present analysis shows that forensic IHC has adopted several methodological safeguards, particularly predefined reading metrics (M1: 75.7%). However, formal assessment of inter-observer agreement remained uncommon (M4: 10%), formal assessment of reproducibility was rarely reported, and predefined positivity thresholds were often absent. Similar concerns emerged from the complementary QUADAS-2 assessment. Collectively, these findings indicate that methodological variability remains a major limitation to the reproducibility and comparability of forensic IHC studies.

Existing frameworks developed for diagnostic pathology and biomarker research, including STARD [52], REMARK [53], QUADAS-2 [29], and recommendations for immunohistochemical assay validation [54], may provide useful methodological references for future forensic IHC research. Based on the gaps identified in this review, greater emphasis should be placed on analytical validation, standardized interpretation criteria, assessment of observer agreement, and systematic reporting of relevant pre-analytical variables [54–57]. Such measures may improve the transparency, reproducibility, and evidentiary robustness of forensic immunohistochemistry.

5. Conclusion

This systematic appraisal of 144 studies confirms that contemporary forensic immunohistochemistry is a scientifically active but methodologically heterogeneous field. IHC has been applied across several forensic domains—including traumatic brain injury, cardiac pathology, cutaneous wound vitality, asphyxial and hypoxia-related deaths, and post-mortem interval estimation—frequently with an inferential rather than a purely descriptive purpose. The central finding of this review is not that forensic IHC lacks biological plausibility or forensic utility, but that its potential remains unevenly supported by the methodological safeguards required to ensure reproducibility and comparability across studies.

The structured M1–M6 assessment revealed a consistent pattern. Predefined slide-reading metrics were widely adopted (M1: 75.7%) and quality controls were moderately implemented (M5: 52.8%), whereas safeguards more directly related to interpretative reproducibility—including blinded reading, formal assessment of inter-observer agreement (M4: 10%), and predefined positivity thresholds—were reported only infrequently. A complementary QUADAS-2 appraisal supported these findings, with most eligible studies judged at high overall risk of bias, predominantly due to two-gate case selection and limitations in study design and reference standards.

Taken together, these findings indicate that the current limitations of forensic immunohistochemistry are primarily methodological rather than biological. Accordingly, immunohistochemical findings should be interpreted as complementary to conventional histopathological,

clinical, and circumstantial evidence rather than as independently decisive. Future studies should place greater emphasis on analytical validation, standardized interpretation criteria, assessment of observer agreement, and systematic reporting of relevant pre-analytical variables in order to improve the reproducibility, comparability, and evidentiary value of forensic immunohistochemical investigations.

PROSPERO registration

CRD420251063965 — https://www.crd.york.ac.uk/PROSPERO/view/CRD420251063965

Ethics approval

Not applicable. This study is based exclusively on published literature and did not involve human participants, animal subjects, or identifiable patient data.

Declaration of generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this manuscript the authors used Claude (Anthropic) to assist with text editing and language revision. After using this tool, the authors reviewed and edited all content as needed and take full responsibility for the content of the published article.

Funding

No funds, grants, or other support was received.

Authors' contribution

L.T. and M.O. conceived and designed the study. L.T. coordinated the overall research activity and supervised all phases of the study. M.O. contributed substantially to data analysis and interpretation and drafted the first version of the manuscript. M.L. contributed to the acquisition and interpretation of pathological and laboratory data. P.F. and R.S. provided critical forensic and medico-legal expertise and contributed to case evaluation. N.P. contributed to methodological support and critical revision of the manuscript. All authors contributed to manuscript revision, approved the final version, and agree to be accountable for all aspects of the work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.legalmed.2026.102894.

References

[1] L.L. De Matos, D.C. Trufelli, De Matos MGL, M.A. Da Silva Pinhal, Immunohistochemistry as an important tool in biomarkers detection and clinical practice, Biomark. Insights 5 (2010) 9–20, https://doi.org/10.4137/bmi.s2185 PubMed PMID: 20212918; PubMed Central PMCID: PMC2832341.

[2] H.M. Hussaini, B. Seo, A.M. Rich, Immunohistochemistry and immunofluorescence, Methods Mol. Biol. 2588 (2023) 439–450, https://doi.org/10.1007/978-1-0716-2780-8_26 (PubMed PMID: 36418703).

[3] J. Duraiyan, R. Govindarajan, K. Kaliyappan, M. Palanisamy, Applications of immunohistochemistry, J. Pharm. Bioallied Sci. 4 (Suppl. 2) (2012) S307–S309, https://doi.org/10.4103/0975-7406.100281 PubMed PMID: 23066277; PubMed Central PMCID: PMC3467869.

[4] B. Brattoli, M. Mostafavi, T. Lee, W. Jung, J. Ryu, S. Park, et al., A universal immunohistochemistry analyzer for generalizing AI-driven assessment of immunohistochemistry across immunostains and cancer types, NPJ Precis. Oncol. 8 (1) (2024) 277, https://doi.org/10.1038/s41698-024-00770-z PubMed PMID: 39627299; PubMed Central PMCID: PMC11615360.

[5] P.K. Nakane, G.B. Pierce, Enzyme-labeled antibodies: preparation and application for the localization of antigens, J. Histochem. Cytochem. 14 (12) (1966) 929–931, https://doi.org/10.1177/14.12.929 (PubMed PMID: 17121392).

[6] S.R. Shi, C. Liu, C.R. Taylor, Standardization of immunohistochemistry for formalin-fixed, paraffin-embedded tissue sections based on the antigen-retrieval technique: from experiments to hypothesis, J. Histochem. Cytochem. 55 (2) (2007) 105–109, https://doi.org/10.1369/jhc.6P7080.2006 (PubMed PMID: 16982846).

[7] J.M. Tortosa Lopez, Histopathology in medico-legal autopsies: clinical foundation, legal mandate, and judicial guarantee, J. Forensic Leg. Med. 115 (2025) 102974, https://doi.org/10.1016/j.jflm.2025.102974 (PubMed PMID: 40967093).

[8] D. Jin, L. Zhang, Forensic wound age estimation (2003-2024): evolving perspectives on tissue viability and molecular pathology, Leg. Med. (Tokyo) 77 (2025), https://doi.org/10.1016/j.legalmed.2025.102677 (PubMed PMID: 40716191).

[9] E. Buratti, G. Ricchezze, A. Raimondi, M. Bacci, Asphyxial death biomarkers: a narrative review, J. Forensic Sci. Medicine. 11 (3) (2025) 157–162, https://doi.org/10.4103/jfsm.jfsm_99_24.

[10] A. Natanti, R. Mazzanti, M. Palpacelli, C. Turchi, A. Tagliabracci, M. Pesaresi, Death following extreme temperature exposure: histological, biochemical and immunohistochemical markers, Med. Sci. Law 61 (1_suppl) (2021) 36–41, https://doi.org/10.1177/0025802420942423.

[11] M. Nogami, T. Hoshi, Y. Toukairin, T. Arai, T. Nishio, Immunohistochemistry of advanced glycation end product Nε-(carboxymethyl)lysine in coronary arteries in relation to cardiac fibrosis and serum N-terminal-pro basic natriuretic peptide in forensic autopsy cases, BMC Res. Notes 13 (1) (2020) 239, https://doi.org/10.1186/s13104-020-05082-6 PubMed PMID: 32398121; PubMed Central PMCID: PMC7216670.

[12] M. Olczak, Ł. Poniatowski, M. Kwiatkowska, D. Samojłowicz, S. Tarka, T. Wierzba-Bobrowicz, Immunolocalization of dynein, dynactin, and kinesin in the cerebral tissue as a possible supplemental diagnostic tool for traumatic brain injury in postmortem examination, Folia Neuropathol. 57 (1) (2019) 51–62, https://doi.org/10.5114/fn.2019.83831 (PubMed PMID: 31038188).

[13] R. Scendoni, L. Ferrante, D. Stramazzotti, A. Tagliabracci, Analysis of immunohistochemical expression of inducible nitric oxide synthase for the evaluation of agonal time in forensic medicine, Int. J. Legal Med. 130 (6) (2016) 1639–1646, https://doi.org/10.1007/s00414-016-1402-8 (PubMed PMID: 27329069).

[14] A.B. Bromley, M.J. Trotter, Immunohistochemistry utilization in autopsy pathology: a Canadian experience, Pathol. Res. Pract. 207 (4) (2011) 241–246, https://doi.org/10.1016/j.prp.2011.02.006 (PubMed PMID: 21440379).

[15] S. Sabatasso, P. Mangin, T. Fracasso, M. Moretti, M. Docquier, V. Djonov, Early markers for myocardial ischemia and sudden cardiac death, Int. J. Legal Med. 130 (5) (2016) 1265–1280, https://doi.org/10.1007/s00414-016-1401-9 (PubMed PMID: 27392959).

[16] H. Bi, Y. Yang, J. Huang, Y. Li, C. Ma, B. Cong, Immunohistochemical detection of S100A1 in the postmortem diagnosis of acute myocardial infarction, Diagn. Pathol. 8 (2013) 84, https://doi.org/10.1186/1746-1596-8-84 PubMed PMID: 23683996; PubMed Central PMCID: PMC3663776.

[17] Buğra A, Daş T. The role of immunohistochemical markers in the diagnosis of early myocardial infarction. Cureus 14(2):e22391. doi:https://doi.org/10.7759/cureus.22391 PubMed PMID: 35371665; PubMed Central PMCID: PMC8938247 (web archive link,).

[18] Taylor CR, Levenson RM. Quantification of immunohistochemistry–issues concerning methods, utility and semiquantitative assessment II. Histopathology 2006 49(4):411–124. doi:https://doi.org/10.1111/j.1365-2559.2006.02513.x (PubMed PMID: 16978205).

[19] Tomassini L, Ricchezze G, Gambelunghe C, Lancia M, Goracci V, De Micco F, et al. Immunohistochemical insights into hyperthermia-related deaths: a systematic review. Int. J. Legal Med. 2025 139(5):2467–2501. doi:https://doi.org/10.1007/s00414-025-03485-3 (PubMed PMID: 40189649).

[20] J.A. Ramos-Vara, Principles and methods of immunohistochemistry, Methods Mol. Biol. 691 (2011) 83–96, https://doi.org/10.1007/978-1-60761-849-2_5 (PubMed PMID: 20972748).

[21] S. Mauriello, M. Treglia, M. Pallocci, R. Bonfiglio, E. Giacobbi, P. Passalacqua, et al., Antigenicity preservation is related to tissue characteristics and the post-mortem interval: immunohistochemical study and literature review, Healthcare 10 (8) (2022) 1495, https://doi.org/10.3390/healthcare10081495 PubMed PMID: 36011152; PubMed Central PMCID: PMC9408092.

[22] J. Mnookin, Science, justice, and evidence, Science 382 (6672) (2023) 741, https://doi.org/10.1126/science.adm8834 (PubMed PMID: 37972152).

[23] R. Harmon, CHAPTER 18 - admissibility standards for scientific evidence, in: R. G. Breeze, B. Budowle, S.E. Schutzer (Eds.), Microbial Forensics [Internet], Academic Press, Burlington, 2005, pp. 381–392 [cited 2026 Jun 13]. Available from: https://www.sciencedirect.com/science/article/pii/B9780120884834500214 https://doi.org/10.1016/B978-012088483-4/50021-4.

[24] Blackmun. William Daubert, Et Ux., Etc., Et Al., Petitioners v. Merrell Dow Pharmaceuticals, Inc. U.S. [Internet] [cited 2026 Jun 15]. Available from: https://www.law.cornell.edu/supct/html/92-102.ZO.html, 1993.

[25] J. Pergolizzi, LeQuang JAK, Black robes and white coats: Daubert standard and medical and legal considerations for medical expert witnesses, Cureus 16 (9) (2024), https://doi.org/10.7759/cureus.69346 PubMed PMID: 39398691; PubMed Central PMCID: PMC11471272.

[26] F. De-Giorgio, E. Bergamin, A. Baldi, R. Gatta, V.L. Pascali, Immunohistochemical expression of HMGB1 and related proteins in the skin as a possible tool for determining post-mortem interval: a preclinical study, Forensic Sci. Med. Pathol. 20 (1) (2024) 149–165, https://doi.org/10.1007/s12024-023-00634-1 PubMed PMID: 37490201; PubMed Central PMCID: PMC10944391.
[27] AI-powered systematic review & literature review platform [Internet]. [cited 2026 Jun 13]. Available from: https://www.rayyan.com/.
[28] M. Ouzzani, H. Hammady, Z. Fedorowicz, A. Elmagarmid, Rayyan-a web and mobile app for systematic reviews, Syst. Rev. 5 (1) (2016) 210, https://doi.org/10.1186/s13643-016-0384-4 PubMed PMID: 27919275; PubMed Central PMCID: PMC5139140.
[29] P.F. Whiting, A.W.S. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, et al., QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (8) (2011) 529–536, https://doi.org/10.7326/0003-4819-155-8-201110180-00009 (PubMed PMID: 22007046).
[30] J. Matschke, A. Büttner, M. Bergmann, C. Hagel, K. Püschel, M. Glatzel, Encephalopathy and death in infants with abusive head trauma is due to hypoxic-ischemic injury following local brain trauma to vital brainstem centers, Int. J. Legal Med. 129 (1) (2015) 105–114, https://doi.org/10.1007/s00414-014-1060-7 (PubMed PMID: 25107298).
[31] C. Mondello, A. Micali, G. Baldino, L. Cardia, A. Alibrandi, A. Asmundo, et al., Immunohistochemical analysis of Sigma-1 receptor (σ-1R) expression in human pineal gland in relation to different causes of death, Leg. Med. (Tokyo) 69 (2024) 102434, https://doi.org/10.1016/j.legalmed.2024.102434 (PubMed PMID: 38493555).
[32] J. Ortmann, E. Doberentz, B. Madea, Immunohistochemical methods as an aid in estimating the time since death, Forensic Sci. Int. 273 (2017) 71–79, https://doi.org/10.1016/j.forsciint.2017.02.004 (PubMed PMID: 28236731).
[33] K.B. Engel, H.M. Moore, Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue, Arch. Pathol. Lab Med. 135 (5) (2011) 537–543, https://doi.org/10.5858/2010-0702-RAIR.1 (PubMed PMID: 21526952).
[34] B.P. Bass, K.B. Engel, S.R. Greytak, H.M. Moore, A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? Arch. Pathol. Lab Med. 138 (11) (2014) 1520–1530, https://doi.org/10.5858/arpa.2013-0691-RA (PubMed PMID: 25357115).
[35] S.R. Greytak, K.B. Engel, B.P. Bass, H.M. Moore, Accuracy of molecular data generated with FFPE biospecimens: lessons from the literature, Cancer Res. 75 (8) (2015) 1541–1547, https://doi.org/10.1158/0008-5472.CAN-14-2378 PubMed PMID: 25836717; PubMed Central PMCID: PMC4636024.
[36] M. Sato, M. Kojima, A.K. Nagatsuma, Y. Nakamura, N. Saito, A. Ochiai, Optimal fixation for total preanalytic phase evaluation in pathology laboratories: a comprehensive study including immunohistochemistry, DNA, and mRNA assays, Pathol. Int. 64 (5) (2014) 209–216, https://doi.org/10.1111/pin.12164 (PubMed PMID: 24888774).
[37] Koga S, Kojima A, Kuwabara S, Yoshiyama Y. Immunohistochemical analysis of tau phosphorylation and astroglial activation with enhanced leptin receptor expression in diet-induced obesity mouse hippocampus. Neurosci. Lett. 2014 571:11–16. doi: https://doi.org/10.1016/j.neulet.2014.04.028 (PubMed PMID: 24785100).
[38] R. Ding, Y. Hase, K.E. Ameen-Ali, M. Ndung'u, W. Stevenson, J. Barsby, et al., Loss of capillary pericytes and the blood-brain barrier in white matter in poststroke and vascular dementias and Alzheimer's disease, Brain Pathol. 30 (6) (2020) 1087–1101, https://doi.org/10.1111/bpa.12888 (PubMed PMID: 32705757; PubMed Central PMCID: PMC8018063).
[39] P. Kivisäkk, H. Fatima, C.Y. Wu, N. Padmanabhan, D. Romero, T. Gorham, et al., Postmortem associations between Alzheimer disease pathology and plasma pTau217, GFAP, and NfL in AD and AD-related dementias, Neurology 106 (4) (2026) e214351, https://doi.org/10.1212/WNL.0000000000214351 PubMed PMID: 41610380; PubMed Central PMCID: PMC12857747.
[40] E.A. Ogłodek, M. Bar, Astroglial and neuronal injury markers (GFAP, UCHL-1, NfL, tau, S100B) as diagnostic and prognostic biomarkers in PTSD and neurological disorders, Int. J. Mol. Sci. 27 (5) (2026) 2374, https://doi.org/10.3390/ijms27052374 PubMed PMID: 41828591; PubMed Central PMCID: PMC12986300.
[41] M. Sharma, A. Subramaniam, K. Sengar, V. Suri, D. Agrawal, N. Chakraborty, et al., Pathological spectrum and β-APP immunoreactivity as a diagnostic tool of diffuse axonal injury following traumatic brain injury: a novel classification, J. Lab. Physicians 15 (3) (2023) 399–408, https://doi.org/10.1055/s-0043-1761926 PubMed PMID: 37564231; PubMed Central PMCID: PMC10411120.
[42] I. Gaaloul, S. Riabi, M. Evans, T. Hunter, S. Huber, M. Aouni, Postmortem diagnosis of infectious heart diseases: a mystifying cause of sudden infant death, Forensic Sci. Int. 1 (262) (2016) 166–172, https://doi.org/10.1016/j.forsciint.2016.03.002.
[43] S. Hiyamizu, Y. Ishida, H. Yasuda, Y. Kuninaka, M. Nosaka, A. Ishigami, et al., Forensic significance of intracardiac expressions of Nrf2 in acute myocardial ischemia, Sci. Rep. 14 (1) (2024) 4046, https://doi.org/10.1038/s41598-024-54530-x.
[44] Y. Kuninaka, Y. Ishida, M. Nosaka, A. Ishigami, A. Taruya, E. Shimada, et al., Forensic significance of intracardiac heme oxygenase-1 expression in acute myocardial ischemia, Sci. Rep. 11 (1) (2021) 21828, https://doi.org/10.1038/s41598-021-01102-y PubMed PMID: 34750390; PubMed Central PMCID: PMC8575909.
[45] K. Olds, R.W. Byard, C. Winskog, N.E.I. Langlois, Validation of alternate light sources for detection of bruises in non-embalmed and embalmed cadavers, Forensic Sci. Med. Pathol. 13 (1) (2017) 28–33, https://doi.org/10.1007/s12024-016-9822-9 (PubMed PMID: 28019003).
[46] Peyron PA, Colomb S, Becas D, Adriansen A, Gauchotte G, Tiers L, et al. Cytokines as new biomarkers of skin wound vitality. Int. J. Legal Med. 2021 135(6): 2537–2545. doi:https://doi.org/10.1007/s00414-021-02659-z (PubMed PMID: 34313846).
[47] J. Lier, B. Ondruschka, I. Bechmann, J. Dreßler, Fast microglial activation after severe traumatic brain injuries, Int. J. Legal Med. 134 (6) (2020) 2187–2193, https://doi.org/10.1007/s00414-020-02308-x PubMed PMID: 32372233; PubMed Central PMCID: PMC7578125.
[48] Jebur1 MAM, Hashim2 NG, Al-Obeidi3 SRH, Mast cell, IL-1 beta and IL-6 for wound timing and vitality in forensic practice, Indian J. Forensic Med. Toxicol. 14 (2) (2020) 2149–2153, https://doi.org/10.37506/ijfmt.v14i2.3333.
[49] C. Mondello, A. Micali, G. Baldino, L. Cardia, A. Alibrandi, A. Asmundo, et al., Immunohistochemical analysis of Sigma-1 receptor (σ-1R) expression in human pineal gland in relation to different causes of death, Legal Med. 69 (2024) 102434, https://doi.org/10.1016/j.legalmed.2024.102434.
[50] B. Baldari, S. De Simone, L. Cipolloni, P. Frisoni, L. Alfieri, D'Errico S, et al., Oxidative stress markers in human brain and placenta may reveal the timing of hypoxic-ischemic injury: evidence from an immunohistochemical study, Int. J. Mol. Sci. 24 (15) (2023) 12221, https://doi.org/10.3390/ijms241512221 PubMed PMID: 37569597; PubMed Central PMCID: PMC10418753.
[51] C. Reisinger, J. Sölkner, D.U. Risser, M.H. Seltenhammer, Diagnosis of asphyxiation in advanced decomposed corpses by using oil-red-O stain and immunohistochemical technology, Leg. Med. (Tokyo) 62 (2023) 102214, https://doi.org/10.1016/j.legalmed.2023.102214 (PubMed PMID: 36905850).
[52] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L. Irwig, et al., STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies, BMJ 351 (2015) h5527, https://doi.org/10.1136/bmj.h5527 PubMed PMID: 26511519; PubMed Central PMCID: PMC4623764.
[53] Reporting recommendations for tumor marker prognostic studies (REMARK): an abridged explanation and elaboration - PubMedReporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration - PubMed [Internet] [cited 2026 Jun 13]. Available from, https://pubmed.ncbi.nlm.nih.gov/29873743/.
[54] P.L. Fitzgibbons, L.A. Bradley, L.A. Fatheree, R. Alsabeh, R.S. Fulton, J. D. Goldsmith, et al., Principles of analytic validation of immunohistochemical assays: guideline from the College of American Pathologists Pathology and Laboratory Quality Center, Arch. Pathol. Lab Med. 138 (11) (2014) 1432–1443, https://doi.org/10.5858/arpa.2013-0610-CP (PubMed PMID: 24646069).
[55] W. Van Den Bogaert, J. Wuestenbergs, B. Bekaert, E. Dequeker, W. Van de Voorde, Enhancing postmortem diagnostics: over a decade of ISO 17020 accreditation and guidelines implementation in forensic pathology, Forensic Sci. Med. Pathol. 21 (2) (2025) 682–689, https://doi.org/10.1007/s12024-024-00918-0 PubMed PMID: 39607669; PubMed Central PMCID: PMC12325508.
[56] E.E. Torlakovic, S. Nielsen, M. Vyberg, C.R. Taylor, Getting controls under control: the time is now for immunohistochemistry, J. Clin. Pathol. 68 (11) (2015) 879–882, https://doi.org/10.1136/jclinpath-2014-202705 PubMed PMID: 26286753; PubMed Central PMCID: PMC4680144.
[57] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem Med (Zagreb). 22 (3) (2012) 276–282 (PubMed PMID: 23092060; PubMed Central PMCID: PMC3900052).