TREND AND ADVANCED STUDIES IN HEALTH SCIENCES

All Sciences Academy

TREND AND ADVANCED STUDIES IN HEALTH SCIENCES

Editor
Prof. Dr. Fatih HATİPOĞLU

Trend and Advanced Studies in Health Sciences
Editor: Prof. Dr. Fatih HATİPOĞLU

Design: All Sciences Academy Design
Published Date: July 2026
Publisher's Certification Number: 72273
ISBN: 978-625-8839-34-0

www.allsciencesacademy.com
allsciencesacademy@gmail.com

CONTENT

Trustworthy Artificial Intelligence for Dental and Maxillofacial Imaging: Validation, Bias, Explainability, and Clinical Translation

Alperen YALIM[1*]

Hilal YILMAZ YALIM[2]

[1] Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Türkiye; E-mail: alperen.yalim@ikcu.edu.tr; alpyalim95@gmail.com; ORCID: 0009-0006-8817-1606. * Corresponding author.

[2] Department of Pediatric Dentistry, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Türkiye; E-mail: dt.hilalyilmaz95@gmail.com; ORCID: 0009-0006-2694-057X.

ABSTRACT

Artificial intelligence (AI) can detect, segment, measure, and describe findings on panoramic, bitewing, periapical, and cone-beam computed tomography images. Multimodal systems and foundation models, which are broadly pretrained and adaptable to multiple downstream tasks, may also combine images with symptoms, records, and prior examinations. High retrospective accuracy, however, does not establish safe clinical use. Devices, acquisition protocols, populations, disease spectra, workflows, and uncertain reference diagnoses all affect performance, and an acceptable average can conceal failure in an important subgroup. This chapter presents a lifecycle framework for trustworthy dental and maxillofacial imaging AI. It examines modality-specific hazards, data provenance, patient-level data separation, task-appropriate reference standards, external validation, calibration, clinical utility, fairness, uncertainty, explainability, and clinician-AI interaction. The framework then connects silent prospective evaluation with controlled clinical-impact studies, regulatory conformity, post-market surveillance, and versioned change control. Current evidence supports the technical feasibility of several narrow detection and segmentation tasks, but independent external validation and prospective studies of patient outcomes remain sparse. Foundation and vision-language models expand capability while adding risks from hallucination, prompt sensitivity, provenance, privacy, and drift. Safe deployment therefore requires a defined use case, representative evidence, effective failure controls, clinical oversight, and continuing revalidation. Clinical readiness must be judged against the intended decision and consequences of error.

Keywords: Artificial intelligence, dental radiology, cone-beam computed tomography, external validation, multimodal learning, clinical translation

1. INTRODUCTION

Artificial intelligence has moved from experimental classification toward localization, three-dimensional segmentation, automated measurement, decision support, and report generation in dentistry. Panoramic and intraoral radiographs generate large volumes of digital data, cone-beam computed tomography (CBCT) depicts volumetric anatomy, and charts, symptoms, treatment history, and prior images add clinical context. Yet repeated images, device-specific processing, narrow disease spectra, and uncertain labels can make a model appear successful without learning a transportable clinical signal.

Early reviews of dental AI described opportunities to improve consistency and support clinical decisions, while also emphasizing data quality, generalizability, interpretability, and accountability [1]. Subsequent reporting

guidance called for transparent descriptions of datasets, reference standards, modelling choices, and evaluation [2]. These concerns become more consequential as analysis expands from a tooth or lesion to an imaging-informed clinical narrative.

Trustworthiness cannot be inferred from one area under the receiver operating characteristic (ROC) curve (AUROC). It is justified confidence that a defined clinical system, in which people use AI, will provide more benefit than harm for its intended population, user, setting, input, and action. Evidence must extend from data governance and locked testing through external validation, human-factors assessment, prospective evaluation, deployment, monitoring, and controlled updates. Technical performance alone does not resolve transportability or workflow risk: a small segmentation error beside the mandibular canal can matter, and even a calibrated classifier has little value if its output arrives after the decision.

This chapter reviews AI systems for panoramic, bitewing, periapical, and CBCT imaging, including multimodal and foundation-model approaches. It focuses on three clinical questions: what evidence is required before an output influences care, how failures should be measured, and how a deployed system can remain auditable over time.

2. LITERATURE SEARCH AND SYNTHESIS APPROACH

This critical narrative review used targeted literature searches, updated through 24 July 2026. Search concepts combined dentistry or dental radiology with artificial intelligence, machine learning or deep learning; panoramic, bitewing, periapical or cone-beam computed tomography imaging; and multimodal learning, foundation models, external validation, calibration, explainability, clinical implementation, regulation, or post-market monitoring. Priority was given to systematic reviews and international guidance, independent external-validation studies, reader studies, prospective evaluations, and studies reporting clinically interpretable failure modes or outcomes; influential technical studies were retained when they defined the state of a specific task. Evidence was compared by intended use, imaging modality, reference standard, independence of evaluation, and proximity of the endpoint to patient care. This was not a systematic review: exhaustive dual screening, protocol registration, and formal study-level risk-of-bias scoring were not undertaken. Substantial heterogeneity in tasks, populations, thresholds, reference procedures, and reported metrics precluded meaningful meta-analysis, so findings were synthesized critically and narratively.

3. THE CLINICAL AND IMAGING PROBLEM SPACE

3.1. Panoramic radiography

Panoramic radiographs provide a broad survey of teeth, jaws, sinuses, and supporting structures. Their large field of view supports tooth numbering, anatomical segmentation, detection of impacted teeth, assessment of periodontal bone loss, and broad screening. Yet panoramic formation is projection dependent. Magnification, superimposition, ghosting, head positioning, focal-trough errors, and vendor post-processing can alter appearance. Early proximal caries may be intrinsically difficult to resolve, while appliances, implants, edentulous spaces, and mixed dentition alter the range of appearances encountered in practice.

Technical studies illustrate both progress and residual heterogeneity. Panoptic segmentation has delineated multiple structures, including teeth, maxillary sinuses, and mandibular canals [3]. A tooth identification system developed on 6,046 panoramic radiographs reported precision and recall above 97% [4]. Detection of periodontal bone loss achieved an AUROC of 0.951 for total bone loss but only 0.733 for vertical loss, showing that a strong aggregate result can conceal a clinically important weak subclass [5]. When two public panoramic datasets were combined, investigators detected missing and incorrect labels; detection of impacted teeth remained much easier than early caries detection [6]. Dataset scale therefore does not remove ontology, annotation, or spectrum problems.

3.2. Bitewing and periapical radiography

Bitewings offer detailed views of proximal surfaces and crestal bone. They are a natural target for caries detection, but overlap, exposure, cervical burnout, restoration margins, lesion stage, and prevalence strongly influence performance. In one study using 3,686 bitewings, a deep-learning system showed greater sensitivity than the dentist comparison group on an internal test set [7]. In another study of early caries, AI assistance increased readers' sensitivity but reduced positive predictive value [8]. The randomized ADEPT reader study similarly increased sensitivity for enamel lesions from 44.3% to 75.8%, while false-positive responses increased from 3.7% to 14.6% [9]. Assistance may therefore reduce missed lesions at the cost of more false-positive findings. The clinical value depends on whether extra detections lead to appropriate monitoring or unnecessary operative treatment.

Independent external validation remains particularly important for commercial systems. A 2026 external pilot evaluated two FDA-cleared decision support systems in 90 patients and reported high negative predictive values for selected caries-detection and periodontal-bone-loss tasks [10].

These findings support further evaluation of the systems as adjunctive screening tools, but a rule-out indication would require prespecified high sensitivity, representative prevalence, calibration, and prospective safety evidence. The small retrospective sample does not establish patient benefit or broad transportability.

Periapical imaging presents a different challenge: the reference diagnosis for apical disease may be uncertain. Radiographic consensus can support a radiographic detection claim, whereas a clinical-disease claim may require CBCT, operative findings, histopathology, records, or follow-up. In a 2025 randomized crossover study using CBCT as the reference, AI assistance increased overall accuracy for periapical radiolucencies from 91.6% to 93.3% and reduced false positives, but sensitivity remained approximately 46% [11]. Average accuracy alone would have obscured that limitation.

3.3. Cone-beam computed tomography

CBCT allows three-dimensional mapping of jaws, teeth, canals, sinuses, and lesions. It also multiplies sources of variation: scanner, field of view, voxel size, tube settings, reconstruction kernel, dose, motion, metal artifacts, and export pipeline. Producing reliable labels is resource intensive, and slice-based pipelines are vulnerable to leakage if slices from one patient are assigned to different data subsets.

Multiclass segmentation of the jaws and teeth has shown strong technical feasibility [12]. Yeshua et al. reported high detection and three-dimensional segmentation performance for maxillofacial bone lesions in 82 subjects imaged with three devices. The sample included histologically confirmed benign lesions and controls. Its narrow lesion spectrum and small size limit inference about incidental lesions, malignancy, and community prevalence [13]. A clinically oriented 3D network for periapical lesions showed high performance in both internal and geographically external validation. In an assisted-reader study, the model improved dentists' AUROC and reduced evaluation time [14].

Boundary metrics require clinical interpretation. Automated segmentation of the mandibular canal was substantially faster than manual tracing, but discrepancies ranged from 0.21 to 7.65 mm and were larger in the posterior canal and anterior loop regions [15]. A later study reported a Dice score of 0.77, a mean average symmetric surface distance of 0.31 mm, and qualitative performance considered comparable with human segmentation, but used only 20 scans for testing [16]. The ToothSeg study included greater device diversity, with 1,282 in-house CBCT scans from more than 25 devices and an additional public dataset [17]. Together, these studies strengthen confidence

in technical feasibility, but they do not replace evaluation of workflow performance, errors in critical regions, or downstream treatment decisions.

Because each imaging modality presents distinct clinical tasks and failure mechanisms, validation cannot rely on a single generic design. Table 1 compares the principal tasks, common failure sources, external-validation priorities, and clinically relevant outcomes for panoramic, bitewing, periapical, and CBCT imaging.

Table 1. Modality-specific tasks, common failure sources, external-validation priorities, and key evaluation outcomes. PPV, positive predictive value; NPV, negative predictive value; FOV, field of view; HD95, 95th percentile Hausdorff distance.

Imaging modality	**Main clinical tasks**	**Common failure sources**	**External-validation priorities**	**Key evaluation outcomes**
Panoramic radiography	Primary applications include tooth numbering, anatomical segmentation, and assessment of impacted teeth and periodontal bone loss.	Positioning error, superimposition, distortion, vendor processing, and poor visibility of early lesions can reduce accuracy.	Validation should include multiple sites, detectors, geographic regions, dentitions, and common artifacts.	Reports should include tooth- and lesion-level sensitivity, false positives per image, subclass performance, and calibration.
Bitewing radiography	Primary applications include assessment of proximal caries, restorations, and crestal bone levels.	Overlap, exposure error, cervical burnout, restorations, and unrepresentative disease prevalence can distort performance.	Validation should include multiple sites, detectors, age groups, and disease stages.	Reports should include surface-level sensitivity and specificity, PPV, NPV, unnecessary interventions, and net benefit.
Periapical radiography	Primary applications include assessment of apical radiolucencies and localized endodontic or periodontal findings.	Projection error, overlap, an uncertain reference standard, and previous treatment can reduce reliability.	Validation should include multiple sites, devices, tooth types, treatment states, and time periods.	Reports should include tooth-level errors and effects on reader performance, referral decisions, and treatment.

Imaging modality	Main clinical tasks	Common failure sources	External-validation priorities	Key evaluation outcomes
Cone-beam CT (CBCT)	Primary applications include segmentation, canal mapping, lesion assessment, and linear measurement.	Scanner and protocol differences, FOV, voxel size, metal artifacts, limited labels, and slice leakage can reduce performance.	Validation should include multiple scanners, protocols, sites, geographic regions, and lesion spectra.	Reports should include Dice, surface distance, HD95, millimetre error, critical-region misses, and time saved without an increase in clinically important errors.

4. FROM TASK-SPECIFIC NETWORKS TO MULTIMODAL SYSTEMS AND FOUNDATION MODELS

Most dental imaging systems are task specific: a single-modality image is processed by a convolutional or transformer-based model, which returns a defined class, box, mask, landmark, or measurement. Their inputs, thresholds, and failure modes can be constrained and locally verified, but separate systems may produce inconsistent findings without access to symptoms, history, or prior examinations.

A foundation model is pretrained on broad data and adapted to multiple downstream tasks. A vision-language model (VLM) links image and text representations, while a multimodal system may also integrate structured clinical data. An out-of-distribution (OOD) input differs materially from the data represented during development or intended use. Abstention means deliberately withholding an output when input quality, evidential support, or confidence is inadequate. DentFound was developed with panoramic images from more than 101,000 patients and evaluated diagnosis and report generation across 98 disease and 11 post-treatment categories in multicentre cohorts [18]. Although this scale is notable, autonomous reporting would still require local validation of calibration, OOD behaviour, rare and omitted findings, reporting conventions, and management recommendations.

In a 100-patient study of radiolucent jaw lesions, multimodal large language models were more accurate when computed tomography and histopathology supplemented panoramic imaging; multiple-choice performance also exceeded short-answer performance [19]. Thus, both clinical context and evaluation format influenced measured performance; however, the constrained retrospective sample did not establish safe autonomous diagnosis or a validated clinician-in-the-loop pathway.

MMOral assembled 20,563 panoramic images and approximately 1.3 million instructions to assess numerous VLMs; even the strongest general models achieved only modest overall performance [20]. Fluent language can coexist with weak perception or localization. Evaluation should therefore separate view recognition, tooth or region localization, finding characterization, calibrated uncertainty, and any authorized action. Assigning a plausible lesion label to the wrong tooth remains a clinically meaningful error.

Compared with narrow classifiers, foundation systems add risks from hallucination, prompt sensitivity, report-template shortcuts, benchmark contamination, opaque training provenance, protected-information leakage, prompt injection, and silent version changes. Prompts, retrieval sources, guardrails, language, and report templates become controlled configuration. Broader capability expands, rather than narrows, the set of claims that require evidence.

5. DATASET DESIGN, PROVENANCE, AND REFERENCE STANDARDS

5.1. Design backward from intended use

The dataset protocol should start with the intended user, setting, population, imaging modality and acceptable acquisition parameters, target finding, output, anticipated action, and abstention conditions. “Detects dental disease” is not testable; “highlights possible proximal caries on adult posterior bitewings for dentist review during routine examination” is. The consequences of false negatives, false positives, localization errors, and unavailable outputs should be specified before sampling begins.

Consecutive or quasi-consecutive sampling supports estimates that are representative of clinical practice. Enriched case-control samples help development and rare-failure stress testing but distort prevalence; predictive values do not transfer without accounting for clinical prevalence. Excluding poor images creates a hidden contraindication unless quality criteria are explicit and the deployed system can detect and manage comparable inputs.

5.2. Patient-level data separation and leakage control

Data must be separated at the patient level rather than at the image, tooth, crop, patch, or slice level. Separation must precede augmentation, crop generation, bilateral extraction, or CBCT slice production, and one patient's visits and derivatives should remain together unless a longitudinal design requires otherwise. Merged repositories require duplicate detection. Labels may also leak through annotations, filenames, burned-in markers, scanner signatures, or post-processing.

The internal test set should remain untouched during model and threshold selection; a tuning subset labelled "validation" is not evidence of generalizability. External validation requires an independent acquisition ecosystem, ideally an unseen institution and device family, with geographic, temporal, or protocol holdouts providing complementary stress. In a review of externally evaluated radiology models, 81% of studies showed some performance decrease and almost one quarter a substantial decrease [21]. Dental settings plausibly shift in detector, patient mix, restoration prevalence, positioning, and referral spectrum.

The 2024 CLAIM update asks investigators to distinguish internal testing from external validation and describe data separation, independence, acquisition, preprocessing, reference standards, analysis, and failures [22]. Multicentre training is not external validation when every centre informs tuning or model selection; the external cohort must remain locked until the analysis plan and operating threshold are fixed.

5.3. Provenance and representativeness

Image provenance is a scientific variable. Records should capture site, manufacturer and model, detector or scanner, view, image dimensions, compression, processing software and version, and acquisition period. CBCT records also need field of view, voxel size, reconstruction, available dose-related settings, motion, and metal burden. Scientifically necessary DICOM metadata should be retained under governance for reproducibility and drift detection.

The clinical composition of the dataset should cover age, relevant sex or gender, dentition, tooth and treatment status, implants or appliances, artifacts, disease subtype/severity, and route into the dataset. Race, ethnicity, ancestry, socioeconomic status, and geography may reveal inequity but require lawful, locally meaningful definitions; the absence of these data is a limitation. Site and device are imperfect proxies and cannot prove demographic fairness.

5.4. Reference standards and annotation quality

The reference procedure must fit the model claim. Expert consensus on bitewings can support radiographic lesion detection; it is weaker for a claim about biological lesion activity. CBCT may be a stronger reference for selected periapical findings but is not ethically obtained merely to label every two-dimensional image. Operative findings, histopathology, clinical records, or longitudinal progression may strengthen selected endpoints; however, they may also introduce verification bias because only suspicious cases receive them.

Higher-stakes studies should use multiple qualified readers, independent assessment before consensus, adjudication of disagreement, explicit uncertainty categories, and measured interobserver agreement. Annotators should not see model outputs when establishing the primary reference. Label version, ontology, annotation tool, reader experience, and quality control sampling should be recorded. Public labels require expert re-audit, as demonstrated by errors found when panoramic datasets were combined [6]. The term "reference standard" appropriately acknowledges this uncertainty, whereas "ground truth" does not.

6. VALIDATION ARCHITECTURE AND CLINICALLY MEANINGFUL METRICS

Validation answers different questions at successive stages. Technical verification confirms reproducible processing of specified inputs; internal testing evaluates a locked model within its development ecosystem, and external validation examines transportability. Reader studies assess assisted decisions, silent prospective studies test live workflow without influencing care, controlled studies evaluate patient-relevant benefit and harm, and surveillance examines persistence after release. These stages are complementary; none substitutes for another.

6.1. Match metrics to the task

For classification, report sensitivity, specificity, predictive values, and likelihood ratios at prespecified operating points with confidence intervals. AUROC summarizes ranking across thresholds but can hide unacceptable performance at the clinical threshold; area under the precision-recall curve is useful for rare findings. Thresholds should reflect the harms of missed disease, unnecessary treatment, added imaging, referral, and surveillance rather than maximize a dataset statistic.

Detection requires lesion- or tooth-level sensitivity, false positives per image, and mean average precision at stated intersection-over-union thresholds, with stratification by lesion, tooth, overlap, restoration, and image quality. Segmentation should combine Dice with surface distance, HD95, and clinically meaningful millimetre error. Measurement tasks require absolute error and agreement analysis because correlation can remain high despite systematic bias. Experts should assess generated reports for factuality, localization, omission, hallucination, severity, and unsafe recommendations; lexical similarity is insufficient. Table 2 links each output type to its reference unit, primary measures, and the failure that can change care.

Table 2. Matching dental imaging AI outputs to reference units, performance measures, and clinically meaningful failures

Task and output	Reference unit	Primary measures	Failure that can change care
Classification	Patient-, tooth-, surface-, or lesion-level truth matched to the intended claim	Sensitivity, specificity, predictive values, likelihood ratios, and AUROC or AUPRC at prespecified thresholds, with confidence intervals	AUROC can hide poor performance at the clinical threshold; prevalence changes predictive values.
Detection	Tooth or lesion boxes matched at a stated intersection-over-union threshold	Object-level sensitivity, false positives per image, and mean average precision at a stated threshold	Overall performance can conceal errors by lesion, tooth, overlap, restoration, or image quality.
Segmentation	Expert mask or surface for the intended anatomy or pathology	Dice, surface distance, HD95, and clinically meaningful millimetre error	High overlap can coexist with a clinically important boundary error.
Measurement	Paired reference measurement with repeatability assessed	Absolute error, bias, and agreement limits	Correlation can remain high despite systematic bias.
Generated reports	Expert review against the source image and authorized claim	Factuality, localization, omission, hallucination, severity, and unsafe recommendation	Lexical similarity is insufficient; fluent text can mislocalize a finding or overstate the claim.

6.2. Calibration and clinical utility

Discrimination is not calibration: among comparable cases assigned a probability of 0.8, about 80% should have the finding. Assessment should include calibration plots, calibration intercept and slope, and a proper scoring rule such as the Brier score, including for relevant subgroups [23]. A confidence score should not be presented as probability without this evidence. Recalibration after documented shift creates a changed version requiring verification.

Clinical utility is determined by how predictions influence decisions and actions. Decision-curve analysis can estimate net benefit, provided harm weighting is explicit. Reader studies should measure changes in errors, time,

confidence, referral, imaging, and treatment. Bitewing evidence shows why: higher sensitivity can coexist with lower positive predictive value or more false positives [8, 9]. Average accuracy cannot determine whether that trade-off is beneficial.

6.3. Hidden stratification and error auditing

Subgroup reporting should prespecify plausible axes and search for unexpected error clusters. Clinically meaningful imaging subclasses have differed by more than 20 percentage points despite acceptable parent-category performance [24]. Relevant dental subclasses include enamel versus dentine caries, horizontal versus vertical bone loss, primary versus permanent dentition, anterior loop versus canal body, treatment status, rare pathology, metal, and motion.

Each subgroup estimate needs a denominator and uncertainty interval; very small samples warrant "insufficient evidence," not equivalence. Error review should identify whether failures arise from acquisition, annotation, anatomy, prevalence, thresholds, interface, or workflow. Appropriate responses may include restricting the intended use, collecting targeted data, revising the reference standard, recalibrating the model, redesigning the interface, or requiring abstention; retraining is not the only remedy.

Before deployment, evidence should progress from dataset fitness and locked testing through external and prospective clinical evaluation, and at each stage the evidence required should be separated from the decision gate that permits progression.

7. BIAS, FAIRNESS, UNCERTAINTY, AND ROBUSTNESS

Bias can arise before model development: access affects who is imaged, referral affects severity, equipment and operators affect quality, suspicious cases more often receive definitive reference procedures, and annotators may apply local thresholds. Excluding artifacts, children, extensive restorations, or rare pathology may produce a clean benchmark that is unsafe for the intended population.

Fairness cannot be reduced to a single universal mathematical constraint: equal sensitivity, false-positive rates, calibration, and treatment benefit can conflict when prevalence and consequences differ. The objective is equitable benefit and acceptable harm within intended use. Studies should report composition and performance across clinically relevant demographic, dentition, site, device, disease, treatment, artifact, and intersectional groups. Without ethically collected demographic attributes, overall performance cannot establish fairness.

Uncertainty must trigger action. Input checks should detect unsupported formats, missing views, severe cropping, low resolution, extreme exposure, motion, metal, and processing outside specification. Prediction uncertainty may reflect intrinsic ambiguity rather than distribution shift, while OOD detectors can themselves fail. Safe operation therefore combines quality control, calibrated confidence where possible, contraindications, clinician verification, and abstention or escalation. A report generator must be able to state that evidence is insufficient.

Robustness testing should vary plausible compression, brightness, noise, rotation, resolution, field of view, metal burden, and device processing without changing the label. Synthetic perturbations do not replace naturally shifted external cohorts. CBCT testing should cover scanner/protocol combinations and critical anatomy; multimodal testing should cover missing or contradictory text, chart errors, prompt variation, retrieval failure, and prompt injection.

Model output should complement, not replace, clinical evidence. Inadequate input, use outside the intended population, or conflict with clinical findings should default to qualified review. Logging these cases creates a governed failure registry without silently expanding the indication.

8. EXPLAINABILITY AND INTERACTION BETWEEN CLINICIANS AND AI

Explainability is valuable when it supports localization checks, uncertainty interpretation, detection of unsupported output, or version traceability. A box or mask may be more actionable than a post hoc saliency map. Explanations should be tested for localization accuracy, stability, clinician comprehension, and their effects on decisions and automation bias.

Saliency maps are not causal proof. Several methods showed low sensitivity and robustness in a radiology study, and heatmaps could remain stable despite materially worse model discrimination [25]. An explanation that appears plausible to a clinician is not necessarily faithful to the model; counterfactual tests, failures, uncertainty, quality flags, provenance, audit trails, and use limits may be more informative.

Clinicians using AI must be measured directly. Reader studies should include intended users, randomize or counterbalance assistance, use washout where appropriate, prespecify thresholds and endpoints, and record time, confidence, errors, overrides, and downstream decisions. ADEPT and periapical-radiolucency studies show that assistance can shift the error balance and affect junior and senior readers differently [9, 11]. Human performance is a comparator, not an infallible reference.

Interface design affects safety: alerts can cause fatigue, overlays can anchor attention, scores can be mistaken for calibrated probabilities, and generated recommendations can exceed the authorized claim. The interface should display the source image, exact tooth or region, missing inputs, and the distinction between observation and recommendation, while allowing documented override. Early evaluation should follow guidance such as DECIDE-AI on workflow, human factors, safety, and iterative assessment before a definitive trial [26].

9. CLINICAL TRANSLATION, MONITORING, AND CONTROLLED CHANGE

A hazard-based deployment plan should assign every anticipated harm an owner, control, evidence artifact, and stop criterion. The FUTURE-AI consensus organizes this work around fairness, universality, traceability, usability, robustness, and explainability [27].

The first live deployment step should generally be a protocol-defined silent prospective evaluation: the model receives real inputs without influencing care. Prespecified measures should cover interoperability, latency, missing or unsupported inputs, OOD frequency, prevalence, calibration, and failure recovery. The study needs a sample-size rationale and completion criteria; it is not informal observation.

Controlled impact evaluation may then use a counterbalanced reader study, prospective paired comparison, cluster design, or randomized trial. Endpoints should include missed disease, false intervention, added radiation or referral, reading time, confidence, equity, and patient-relevant outcomes; any surrogate limitation must be explicit. Deployment can proceed in stages for restricted users, sites, or indications, with enhanced review and rollback.

The IMDRF SaMD framework separates valid clinical association, analytical or technical performance, and clinical performance [28]. Good machine learning practice principles also emphasize multidisciplinary expertise, representative datasets, independent tests, clinician-AI performance, user information, monitoring, and controlled retraining [29]. FDA guidance similarly requires planned modifications, a protocol, and impact assessment in a predetermined change control plan [30]. Thresholds, training distributions, foundation models, prompts, retrieval corpora, and interfaces can all alter clinical behaviour.

After release, monitoring should track drift, calibration, operating-point errors when reference outcomes become available, subgroup outcomes, OOD and abstention rates, overrides, complaints, incidents, downtime, cybersecurity, and configuration versions. Prespecified triggers should govern investigation,

pause, rollback, correction, and revalidation. Feedback cases require lawful governance and must not flow automatically into retraining.

Trustworthy deployment is a continuous evidence cycle that links a defined clinical purpose with governed data, locked testing, external and prospective validation, controlled use, monitoring, and revalidation. The cross-stage controls reflect a sociotechnical system consistent with WHO ethical principles for health AI [31] and the NIST functions Govern, Map, Measure, and Manage [32].

To translate this lifecycle into auditable practice, Table 3 aligns each trustworthiness domain with its warning signals, primary control, and the evidence that should be retained.

Table 3. Trustworthiness controls and auditable evidence for clinical dental imaging AI. OOD, out-of-distribution.

Trust domain	Warning signal	Primary control	Evidence to retain
Bias and fairness	A subgroup shows lower sensitivity or a higher false-positive rate.	Use representative sampling, intersectional analysis, and harm-weighted thresholds.	Retain subgroup metrics, denominators, and confidence intervals.
Calibration and uncertainty	The model is overconfident or loses calibration at a new site.	Assess calibration and input quality; permit abstention; govern any recalibration.	Retain calibration plots, intercept and slope estimates, and OOD or abstention logs.
Robustness	Performance declines across devices, protocols, artifacts, or prompts.	Use naturally shifted external cohorts and prespecified stress tests.	Retain acquisition and configuration matrices and a structured failure summary.
Explainability	A localization or rationale is unstable, misleading, or unrelated to the output.	Test fidelity and stability, and assess clinician comprehension.	Retain the validation report and representative failure examples.
Human factors	Automation bias, alert fatigue, or unsafe overrides emerge.	Conduct usability and counterbalanced reader studies; define escalation rules.	Retain combined-performance and usability reports.

Trust domain	Warning signal	Primary control	Evidence to retain
Privacy and security	Data leakage, prompt injection, or unauthorized access occurs.	Minimize data, control access, de-identify records, and conduct security testing.	Retain the data-flow map, security report, and audit log.
Traceability	The deployed data, model, prompt, or threshold version cannot be identified.	Use versioned provenance and configuration logging.	Retain the dataset record, model card, and deployment manifest.
Monitoring and change	Performance drifts or regresses after release or an update.	Define investigation triggers, rollback rules, impact assessment, and revalidation requirements.	Retain the monitoring plan and signed change report.

10. REGULATION, STANDARDS, PRIVACY, AND REPORTING

Regulatory requirements depend on intended use, risk, and jurisdiction. Under Regulation (EU) 2024/1689, an AI system can be high risk when it is a safety component of a product, or is itself a product, covered by the Union harmonization legislation listed in Annex I and the product is required to undergo third-party conformity assessment under that legislation. Relevant duties include lifecycle risk management, data governance, technical documentation, logging, transparency, human oversight, accuracy, robustness, cybersecurity, and post-market monitoring [33]. Application is phased, so the operative duties and dates must be checked against the law in force for the product and planned deployment date rather than treated as fully applicable at all times. MDCG 2019-11 rev.1 supports qualification and classification of software under the Medical Device Regulation and In Vitro Diagnostic Medical Device Regulation [34]. The interaction is product specific; the label “AI” does not determine classification by itself.

In the United States, an entry on FDA's AI-enabled medical device list indicates authorization of a defined device and intended use, not blanket endorsement of a model family or proof of effectiveness in every clinic. The list is periodically updated and FDA states that it is not comprehensive [35]. Local governance should verify the authorized version, input, user, output, limitations, and update pathway. Research prototypes and general-purpose foundation models should not be represented as cleared clinical devices without an applicable authorization.

ISO 14971 provides a general process for identifying hazards, estimating and controlling risk, and evaluating residual risk across the medical device lifecycle [36]. ISO/TS 24971-2:2026 adds guidance for machine learning in medical devices but explicitly excludes large language models and generative AI; it should not be presented as a complete standard for multimodal report generators [37]. IEC 62304 addresses medical device software lifecycle processes [38]. Standards support a quality system; conformance does not replace clinical evidence.

Privacy and security require controls tailored to imaging data. DICOM headers can contain identifiers and acquisition provenance; pixels may contain burned-in text or recognizable facial features in volumetric data. DICOM PS3.15 defines confidentiality profiles and options for identifiers, descriptors, structured content, visual features, and unique identifiers [39]. The correct profile depends on the research or clinical purpose. Under the General Data Protection Regulation, health data receive special-category protection [40]. HIPAA guidance recognizes Safe Harbor and Expert Determination routes for de-identification and notes that residual re-identification risk is not literally zero [41]. Data minimization should therefore coexist with controlled retention of scientifically necessary provenance.

Reporting guidance should be selected according to study design. CLAIM 2024 applies broadly to AI in medical imaging [22]. STARD-AI addresses AI diagnostic accuracy studies [42]; TRIPOD+AI addresses prediction model development and evaluation [43]; PROBAST+AI supports structured assessment of quality, risk of bias, and applicability [44]; and CONSORT-AI applies when AI interventions are evaluated in clinical trials [45]. Ethical guidance for dental AI highlights autonomy, privacy, fairness, transparency, accountability, and professional responsibility [46]. A checklist improves reporting, but cannot repair biased sampling or an inappropriate reference standard. Registration, a prespecified analysis plan, accessible code or model details where possible, and complete reporting of exclusions and failures make independent scrutiny more credible.

11. CONCLUSION

Dental and maxillofacial imaging AI has demonstrated credible technical strengths, but evidence remains uneven: retrospective task performance is more common than independent external validation, prospective workflow evaluation, controlled clinical impact, and post-market outcomes. Multimodal capability does not remove these gaps.

Trustworthiness is a property of the complete clinical system. A specified version should be judged by whether defined clinicians can use it on intended patients and devices to improve decisions without unacceptable harm, and

whether the organization can detect deterioration and govern change. This provides a practical standard for earning and retaining clinical trust.

REFERENCES

[1]. Schwendicke, F., Samek, W., and Krois, J. (2020). Artificial intelligence in dentistry: Chances and challenges. Journal of Dental Research, 99(7), 769–774. https://doi.org/10.1177/0022034520915714

[2]. Schwendicke, F., Singh, T., Lee, J. H., Gaudin, R., Chaurasia, A., Wiegand, T., Uribe, S., and Krois, J. (2021). Artificial intelligence in dental research: Checklist for authors, reviewers, readers. Journal of Dentistry, 107, 103610. https://doi.org/10.1016/j.jdent.2021.103610

[3]. Cha, J. Y., Yoon, H. I., Yeo, I. S., Huh, K. H., and Han, J. S. (2021). Panoptic segmentation on panoramic radiographs: Deep learning-based segmentation of various structures including maxillary sinus and mandibular canal. Journal of Clinical Medicine, 10(12), 2577. https://doi.org/10.3390/jcm10122577

[4]. Xu, M., Wu, Y., Xu, Z., Ding, P., Bai, H., and Deng, X. (2023). Robust automated teeth identification from dental radiographs using deep learning. Journal of Dentistry, 136, 104607. https://doi.org/10.1016/j.jdent.2023.104607

[5]. Kurt-Bayrakdar, S., Bayrakdar, İ. Ş., Yavuz, M. B., Sali, N., Çelik, Ö., Köse, O., Uzun Saylan, B. C., Kuleli, B., Jagtap, R., and Orhan, K. (2024). Detection of periodontal bone loss patterns and furcation defects from panoramic radiographs using deep learning algorithm: A retrospective study. BMC Oral Health, 24(1), 155. https://doi.org/10.1186/s12903-024-03896-5

[6]. van Nistelrooij, N., Ghoul, K. E., Xi, T., Saha, A., Kempers, S., Cenci, M., Loomans, B., Flügge, T., van Ginneken, B., and Vinayahalingam, S. (2024). Combining public datasets for automated tooth assessment in panoramic radiographs. BMC Oral Health, 24(1), 387. https://doi.org/10.1186/s12903-024-04129-5

[7]. Cantu, A. G., Gehrung, S., Krois, J., Chaurasia, A., Rossi, J. G., Gaudin, R., Elhennawy, K., and Schwendicke, F. (2020). Detecting caries lesions of different radiographic extension on bitewings using deep learning. Journal of Dentistry, 100, 103425. https://doi.org/10.1016/j.jdent.2020.103425

[8]. Lee, S., Oh, S. I., Jo, J., Kang, S., Shin, Y., and Park, J. W. (2021). Deep learning for early dental caries detection in bitewing radiographs. Scientific Reports, 11(1), 16807. https://doi.org/10.1038/s41598-021-96368-7

[9]. Devlin, H., Williams, T., Graham, J., and Ashley, M. (2021). The ADEPT study: A comparative study of dentists' ability to detect enamel-only proximal caries in bitewing radiographs with and without the use of AssistDent artificial intelligence software. British Dental Journal, 231(8), 481–485. https://doi.org/10.1038/s41415-021-3526-6

[10]. Farooqi, O. A., Fru, G. A., Gong, Y. M., Oswald, L. E., and DeNucci, D. J. (2026). Evaluation of artificial intelligence-based clinical decision support systems for caries and periodontal bone loss: An external validation study. Journal of the American Dental Association, 157(6), 611–618. https://doi.org/10.1016/j.adaj.2025.11.005

[11]. Pul, U., Tichy, A., Pitchika, V., and Schwendicke, F. (2025). Impact of artificial intelligence assistance on diagnosing periapical radiolucencies: A randomized

controlled trial. Journal of Dentistry, 160, 105868. https://doi.org/10.1016/j.jdent.2025.105868

[12]. Wang, H., Minnema, J., Batenburg, K. J., Forouzanfar, T., Hu, F. J., and Wu, G. (2021). Multiclass CBCT image segmentation for orthodontics with deep learning. Journal of Dental Research, 100(9), 943–949. https://doi.org/10.1177/00220345211005338

[13]. Yeshua, T., Ladyzhensky, S., Abu-Nasser, A., Abdalla-Aslan, R., Boharon, T., Itzhak-Pur, A., Alexander, A., Chaurasia, A., Cohen, A., Sosna, J., Leichter, I., and Nadler, C. (2023). Deep learning for detection and 3D segmentation of maxillofacial bone lesions in cone beam CT. European Radiology, 33(11), 7507–7518. https://doi.org/10.1007/s00330-023-09726-6

[14]. Fu, W. T., Zhu, Q. K., Li, N., Wang, Y. Q., Deng, S. L., Chen, H. P., Shen, J., Meng, L. Y., and Bian, Z. (2024). Clinically oriented CBCT periapical lesion evaluation via 3D CNN algorithm. Journal of Dental Research, 103(1), 5–12. https://doi.org/10.1177/00220345231201793

[15]. Ntovas, P., Marchand, L., Finkelman, M., Revilla-León, M., and Att, W. (2024). Accuracy of artificial intelligence-based segmentation of the mandibular canal in CBCT. Clinical Oral Implants Research, 35(9), 1163–1171. https://doi.org/10.1111/clr.14307

[16]. Semper-Hogg, W., Rau, A., Fuessinger, M. A., Zimmermann, S., Bamberg, F., Metzger, M. C., Schmelzeisen, R., Rau, S., Reisert, M., and Russe, M. F. (2025). Deep learning-based segmentation of the mandibular canals in cone-beam CT reaches human-level performance. Dentomaxillofacial Radiology, 54(4), 279–285. https://doi.org/10.1093/dmfr/twae069

[17]. van Nistelrooij, N., Krämer, L., Kempers, S., Beyer, M., Bolelli, F., Xi, T., Bergé, S., Heiland, M., Maier-Hein, K. H., Vinayahalingam, S., and Isensee, F. (2026). ToothSeg: Robust tooth instance segmentation and numbering in CBCT using deep learning and self-correction. IEEE Journal of Biomedical and Health Informatics. Advance online publication. https://doi.org/10.1109/JBHI.2025.3650444

[18]. Zhu, Q., Lin, Y., Fu, W., Tang, W., Li, J., Zhang, Y., Li, B., Guo, X., Wang, F., Qi, H., Sun, C., Zhu, X., Liu, Z., He, L., Zheng, Z., Du, B., Yang, J., Bian, Z., and Meng, L. (2026). Towards clinical-level interpretation of dental panoramic radiography using an instance-guided vision-language model. Nature Biomedical Engineering. Advance online publication. https://doi.org/10.1038/s41551-026-01713-8 (Author correction published July 8, 2026: https://doi.org/10.1038/s41551-026-01759-8)

[19]. Kim, K., and Kim, B. C. (2025). Diagnostic performance of large language models in multimodal analysis of radiolucent jaw lesions. International Dental Journal, 75(6), 103910. https://doi.org/10.1016/j.identj.2025.103910

[20]. Hao, J., Fan, Y., Sun, Y., Guo, K., Lin, L., Yang, J., Ai, Q. Y. H., Wong, L. M., Tang, H., and Hung, K. F. (2025). Towards better dental AI: A multimodal benchmark and instruction dataset for panoramic X-ray analysis. Advances in Neural Information Processing Systems, 38. https://proceedings.neurips.cc/paper_files/paper/2025/hash/3797b3a3c939156757857cf1b9afde3c-Abstract-Datasets_and_Benchmarks_Track.html

[21]. Yu, A. C., Mohajer, B., and Eng, J. (2022). External validation of deep learning algorithms for radiologic diagnosis: A systematic review. Radiology: Artificial Intelligence, 4(3), e210064. https://doi.org/10.1148/ryai.210064

[22]. Tejani, A. S., Klontzas, M. E., Gatti, A. A., Mongan, J. T., Moy, L., Park, S. H., Kahn, C. E., Abbara, S., Afat, S., Anazodo, U. C., Andreychenko, A., Asselbergs, F. W., Badano, A., Baessler, B., Bold, B., Bisdas, S., Brismar, T. B., Cacciamani, G. E., Carrino, J. A., … Zins, M. (2024). Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. Radiology: Artificial Intelligence, 6(4), e240300. https://doi.org/10.1148/ryai.240300

[23]. Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., Steyerberg, E. W., and Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17(1), 230. https://doi.org/10.1186/s12916-019-1466-7

[24]. Oakden-Rayner, L., Dunnmon, J., Carneiro, G., and Ré, C. (2020). Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. Proceedings of the ACM Conference on Health, Inference, and Learning, 151–159. https://doi.org/10.1145/3368555.3384468

[25]. Zhang, J., Chao, H., Dasegowda, G., Wang, G., Kalra, M. K., and Yan, P. (2024). Revisiting the trustworthiness of saliency methods in radiology AI. Radiology: Artificial Intelligence, 6(1), e220221. https://doi.org/10.1148/ryai.220221

[26]. Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., … DECIDE-AI expert group. (2022). Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ, 377, e070904. https://doi.org/10.1136/bmj-2022-070904

[27]. Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., Weicken, E., Asselbergs, F. W., Prior, F., Collins, G. S., Kaissis, G., Tsakou, G., Buvat, I., Kalpathy-Cramer, J., Mongan, J., Schnabel, J. A., Kushibar, K., Riklund, K., Marias, K., … FUTURE-AI Consortium. (2025). FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ, 388, e081554. https://doi.org/10.1136/bmj-2024-081554

[28]. International Medical Device Regulators Forum. (2017). Software as a medical device (SaMD): Clinical evaluation (IMDRF/SaMD WG/N41FINAL:2017). https://www.imdrf.org/documents/software-medical-device-samd-clinical-evaluation

[29]. International Medical Device Regulators Forum. (2025). Good machine learning practice for medical device development: Guiding principles (IMDRF/AIML WG/N88 FINAL:2025). https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles

[30]. U.S. Food and Drug Administration. (2025). Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions (Guidance for industry and Food and Drug Administration staff). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence

[31]. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. https://www.who.int/publications/i/item/9789240029200

[32]. Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1
[33]. European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
[34]. European Commission Medical Device Coordination Group. (2025). MDCG 2019-11 rev.1: Guidance on qualification and classification of software in Regulation (EU) 2017/745—MDR and Regulation (EU) 2017/746—IVDR. https://health.ec.europa.eu/latest-updates/update-mdcg-2019-11-rev1-qualification-and-classification-software-regulation-eu-2017745-and-2025-06-17_en
[35]. U.S. Food and Drug Administration. (n.d.). Artificial intelligence-enabled medical devices. Retrieved July 21, 2026, from https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices
[36]. International Organization for Standardization. (2019). ISO 14971:2019 Medical devices—Application of risk management to medical devices. https://www.iso.org/standard/72704.html
[37]. International Organization for Standardization. (2026). ISO/TS 24971-2:2026 Medical devices—Guidance on the application of ISO 14971—Part 2: Machine learning in artificial intelligence. https://www.iso.org/standard/24971-2
[38]. International Electrotechnical Commission. (2015). IEC 62304:2006+AMD1:2015 Medical device software—Software life cycle processes. https://webstore.iec.ch/en/publication/6792
[39]. DICOM Standards Committee. (2026). DICOM PS3.15 2026b: Security and system management profiles. https://dicom.nema.org/medical/Dicom/2026b/output/pdf/part15.pdf
[40]. European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679
[41]. U.S. Department of Health and Human Services, Office for Civil Rights. (n.d.). Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. Retrieved July 21, 2026, from https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
[42]. Sounderajah, V., Guni, A., Liu, X., Collins, G. S., Karthikesalingam, A., Markar, S. R., Golub, R. M., Denniston, A. K., Shetty, S., Moher, D., Bossuyt, P. M., Darzi, A., Ashrafian, H., and STARD-AI Steering Committee. (2025). The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nature Medicine, 31(10), 3283–3289. https://doi.org/10.1038/s41591-025-03953-8
[43]. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A. L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., … Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use

regression or machine learning methods. BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378

[44]. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., … van Smeden, M. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 388, e082505. https://doi.org/10.1136/bmj-2024-082505

[45]. Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., and SPIRIT-AI and CONSORT-AI Working Group. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nature Medicine, 26(9), 1364–1374. https://doi.org/10.1038/s41591-020-1034-x

[46]. Rokhshad, R., Ducret, M., Chaurasia, A., Karteva, T., Radenkovic, M., Roganovic, J., Hamdan, M., Mohammad-Rahimi, H., Krois, J., Lahoud, P., and Schwendicke, F. (2023). Ethical considerations on artificial intelligence in dentistry: A framework and checklist. Journal of Dentistry, 135, 104593. https://doi.org/10.1016/j.jdent.2023.104593

Biologically Based Vital Pulp Therapy in Immature Permanent Teeth: Diagnostic Thresholds, Calcium Silicate Biomaterials, and Long-Term Outcomes

Hilal YILMAZ YALIM[1]*

[1] Hilal Yılmaz Yalım, Research Assistant; Department of Pediatric Dentistry, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Türkiye; E-mail: dt.hilalyilmaz95@gmail.com; ORCID: 0009-0006-2694-057X. Corresponding author.

ABSTRACT

Preserving pulpal vitality in an immature permanent tooth allows physiological root development to continue, increasing root length, thickening fragile dentinal walls, and improving structural prognosis. The principal diagnostic challenge is that symptoms, sensibility tests, and radiographs infer rather than directly measure pulpal inflammation. This focused narrative review synthesizes clinical practice guidelines, systematic reviews, randomized trials, and selected mechanistic evidence on deep caries and traumatic pulp exposure in vital immature permanent teeth. Selective caries removal is preferred when a radiographic dentine barrier remains and the pulp is judged normal or reversibly inflamed. When exposure occurs, direct pulp capping, partial pulpotomy, or full pulpotomy can provide favorable short-term outcomes in appropriately selected teeth, provided that asepsis, hemorrhage control, a conventional hydraulic calcium silicate cement, and a durable coronal seal are achieved. Pooled success rates should not be interpreted as high-certainty evidence because definitions, populations, procedures, and outcomes vary, while long-term data specific to immature teeth remain limited. Diagnostic decisions should therefore be staged and revised after direct assessment of exposed tissue. Uncontrolled hemorrhage, suppuration, necrotic tissue, or infection beyond the pulp redirects care away from vital pulp therapy. Follow-up should evaluate symptoms, restoration integrity, periapical health, and continued root maturation rather than apical closure alone. Regenerative endodontic procedures form a separate pathway for necrotic immature teeth. The proposed decision pathway relates tissue preservation to disease extent and the certainty of available evidence.

Keywords: Immature permanent tooth, apexogenesis, vital pulp therapy, partial pulpotomy, calcium silicate cement, deep caries, dental trauma

1. INTRODUCTION

An immature permanent tooth begins to function before root formation is complete. Its open apex, thin radicular dentine, and relatively rich pulpal blood supply create both therapeutic opportunity and structural vulnerability. Loss of vitality arrests physiological root maturation and may leave a tooth that is difficult to disinfect, obturate, and protect from fracture. Successful vital pulp therapy (VPT), by contrast, preserves enough healthy pulp to support apexogenesis, the continued physiological formation of root dentine and apical tissues. The aim is therefore not simply to retain a positive sensibility response, but to maintain a biologically competent pulp in a symptom-free, functional, restorable tooth without progressive apical disease.

The treatment spectrum extends from indirect pulp treatment, more precisely described in permanent teeth as selective caries removal followed by a sealed restoration, to direct pulp capping, partial pulpotomy, and full pulpotomy. Although these procedures remove different amounts of inflamed coronal tissue, they share four requirements: a defensible diagnosis, controlled asepsis, a biocompatible pulp-contact material, and an effective coronal seal. Modern hydraulic calcium silicate cements have broadened the indications for VPT, while advances in caries biology have challenged the assumption that all softened dentine must be removed at the first visit.

This apparently straightforward sequence contains substantial uncertainty. Cold and electric tests assess neural response rather than pulpal blood flow or histological inflammation, and recently traumatized or immature teeth may respond inconsistently. Symptoms historically classified as "irreversible pulpitis" do not invariably indicate irreparable inflammation throughout the pulp. Bleeding behavior provides useful intraoperative information after exposure, but it has not been validated as an independent histological test. The treatment evidence is also heterogeneous because studies combine mature and immature teeth, carious and traumatic exposures, different cements, and varying definitions of success.

This chapter addresses three related questions: which diagnostic threshold supports an attempt at VPT in an immature permanent tooth; how tissue preservation, pulpotomy depth, hemorrhage control, and biomaterial choice should be matched to the clinical presentation; and how success should be evaluated over time. The discussion focuses on deep and extremely deep caries and traumatic pulp exposure. Regenerative endodontic procedures are considered only to define the boundary between treatment pathways for vital and necrotic immature teeth.

2. FOCUSED NARRATIVE REVIEW METHOD

Targeted searches of PubMed/MEDLINE and the official websites of professional organizations were updated through 23 July 2026. Search concepts combined immature permanent teeth, vital pulp therapy, indirect pulp treatment, selective caries removal, direct pulp capping, partial pulpotomy, full pulpotomy, apexogenesis, calcium silicate cement, traumatic pulp exposure, diagnosis, and outcome. Priority was given to evidence-based clinical practice guidelines, international consensus documents, systematic reviews and meta-analyses, randomized or prospective studies of permanent teeth, and investigations reporting root-development or long-term outcomes. Human histological evidence and authoritative position statements were included when they clarified biological mechanisms or areas of disagreement.

Evidence was compared according to pulp status, cause of exposure, root maturity, operative method, material, follow-up duration, and definition of success. Guideline recommendations were not treated as equivalent to certainty of evidence; when recommendations differed, their populations, clinical questions, search dates, and evidential bases were examined. This work was not a systematic review: no protocol was registered, screening was not conducted independently by two reviewers, and study-level risk of bias was not recalculated. Numerical estimates are therefore presented as contextual evidence rather than as a new pooled analysis or a guarantee for an individual patient.

3. WHY ROOT IMMATURITY CHANGES THE TREATMENT OBJECTIVE

Root development depends on the survival and coordinated activity of the radicular pulp, odontoblast lineage, epithelial root sheath, and apical tissues. When the coronal pulp is injured but viable radicular tissue remains, an appropriately selected VPT procedure can remove the most inflamed tissue while preserving the biological apparatus required for continued dentine deposition. Favorable maturation is expressed clinically as increasing root length, thickening dentinal walls, progressive apical narrowing, and maintenance of healthy periodontal and periapical tissues. These changes may improve fracture resistance and simplify any later endodontic intervention.

The vascularity and cellularity of an immature pulp may favor healing, but root immaturity does not neutralize infection. Persistent biofilm, procedural contamination, an inadequate seal, or inflammation extending into the radicular pulp can compromise an otherwise conservative procedure. An open apex alone also does not establish vitality: a necrotic immature tooth may retain a wide apical opening and incomplete root walls, yet it requires a different biological and therapeutic pathway.

Evidence specific to immature permanent teeth is more limited than some clinical claims suggest. A systematic review of immature permanent posterior teeth reported high pooled success for direct pulp capping, indirect pulp treatment, and partial pulpotomy, but emphasized the small number and limited quality of randomized trials [1]. In the 2025 American Academy of Pediatric Dentistry (AAPD) guideline, root maturity did not significantly modify the 24-month success of calcium silicate partial or full pulpotomy in teeth diagnosed with normal pulp or reversible pulpitis. Because this finding was indirect and supported by low-certainty evidence, it should not be interpreted as proof that mature and immature pulps behave identically [2].

The appropriate principle is therefore tissue preservation without therapeutic delay. Healthy tissue should be retained when the clinical presentation and intraoperative findings support healing; VPT should not be continued when necrosis, suppuration, uncontrolled hemorrhage, or infection extending to the periapical tissues is evident.

4. DIAGNOSTIC THRESHOLDS: WHAT CAN AND CANNOT BE KNOWN

4.1 A staged rather than binary diagnosis

Pretreatment pulp diagnosis is a working hypothesis derived from several imperfect findings. Assessment should include the pain history and pattern, swelling or a sinus tract, percussion and palpation tenderness, mobility, and periodontal probing when indicated. Sensibility responses should be compared with control teeth, while radiographs should be examined for caries depth, restorability, root development, periodontal ligament space, and periapical status. The AAPD conditionally recommends interpreting cold and electric pulp tests together with symptoms, clinical findings, and radiographs, although the supporting evidence is of very low certainty [2]. The International Association of Paediatric Dentistry (IAPD) similarly concluded that current tests cannot reliably determine the extent of inflammation or predict healing [3].

Sensibility is not synonymous with vitality. Cold and electric tests stimulate sensory pathways but do not measure pulpal blood flow. A systematic review found low diagnostic effectiveness for determining the true inflammatory state of the pulp because accurate, reproducible clinical evidence remains limited [4]. False-negative responses are particularly plausible in recently erupted teeth, incompletely innervated immature teeth, and teeth tested soon after trauma. A response that differs from control teeth is informative, but no single result should determine treatment. After dental trauma, serial testing interpreted alongside clinical and radiographic findings is more reliable than an isolated early response [5].

Pain severity also has limited anatomical specificity. Spontaneous, nocturnal, or lingering pain raises concern that inflammation is extensive, but does not prove that every part of the pulp is incapable of healing. Randomized trials and systematic reviews indicate that pulpotomy can succeed in selected permanent teeth with symptoms traditionally assigned to irreversible pulpitis, although most evidence concerns mature teeth and is of low or very low certainty [6, 7]. In an immature tooth, these symptoms should lower the threshold for removing additional coronal tissue and for referral when case complexity exceeds the operator's experience. They should not be used alone to establish either certain vitality or certain irreversibility.

4.2 Lesion depth and periapical status

The radiographic distinction between deep and extremely deep caries helps define the initial strategy. In the AAPD guideline, deep caries extends into the inner third or quarter of dentine while leaving a radiographically detectable dentine barrier; extremely deep caries extends through the full dentine thickness without a visible barrier [2]. These are clinical decision categories rather than microscopic measurements. Projection geometry and lesion location may cause radiographs to understate or overstate the amount of remaining dentine.

In a vital tooth with deep caries, no spontaneous or lingering symptoms, and normal periapical tissues, avoiding pulp exposure is biologically coherent. Extremely deep caries, or deep caries accompanied by spontaneous, nocturnal, or lingering pain, increases the likelihood of established exposure and extensive coronal inflammation. For this latter group, the 2025 AAPD guideline recommends nonselective removal to expose and directly assess the pulp, provided that the tooth is restorable and the periapical appearance does not indicate infection beyond the pulp [2]. This qualification explains why lesion depth and symptom pattern must be interpreted together rather than applying selective excavation uniformly.

Periapical radiolucency, swelling, a sinus tract, or suppuration strongly redirects treatment away from VPT. Percussion tenderness requires context, particularly after trauma or altered occlusion. Nevertheless, pulpal infection with periapical involvement falls outside the AAPD indication for pulpotomy, as do exposed necrotic tissue and hemorrhage that remains uncontrolled after appropriate tissue removal and disinfection.

4.3 Intraoperative reassessment

Direct inspection refines the diagnosis. Under isolation and, when available, magnification, the operator determines whether tissue is present and vital, whether purulence or necrosis is visible, whether bleeding arises from a defined vital surface, and whether hemorrhage can be controlled after inflamed tissue is removed. Bleeding color and duration are useful pragmatic signals, but neither has a validated one-to-one relationship with histological inflammation. The IAPD consensus therefore advises interpreting bleeding duration together with the other findings rather than as an independent criterion [3].

The AAPD conditionally associates hemostasis achieved within six minutes with greater pulpotomy success, based on low-certainty evidence [2]. This threshold is an operational guide rather than a biological stopwatch. Technique, wound size, irrigant and concentration, depth of tissue removal,

systemic factors, and measurement method can all affect bleeding. Persistent profuse hemorrhage after adequate coronal tissue removal, or absent bleeding from tissue suspected to be necrotic, should prompt diagnostic reassessment rather than repeated capping.

These limitations support a probability-based interpretation of the examination. Table 1 summarizes how each diagnostic pattern informs initial management while emphasizing that no single finding is determinative.

Table 1. Diagnostic findings and their implications for a vital immature permanent tooth

Clinical finding	Clinical interpretation	Initial management	Key limitation
Deep caries with a radiographically visible dentine barrier; no pain or pain only on provocation; normal periapical tissues	Pulp exposure may be avoidable, and a normal or reversibly inflamed pulp remains plausible.	Use selective caries removal and place a durable coronal seal.	Radiographic depth is approximate, and absence of symptoms does not confirm normal histology.
Extremely deep caries, or deep caries with spontaneous, nocturnal, or lingering pain	Established exposure or extensive coronal inflammation is more likely.	If the tooth is restorable and periapical tissues are normal, remove caries to assess the pulp and select the appropriate VPT depth.	Symptom categories do not map precisely to histology, and evidence for symptomatic cases is less certain.
Recent traumatic exposure in an immature permanent tooth	The pulp may retain substantial healing capacity despite exposure.	Preserve vitality, usually by partial pulpotomy when the wound is contaminated or treatment is delayed.	Early sensibility tests may be falsely negative, and concomitant luxation increases uncertainty.
Vital bleeding that becomes controlled after affected coronal tissue is removed	The remaining tissue may be suitable for placement of a pulp-contact material.	Apply a conventional hydraulic calcium silicate cement and establish a durable seal.	Hemostasis time is a pragmatic guide, not an independent histological test.
Suppuration, visible necrosis, uncontrolled hemorrhage, swelling, sinus tract, or infection-related periapical involvement	Pulpal or periradicular disease extends beyond the intended indication for VPT.	Stop the VPT pathway and provide, or refer for, appropriate treatment of the nonvital tooth.	Trauma-related percussion tenderness and early radiographic uncertainty still require clinical interpretation.

5. MANAGING DEEP AND EXTREMELY DEEP CARIES

5.1 Selective caries removal and indirect pulp treatment

Selective removal intentionally retains soft or firm carious dentine on the pulpal wall when further excavation would create a high risk of exposure; peripheral enamel and dentine are prepared to support an effective seal. The biological objective is to deprive the residual lesion of nutrients and avoid introducing bacteria into a pulp that may recover beneath the remaining dentine. Treatment is completed by the sealed restoration, not by the color or hardness of residual dentine at the pulpal wall.

For deep caries in a permanent tooth diagnosed with normal pulp or reversible pulpitis, the AAPD strongly recommends selective rather than nonselective removal because it produces fewer pulp exposures; it also favors a single-visit selective approach over stepwise excavation [2]. The American Dental Association guideline conditionally supports selective removal for advanced lesions in vital permanent teeth, although the certainty of evidence is very low [8]. Five-year randomized evidence supports the durability of restorations placed after selective excavation, and human histological observations show that pulpal healing can occur beneath sealed residual caries [9, 10].

Avoiding unnecessary exposure is particularly valuable in an immature permanent tooth, but incomplete peripheral caries removal that compromises bonding or sealing is not selective treatment. The operator should establish sound or firm peripheral margins compatible with the restorative system, avoid desiccation and thermal injury, and place a restoration that can be monitored and maintained. A planned return solely to excavate closer to the pulp is usually unnecessary and increases the risk of exposure.

The choice of liner beneath indirect treatment appears less important than the quality of the seal. The AAPD found no significant 24-month difference among glass ionomer, calcium hydroxide, and calcium silicate liners, based on moderate-certainty evidence [2]. A liner should therefore be used for a defined material or restorative indication, not to compensate for an unreliable restoration.

5.2 Direct pulp capping

Direct pulp capping places a pulp-contact material over a small exposure without removing additional pulpal tissue. In selected cariously exposed permanent teeth with normal pulp or reversible pulpitis, calcium silicate direct capping can achieve outcomes comparable to more invasive VPT. The 2025 AAPD guideline accepts either indirect treatment or calcium silicate

direct capping after exposure during management of deep caries and reports no clinically meaningful difference at 36 months in the evidence reviewed [2]. It also favors partial pulpotomy over direct capping when a calcium silicate cement is not used.

The indication is narrower than the phrase "vital pulp" might suggest. Direct capping is most defensible when the exposure is limited, the tooth has no features suggesting extensive inflammation, contamination can be controlled, the pulp appears viable, and hemostasis is prompt. Calcium hydroxide should not be regarded as equivalent to a modern hydraulic calcium silicate cement: the AAPD strongly recommends a calcium silicate material for carious direct pulp capping, and meta-analysis shows a greater decline in longer-term success with calcium hydroxide than with mineral trioxide aggregate (MTA) or related materials [2, 11].

After traumatic exposure, direct capping may be considered only for a very small, clean, recent wound with favorable associated injuries. Current AAPD guidance conditionally favors partial or full pulpotomy over direct capping for traumatic exposures, based on low-certainty and partly indirect evidence [2]. Tissue removal is especially valuable when the superficial pulp is contaminated or mechanically damaged.

5.3 Partial pulpotomy

Partial pulpotomy removes a shallow layer of inflamed coronal pulp, commonly about 1–3 mm or until healthy-appearing tissue capable of controlled hemostasis is reached. By preserving more coronal tissue than full pulpotomy, it aligns closely with the goal of apexogenesis. The depth should be guided by tissue findings rather than an arbitrary millimeter value, and the tissue should be removed with a sterile, sharp instrument under coolant to avoid further injury.

For carious exposure in teeth diagnosed with normal pulp or reversible pulpitis, partial pulpotomy with a conventional calcium silicate cement is a predictable option. The AAPD found similar 24-month success for partial and full pulpotomy in this population, although the comparison was supported by low-certainty evidence [2]. A focused systematic review identified only two direct comparative studies and found no significant difference, highlighting the limited evidence for choosing the ideal pulpotomy depth [12]. A broader review of randomized trials also reported high overall success for VPT with calcium silicate materials but rated the evidence as low certainty [13]. Small trials in immature teeth document continued root development after partial pulpotomy, although sample size and follow-up limit precision [14].

When the radicular pulp remains healthy, partial pulpotomy is the principal vital treatment for a complicated crown fracture in an immature permanent tooth. International trauma guidance prioritizes pulp preservation to allow continued root formation [5]. In a randomized trial of traumatized immature permanent teeth, both a calcium silicate material and calcium hydroxide supported favorable root maturation, with no statistically significant difference in survival or root-development measures [15]. This finding supports the feasibility of partial pulpotomy but does not supersede the broader evidence favoring hydraulic calcium silicate cement.

5.4 Full pulpotomy

Full pulpotomy removes the entire coronal pulp to the canal orifices while preserving radicular vitality. It is appropriate when superficial removal does not reveal a healthy wound surface with controllable bleeding, or when symptoms and caries depth suggest that inflammation extends more deeply through the coronal pulp. In selected permanent teeth with spontaneous, nocturnal, or lingering pain, the AAPD conditionally favors full over partial pulpotomy; the evidence is of low certainty and is not restricted to immature teeth [2].

Full pulpotomy is not a rescue procedure for every unsuccessful attempt at partial pulpotomy. The remaining radicular tissue must bleed consistently with vitality, become hemostatic after appropriate disinfection, and show no suppuration or necrosis. A randomized trial in mature symptomatic permanent teeth reported higher 12-month success for full than partial pulpotomy and proposed a short hemostasis threshold, but extrapolation to immature teeth requires caution [7]. A systematic review found favorable pooled outcomes for pulpotomy in cariously exposed permanent teeth; however, certainty remained very low to low, and outcomes were poorer in teeth classified as irreversibly inflamed [6].

Procedure selection should match the depth of inflammation rather than favor conservation for its own sake. Table 2 compares the tissue retained, clinical indications, operative endpoints, and principal limitations of each VPT approach; the preferred procedure is the least tissue removal required to reach a wound capable of healing.

Table 2. Comparison of vital pulp therapies for immature permanent teeth

Procedure	Tissue preserved	Best-supported clinical context	Operative endpoint	Main uncertainty or failure risk
Selective caries removal / indirect pulp treatment	The entire pulp; a dentine barrier remains.	Deep caries with normal pulp or reversible symptoms and no periapical disease	The pulp remains unexposed, peripheral margins are sealable, and a definitive restoration is placed.	Residual caries is acceptable only beneath an effective seal; radiographs may underestimate lesion depth.
Direct pulp capping	The entire pulp after exposure	A small, controlled carious exposure with a normal or reversible presentation; selected recent traumatic exposures	A viable surface, prompt hemostasis, and complete coverage with a calcium silicate cement	Superficial inflamed or contaminated tissue remains; the rationale is weaker after trauma or with more extensive symptoms.
Partial pulpotomy	Most coronal pulp and all radicular pulp	Carious exposure with a normal or reversible presentation; many traumatic exposures	Superficial inflamed pulp is removed, clean vital tissue is reached, and hemorrhage is controlled.	Removal may be too shallow when inflammation is extensive, and evidence often combines root-maturity groups.
Full pulpotomy	Radicular pulp	More extensive coronal inflammation; selected teeth with spontaneous, nocturnal, or lingering pain and normal periapical tissues	Vital radicular tissue remains at the canal orifices, hemostasis is achieved, and necrosis or suppuration is absent.	Evidence for symptomatic cases is mainly short term and includes mature teeth; uncontrolled hemorrhage precludes continued VPT.

6. TRAUMATIC PULP EXPOSURE

Trauma and caries create different biological conditions. Caries exposes the pulp through a biofilm-driven process that may evolve over months. A crown fracture is acute, but the wound may be contaminated by the environment, dentine fragments, or delayed presentation. Associated luxation can impair the neurovascular supply and complicate interpretation of sensibility responses. A recent injury therefore does not make every small exposure suitable for direct capping.

Examination should document the fracture, exposure size and duration, dentine contamination, stage of root development, periodontal injury, mobility, displacement, percussion response, soft-tissue injury, and radiographic evidence of root or alveolar fracture. In a vital immature tooth, the priority is to preserve enough pulp for continued root development. Partial pulpotomy is generally preferred because it removes contaminated and mechanically injured superficial tissue while retaining most of the coronal pulp and all radicular pulp [5]. Full pulpotomy remains appropriate when a shallow procedure does not produce an acceptable tissue endpoint.

Timing is relevant but should not become an absolute exclusion criterion. Favorable healing has been reported after delayed treatment when the remaining pulp is vital, and international trauma guidance does not define a universal interval beyond which partial pulpotomy must be abandoned. The decision should integrate contamination, associated injury, symptoms, radiographic findings, tissue appearance, and hemorrhage control. The crown fracture should be restored promptly, and fragment reattachment or a bonded restoration must provide a maintainable seal.

Follow-up after trauma is longer and more structured than routine restorative review because pulp necrosis, inflammatory root resorption, ankylosis, or disturbed root development may emerge later. A transiently negative sensibility response immediately after injury does not establish necrosis; serial clinical findings and radiographic development are more informative.

7. ASEPSIS, HEMOSTASIS, AND OPERATIVE TECHNIQUE

VPT is biologically conservative but technically exacting. Rubber dam isolation is the standard method for limiting salivary contamination and protecting the airway. The AAPD describes rubber dam as the gold standard, although direct comparative trials of isolation methods are lacking [2]. If adequate isolation, visibility, hemorrhage control, or immediate restoration cannot be achieved, referral or a planned alternative is safer than compromised VPT.

After access, caries should be removed with instruments and burs chosen to preserve sound tissue and minimize thermal injury. Magnification and illumination improve inspection and precision, although randomized evidence that magnification alone increases success is unavailable. A new or sterile high-speed diamond with copious coolant is commonly used for pulpotomy; the tissue should be cut cleanly rather than torn.

Sodium hypochlorite can provide both disinfection and hemorrhage control. For carious direct pulp capping, the AAPD strongly recommends sodium hypochlorite over saline on the basis of moderate-certainty but partly indirect evidence; for pulpotomy, the recommendation is conditional and the certainty very low [2]. Because studies use different concentrations, volumes, and application times, no single concentration can be declared universally superior. Controlled delivery, effective isolation, and attention to the open apex and surrounding tissues are essential. Manufacturer instructions, professional standards, local regulations, and operator training also remain relevant.

Once the wound is clean, a moistened sterile pellet may be applied with light pressure while hemostasis is assessed. Forceful compression can create an impression of control without resolving underlying inflammation. Repeated cutting and chemical application should not continue indefinitely. If hemorrhage cannot be controlled within a clinically reasonable period after sufficient tissue removal, or if suppuration or necrotic tissue is present, the diagnosis and treatment plan must be revised.

The exposed surface should be completely covered with the selected cement at the manufacturer recommended thickness, without voids or displacement. The restorative sequence must avoid washing out or contaminating the unset material. When material characteristics and clinical conditions permit, a definitive, well-sealed restoration at the same visit is preferred, although comparative evidence on restoration timing and type remains inconclusive [2, 16]. The coronal seal is an integral component of VPT rather than a separate procedural detail.

8. CALCIUM SILICATE BIOMATERIALS AND THE CORONAL SEAL

Hydraulic calcium silicate cements interact with tissue fluids, release calcium ions, create an alkaline environment during setting, and can support formation of a mineralized barrier. The material family includes MTA formulations, tricalcium silicate products such as Biodentine, and premixed bioceramic putties. Products differ in radiopacifier, setting behavior, handling, washout resistance, discoloration potential, and the quantity and

quality of clinical evidence. Membership in the same material class therefore does not establish clinical interchangeability.

A conventional non-resin hydraulic calcium silicate cement is the preferred pulp-contact material. The AAPD strongly recommends calcium silicate cement over calcium hydroxide for direct pulp capping of a carious exposure, based on moderate-certainty evidence, and also recommends it for pulpotomy [2]. A systematic review estimated that success after carious direct capping declined over time with all materials but remained consistently lower with calcium hydroxide than with MTA or Biodentine; the underlying studies were heterogeneous and generally of low quality [11]. Randomized evidence also documents favorable three-year VPT outcomes with both MTA and Biodentine, although inclusion of mature teeth limits direct inference for immature teeth [17].

Esthetic consequences are especially important in fractured anterior teeth. Bismuth oxide-containing MTA has been associated with substantial crown discoloration, and a trial in young permanent teeth reported more discoloration with ProRoot MTA than with Biodentine despite similar clinical success [18]. The AAPD strongly recommends a nonstaining calcium silicate material in esthetic areas, supported by high-certainty evidence for the discoloration outcome [2]. The term "nonstaining" remains product- and context-specific: blood contamination, restoration, dentine thickness, and formulation can affect color, and even lower-risk materials are not entirely immune to discoloration [19].

Light-cured resin-modified calcium silicate products offer convenient placement but contain resin components and do not share the same biological or clinical evidence as conventional hydraulic cements. A 2026 systematic review found broadly similar short-term clinical and radiographic outcomes in limited comparisons, followed by a later trend favoring non-resin materials and less consistent dentine-bridge formation with a resin-modified product [20]. These findings do not establish class equivalence. When sustained bioactivity and dentinogenesis are central objectives, a conventional hydraulic calcium silicate cement remains the more defensible choice.

Material selection must be integrated with handling and restoration. A cement supported by strong biological evidence may still fail if it is placed over infected tissue, mixed incorrectly, washed out, applied too thinly, or covered by a leaking restoration. The evidence reviewed above supports four practical distinctions:

Conventional hydraulic calcium silicate cements. This group has the strongest overall support for carious direct pulp capping and pulpotomy.

Bioactivity and sealing potential are advantages, but handling, setting, and discoloration remain formulation specific.

Handling or color optimised products and premixed putties. These materials can simplify delivery and may reduce discoloration risk. Long term comparative evidence in immature permanent teeth and product specific moisture and restorative requirements remain important limitations.

Calcium hydroxide. Historical success, familiarity, and low cost do not offset its poorer and declining performance after carious direct pulp capping, solubility, tunnel defects, and limited seal; it is not the preferred material.

Resin modified calcium silicate products. Light curing permits immediate handling, but the evidence is limited and mainly short term. Resin containing products should not be assumed biologically or clinically equivalent to conventional hydraulic cements.

9. EVIDENCE INFORMED CLINICAL DECISION PATHWAY

The preceding diagnostic and operative thresholds combine into a single clinical pathway that begins with evidence of restorability and preserved vitality. An unrestorable tooth, sinus tract, swelling, suppuration, necrotic tissue, or infection related periapical involvement leaves the vital pulp therapy pathway. Deep caries with a radiographic dentine barrier, together with a normal pulp or reversible pulpitis, supports selective removal and a sealed restoration. Where caries is very deep, or where deep caries is accompanied by spontaneous, nocturnal, or lingering pain in an otherwise suitable tooth, the pulp should be exposed deliberately and the tissue assessed directly. Carious exposure may be managed by calcium silicate direct pulp capping, partial pulpotomy, or full pulpotomy according to the clinical picture and the tissue findings; partial pulpotomy is usually preferred after trauma. If vital tissue is reached and hemostasis is achieved, a conventional hydraulic calcium silicate cement with a durable seal completes treatment. If this outcome is not achieved, a non-vital treatment plan or referral is indicated. This pathway does not replace intraoperative reassessment.

10. OUTCOMES AND LONG TERM FOLLOW-UP

Success is multidimensional. The tooth should remain functional and free of spontaneous pain, swelling, sinus tract, abnormal mobility, and persistent tenderness, while the restoration maintains an effective seal. Radiographs should demonstrate normal or improving periapical and furcal tissues

without pathologic resorption. In an immature tooth, serial images should also assess root length, dentinal wall thickness, and apical narrowing. Apical closure alone is an incomplete endpoint because closure may occur without the desired strengthening of thin root walls.

The European Society of Endodontology S3 guideline supports clinical review at approximately six months and a periapical radiograph at one year, with longer monitoring when healing is uncertain or root development remains incomplete [21]. Follow-up after trauma should also follow injury specific guidance [5]. A practical schedule comprises early review when symptoms or restoration integrity raise concern, clinical assessment at six months, standardized clinical and radiographic assessment at 12 months, and annual review until maturation is stable. Additional imaging should be justified by expected clinical benefit and radiation protection principles.

Sensibility responses may remain altered after pulpotomy or trauma, and a negative result does not independently establish failure. Trends in symptoms, examination findings, sensibility, and radiographs are more informative. Published success rates usually reflect the absence of symptoms and radiographic disease rather than direct evidence of histological health. The AAPD review reported approximate 24-month success of 91%–97% across indirect treatment, direct capping, partial pulpotomy, and full pulpotomy in selected permanent teeth with normal pulp or reversible pulpitis; however, comparisons were often indirect and follow-up beyond two to three years was limited [2, 22]. These estimates support offering VPT but do not justify promising a specific outcome.

11. FAILURE, RESCUE, AND THE BOUNDARY WITH REGENERATIVE ENDODONTICS

New spontaneous pain, swelling, a sinus tract, progressive radiolucency, pathologic resorption, arrested maturation, or loss of restorability requires prompt reassessment. Management should reflect current vitality, anatomy, stage of root development, restorability, and patient factors. If disease has progressed, a failed capping procedure should not be repeated automatically.

Regenerative endodontic procedures are intended for a necrotic immature permanent tooth and do not compete with first-line pulp preservation when biologically competent tissue remains. Once necrosis is established, regenerative treatment, apexification, conventional root canal treatment where feasible, or extraction may be considered according to anatomy, prognosis, expertise, and informed consent. Maintaining this distinction prevents the biological advantages of VPT from being used to undertreat necrosis.

12. EVIDENCE GAPS AND SHARED DECISION MAKING

Important research gaps include long-term trials restricted to immature teeth, standardized definitions of pulp status and success, validated inflammatory biomarkers, product-specific material comparisons, and patient centered outcomes. Guideline positions also differ. The 2021 American Association of Endodontists position emphasized complete caries removal before direct pulp assessment [23], whereas the newer AAPD guideline strongly favors selective removal when a dentine barrier remains and the pulp presents as normal or reversibly inflamed [2]. A 2026 European S3 guideline accepts VPT after exposure and selected pulpotomy for symptomatic teeth but does not resolve uncertainty about the most appropriate pulpotomy depth [24].

Informed consent should explain the rationale for preserving vitality, reasonable alternatives, diagnostic uncertainty, the need for a durable restoration and follow-up, possible discoloration, and the potential need for later endodontic treatment. Family preferences are important, but should be formed after discussing the consequences of both unnecessary pulp removal and delayed management of progressive disease.

13. CONCLUSION

VPT in an immature permanent tooth is a staged biological decision rather than a material driven procedure. Selective caries removal protects an unexposed pulp; direct pulp capping, partial pulpotomy, and full pulpotomy remove progressively more tissue to reach a wound capable of healing. A conventional hydraulic calcium silicate cement and a reliable coronal seal are central to treatment, but neither can compensate for infection beyond the indication. Long-term success requires continued root maturation, periapical health, restoration integrity, and structured follow-up.

REFERENCES

[1]. Tong, H. J., Seremidi, K., Stratigaki, E., Kloukos, D., Duggal, M., and Gizani, S. (2022). Deep dentine caries management of immature permanent posterior teeth with vital pulps: A systematic review and meta-analysis. Journal of Dentistry, 124, 104214. https://doi.org/10.1016/j.jdent.2022.104214

[2]. Coll, J. A., Dhar, V., Guelmann, M., Crystal, Y. O., Chen, C. Y., Marghalani, A. A., AlShamali, S., Xu, Z., Glickman, G., Ather, A., Sabeti, M., and Wedeward, R. (2025a). Guideline for use of vital pulp therapy in permanent teeth. Pediatric Dentistry, 47(5), 299–311. https://www.aapd.org/globalassets/media/policies_guidelines/g_vpt-permanentteeth.pdf

[3]. Dhar, V., El-Karim, I., Coll, J. A., Fouad, A. F., O'Connell, A. C., Asgary, S., Bjørndal, L., Cehreli, Z. C., Crystal, Y. O., Ekambaram, M., Hodgson, B. D., Innes, N. P., Rodrigues, J. A., Taha, N. A., Tewari, N., and Turk, T. (2026).

Consensus-based recommendations on pulp therapies in primary and permanent teeth: IAPD Porto Forum. International Journal of Paediatric Dentistry, 36(2), 371–386. https://doi.org/10.1111/ipd.70068

[4]. Donnermeyer, D., Dammaschke, T., Lipski, M., and Schäfer, E. (2023). Effectiveness of diagnosing pulpitis: A systematic review. International Endodontic Journal, 56(Suppl. 3), 296–325. https://doi.org/10.1111/iej.13762

[5]. Bourguignon, C., Cohenca, N., Lauridsen, E., Flores, M. T., O'Connell, A. C., Day, P. F., Tsilingaridis, G., Abbott, P. V., Fouad, A. F., Hicks, L., Andreasen, J. O., Cehreli, Z. C., Harlamb, S., Kahler, B., Oginni, A., Semper, M., and Levin, L. (2020). International Association of Dental Traumatology guidelines for the management of traumatic dental injuries: 1. Fractures and luxations. Dental Traumatology, 36(4), 314–330. https://doi.org/10.1111/edt.12578

[6]. Li, W., Yang, B., and Shi, J. (2024). Efficacy of pulpotomy for permanent teeth with carious pulp exposure: A systematic review and meta-analysis of randomized controlled trials. PLOS ONE, 19(7), e0305218. https://doi.org/10.1371/journal.pone.0305218

[7]. Taha, N. A., and Albakri, S. W. (2024). Partial versus full pulpotomy in mature permanent teeth with symptoms indicative of irreversible pulpitis: A randomized clinical trial. Journal of Endodontics, 50(7), 889–898. https://doi.org/10.1016/j.joen.2024.03.012

[8]. Dhar, V., Pilcher, L., Fontana, M., González-Cabezas, C., Keels, M. A., Mascarenhas, A. K., Nascimento, M. M., Platt, J. A., Sabino, G. J., Slayton, R., Tinanoff, N., Young, D. A., Zero, D. T., Pahlke, S., Urquhart, O., O'Brien, K. K., and Carrasco-Labra, A. (2023). Evidence-based clinical practice guideline on restorative treatments for caries lesions: A report from the American Dental Association. Journal of the American Dental Association, 154(7), 551–566.e51. https://doi.org/10.1016/j.adaj.2023.04.011

[9]. Jardim, J. J., Mestrinho, H. D., Koppe, B., de Paula, L. M., Alves, L. S., Yamaguti, P. M., Almeida, J. C. F., and Maltz, M. (2020). Restorations after selective caries removal: 5-year randomized trial. Journal of Dentistry, 99, 103416. https://doi.org/10.1016/j.jdent.2020.103416

[10]. Ricucci, D., Siqueira, J. F., Jr., Rôças, I. N., Lipski, M., Shiban, A., and Tay, F. R. (2020). Pulp and dentine responses to selective caries excavation: A histological and histobacteriological human study. Journal of Dentistry, 100, 103430. https://doi.org/10.1016/j.jdent.2020.103430

[11]. Cushley, S., Duncan, H. F., Lappin, M. J., Chua, P., Elamin, A. D., Clarke, M., and El-Karim, I. A. (2021). Efficacy of direct pulp capping for management of cariously exposed pulps in permanent teeth: A systematic review and meta-analysis. International Endodontic Journal, 54(4), 556–571. https://doi.org/10.1111/iej.13449

[12]. Louzada, L. M., Hildebrand, H., Neuhaus, K. W., and Duncan, H. F. (2025). The effectiveness of partial pulpotomy compared with full pulpotomy in managing deep caries in vital permanent teeth with a diagnosis of non-traumatic pulpitis. International Endodontic Journal, 58(1), 37–54. https://doi.org/10.1111/iej.14149

[13]. Sabeti, M., Huang, Y., Chung, Y. J., and Azarpazhooh, A. (2021). Prognosis of vital pulp therapy on permanent dentition: A systematic review and meta-analysis of randomized controlled trials. Journal of Endodontics, 47(11), 1683–1695. https://doi.org/10.1016/j.joen.2021.08.008

[14]. Uyar, D. S., and Alacam, A. (2021). Evaluation of partial pulpotomy treatment in cariously exposed immature permanent molars: Randomized controlled trial. Nigerian Journal of Clinical Practice, 24(10), 1511–1519. https://doi.org/10.4103/njcp.njcp_686_20

[15]. Yang, Y., Xia, B., Xu, Z., Dou, G., Lei, Y., and Yong, W. (2020). The effect of partial pulpotomy with iRoot BP Plus in traumatized immature permanent teeth: A randomized prospective controlled trial. Dental Traumatology, 36(5), 518–525. https://doi.org/10.1111/edt.12563

[16]. Duncan, H. F., Galler, K. M., Tomson, P. L., Simon, S., El-Karim, I., Kundzina, R., Krastl, G., Dammaschke, T., Fransson, H., Markvart, M., Zehnder, M., and Bjørndal, L. (2019). European Society of Endodontology position statement: Management of deep caries and the exposed pulp. International Endodontic Journal, 52(7), 923–934. https://doi.org/10.1111/iej.13080

[17]. Awawdeh, L., Al-Qudah, A., Hamouri, H., and Chakra, R. J. (2018). Outcomes of vital pulp therapy using mineral trioxide aggregate or Biodentine: A prospective randomized clinical trial. Journal of Endodontics, 44(11), 1603–1609. https://doi.org/10.1016/j.joen.2018.08.004

[18]. Uesrichai, N., Nirunsittirat, A., Chuveera, P., Srisuwan, T., Sastraruji, T., and Chompu-Inwai, P. (2019). Partial pulpotomy with two bioactive cements in permanent teeth of 6- to 18-year-old patients with signs and symptoms indicative of irreversible pulpitis: A noninferiority randomized controlled trial. International Endodontic Journal, 52(6), 749–759. https://doi.org/10.1111/iej.13071

[19]. Slaboseviciute, M., Vasiliauskaite, N., Drukteinis, S., Martens, L. C., Rajasekharan, S., and Vasiliauskiene, I. (2021). Discoloration potential of Biodentine: A systematic review. Materials, 14(22), 6861. https://doi.org/10.3390/ma14226861

[20]. Cabrera-Fernández, A., Dominguez-Dominguez, L., Pérez-Pérez, A., Santos, J. M. M., Díaz-Cuenca, A., Torres-Lagares, D., Sequeira, D. B., Segura-Egea, J. J., and Martín-González, J. (2026). Clinical and radiographic outcomes of vital pulp therapy using resin-modified versus conventional calcium silicate-based materials: A systematic review and meta-analysis. Journal of Functional Biomaterials, 17(1), 32. https://doi.org/10.3390/jfb17010032

[21]. Duncan, H. F., Kirkevang, L. L., Peters, O. A., El-Karim, I., Krastl, G., Del Fabbro, M., Chong, B. S., Galler, K. M., Segura-Egea, J. J., Kebschull, M., and European Society of Endodontology Workshop Participants and Methodological Consultant. (2023). Treatment of pulpal and apical disease: The European Society of Endodontology S3-level clinical practice guideline. International Endodontic Journal, 56(Suppl. 3), 238–295. https://doi.org/10.1111/iej.13974

[22]. Coll, J. A., Dhar, V., Guelmann, M., Crystal, Y. O., Chen, C. Y., Marghalani, A. A., AlShamali, S., Xu, Z., Glickman, G., Ather, A., Sabeti, M., and Wedeward, R. (2025b). Vital pulp therapy in permanent teeth: A systematic review and meta-analyses. Pediatric Dentistry, 47(3), 137–150, E3–E75. https://pubmed.ncbi.nlm.nih.gov/40533920/

[23]. American Association of Endodontists. (2021). AAE position statement on vital pulp therapy. Journal of Endodontics, 47(9), 1340–1344. https://doi.org/10.1016/j.joen.2021.07.015

[24]. Schwendicke, F., Kosan, E., Banerjee, A., Baysan, A., Bjørndal, L., Ceballos, L., Duncan, H. F., Herbst, S., Neuhaus, K. W., O'Connell, A. C., Paris, S., and Dujic, H. (2026). Deep caries management: EFCD-ESE-ORCA S3-level clinical practice guideline. Clinical Oral Investigations, 30(5), 186. https://doi.org/10.1007/s00784-025-06727-1

Personnel Hygiene and Food Safety Management in Mass Catering Systems: A Narrative Review

Yusuf ÖZTÜRK[1]

1- Öğr. Gör.; Ardahan Üniversitesi Teknik Bilimler Meslek Yüksekokulu Otel, Lokanta ve İkram Hizmetleri Bölümü, Ardahan, Türkiye. yusufozturk@ardahan.edu.tr ORCID No: 0000-0001-8303-3016

ABSTRACT

Food safety in mass catering systems is a critical public health and operational concern due to the large-scale production and distribution of meals. This narrative review evaluates the role of personnel hygiene and food safety management in institutional catering environments. Relevant literature published between 2000 and 2025 was identified through major databases using keywords related to food safety, hygiene, HACCP, and mass catering. The findings indicate that food safety is a socio-technical outcome shaped by the interaction of human behavior, processes, infrastructure, and management systems. Personnel hygiene is a central determinant, with common risk factors including inadequate hand hygiene, improper glove use, poor illness reporting, and insufficient sanitation during task transitions. Process-related vulnerabilities such as cross-contamination, temperature abuse, and unsafe thawing practices further increase risk. Although preventive systems like HACCP provide a structured framework, their effectiveness depends on proper implementation, staff compliance, and organizational support. The review also highlights that knowledge alone is insufficient to ensure safe practices. Sustainable improvements require continuous training, strong food safety culture, leadership involvement, and real-time monitoring. High staff turnover and operational pressures remain significant challenges. Emerging technologies such as digital monitoring and traceability systems offer potential benefits but require effective integration into daily operations. In conclusion, enhancing food safety in mass catering systems requires a shift from compliance-based approaches to adaptive, behavior-focused strategies that integrate personnel, management, and technology.

Keywords – food safety; personnel hygiene; mass catering systems; HACCP; food safety culture; KAP framework

1. INTRODUCTION

Food safety is both a public health imperative and an operational governance issue in institutional and commercial food service. In mass catering systems, where meals are produced in high volumes under strict time constraints, minor process deviations can rapidly scale into population-level risks. Urbanization, labor market shifts, and increased reliance on out-of-

home eating have amplified the role of centralized catering in hospitals, schools, military services, and industrial canteens (Fariba et al., 2018; Ghezzi and Ayoun, 2013). These settings are characterized by complex workflows, simultaneous preparation streams, frequent staff movement across stations, and significant heterogeneity in worker training profiles.

The burden of foodborne disease remains substantial globally. Reported annual estimates of hundreds of millions of illnesses and hundreds of thousands of deaths underscore that food safety failures continue despite modern regulatory frameworks and technological advances (Roy et al., 2024). Within mass catering operations, risk concentration is magnified because one contaminated batch may expose hundreds or thousands of consumers in a single service period. Consequently, operational resilience requires not only compliance with hygienic standards but also adaptive risk management that anticipates pressure points such as peak service windows, staff shortages, and cold-chain disruptions.

A central insight across both source manuscripts is that personnel hygiene is not a peripheral housekeeping issue but a core control domain that mediates biological, chemical, and physical hazards. Workers are the most frequent contact interface in food systems and thus function simultaneously as barriers and vectors of contamination (Ko, 2013; Putri and Susanna, 2021). Hand hygiene, protective clothing, illness reporting, glove discipline, and station transitions are therefore embedded determinants of outcome quality rather than isolated procedural checkpoints.

Despite the substantial body of literature on food safety in catering environments, evidence remains fragmented across disciplines and settings, with limited integration between behavioral, organizational, and technical perspectives. Many existing reviews address either hygiene practices or management systems in isolation, without sufficiently examining the interplay between personnel factors and structural controls. Furthermore, there is a relative scarcity of synthesized evidence specifically focused on mass catering systems as distinct from retail or domestic food contexts. This review aims to address these gaps by providing a comprehensive, integrated analysis of personnel hygiene and food safety management in mass catering, with the objective of informing both operational practice and future research priorities.

2. METHOD

This study was conducted as a narrative review. Relevant literature was identified through systematic searches of the PubMed, Web of Science, Scopus, and Google Scholar databases, covering publications from 2000 to 2025, with priority given to studies published after 2015. Search terms included combinations of the following: "food safety," "mass catering," "food service," "personnel hygiene," "food handler," "HACCP," "cross-contamination," "temperature control," "foodborne illness," "food safety culture," and "training." Articles not available in English or Turkish, conference abstracts without full-text access, and publications with no clear methodological description were excluded.

3. CONCEPT OF FOOD SAFETY AND FUNDAMENTAL PRINCIPLES

3.1. Definition and Scope of Food Safety

Food safety can be defined as the assurance that food will not cause harm to consumers when prepared, handled, and consumed as intended. In mass catering, this assurance is operationalized through layered controls extending from procurement to final service. Both manuscripts converge on a systems perspective in which food safety is not reducible to end-product testing; instead, it depends on preventive design, process discipline, and continuous verification across the production chain (Başaran, 2015; Özkan, 2021).

This systems perspective aligns with modern management standards in which prerequisite programs (PRPs) provide infrastructural and procedural foundations—including water quality, cleaning regimes, pest control, staff hygiene, and supplier controls—while hazard-specific controls are established through risk analysis methodologies (Başaran, 2015; Kamboj et al., 2020). In practice, deficient PRPs undermine advanced control plans by introducing recurrent instability at the operational base.

3.2. Global Burden of Foodborne Diseases

Foodborne diseases represent a persistent transnational burden with significant clinical, economic, and social consequences. The integrated manuscripts note that morbidity estimates remain high across income settings and that institutional meal systems can amplify outbreak scale when

contamination occurs upstream or within preparation lines (Ali et al., 2026; Roy et al., 2024). Beyond acute gastroenteritis, severe outcomes include hospitalization, long-term sequelae, and mortality among vulnerable groups.

In mass catering operations, consequence severity is shaped by consumer profile and exposure density. Hospital and elder-care settings include immunologically susceptible populations for whom low-dose contamination may produce disproportionate harm. Simultaneously, outbreak externalities extend to organizational continuity, legal liability, reputational damage, and emergency response costs. Therefore, food safety governance in catering systems should be conceptualized as an integrated public health and enterprise risk function rather than an isolated quality assurance task.

3.3. Biological, Chemical, and Physical Hazards

Biological hazards remain dominant in operational incident analyses. Pathogens commonly implicated in food service contamination include Salmonella, Campylobacter, Escherichia coli, Staphylococcus aureus, and Listeria monocytogenes, with transmission facilitated by inadequate temperature control, poor hand hygiene, and cross-contact between raw and ready to eat products (Hartantyo et al., 2023; Putri and Susanna, 2021; Teym et al., 2025). Viral agents and low infectious-dose organisms intensify risks associated with symptomatic employees and contaminated high-touch surfaces.

Chemical hazards include sanitizer residues, pesticide contaminants, packaging migrants, and allergen mismanagement. In high volume kitchens, chemical risk can increase through incorrect dilution procedures, unlabeled transfer containers, and weak segregation between food and cleaning chemical storage zones (Başaran, 2015; Şahin and Çetin, 2017). Allergen control is particularly sensitive because production line sharing and menu diversification increase opportunities for unintended exposure.

Physical hazards involve extraneous materials such as glass fragments, metal particles, plastic shards, hair, jewelry components, and packaging debris. Although often visible, physical hazards may persist in environments with inadequate equipment maintenance, damaged utensils, or absent foreign body prevention checks. Thus, hazard control requires coordinated interventions spanning procurement, maintenance, staff behavior, and verification.

3.4. HACCP and Preventive Approaches

HACCP remains the dominant preventive architecture for managing critical hazards in catering systems. Its value lies in structured hazard identification, critical control point designation, limit setting, monitoring, corrective action, verification, and recordkeeping (Wallace and Mortimore, 2016). However, the synthesized manuscripts emphasize that documentation quality alone is insufficient. Effective HACCP depends on behavioral compliance, realistic workflow integration, and actionable corrective pathways.

Evidence from training and implementation studies indicates that food handlers may report high procedural awareness while failing to apply controls under operational pressure, especially during peak service periods (Ko, 2013; Nik Husain et al., 2016). This disconnect highlights the need to embed preventive controls into task design for example, simplified measurement protocols, station-level prompts, and supervisor feedback loops rather than relying on periodic classroom instruction alone.

3.5. Farm-to-Fork Continuity

The farm-to-fork concept reframes food safety as a chain wide responsibility with interdependent control points. Mass catering facilities operate at the terminal and highly visible stage of this chain, but risk origin may lie in supplier nonconformities, transport failures, or storage deviations before ingredients reach the kitchen (Ellahi et al., 2023; Özkan, 2021). Consequently, robust receiving checks, supplier qualification criteria, lot traceability, and cold-chain integrity verification are essential components of catering risk governance.

This chain oriented logic also supports rapid incident response. When traceability systems are complete and accessible, implicated batches can be isolated faster, exposure can be limited, and root cause analysis becomes more reliable. Conversely, fragmented records prolong uncertainty and magnify public health impact.

4. ROLE OF KITCHEN STAFF IN FOOD SAFETY

4.1. Human Factor in Contamination Dynamics

Across both manuscripts, the human factor emerges as the most consequential and most variable element in food safety performance. Workers interact continuously with raw materials, tools, surfaces, storage units, and

final products; therefore, behavioral consistency directly influences contamination probability. Observational evidence from bulk preparation contexts demonstrates that procedural gaps frequently occur at transition points station changes, mid-task interruptions, and concurrent handling of dissimilar food categories (Hartantyo et al., 2023).

Human factor vulnerability is not merely an issue of individual negligence. It reflects the interaction of staffing intensity, schedule compression, fatigue, communication quality, equipment accessibility, and supervisory practices. In this sense, personnel-related failures should be interpreted as system signals indicating mismatches between expected protocols and real operational conditions.

4.2. Risk Behaviors among Food Handlers

The manuscripts repeatedly identify a set of high-frequency risk behaviors: incomplete handwashing, glove misuse, contact transfer between raw and cooked foods, inadequate disinfection between tasks, and continuation of work despite gastrointestinal symptoms (Fariba et al., 2018; Teym et al., 2025; Tekinçay and Ademoğlu, 2025). Additional nonconformities include insufficient temperature verification, ad hoc thawing practices, and poor compliance with protective clothing standards.

These behaviors are often clustered rather than isolated. For example, understaffed shifts may simultaneously exhibit reduced hand hygiene opportunities, delayed surface sanitization, and shortcuts in temperature logging. Such clustering suggests that risk management should move beyond single-behavior correction toward integrated operational redesign.

4.3. Knowledge Attitude Practice (KAP) Framework

The KAP framework provides a useful but incomplete model for interpreting personnel performance. Empirical findings across restaurant and catering settings show that knowledge can predict safer intentions, yet behavior outcomes are mediated by attitudes, perceived norms, and situational constraints (Ko, 2013; Putri and Susanna, 2021). In other words, cognitive competency is necessary but not sufficient.

Training interventions improve knowledge and selected practices, particularly when structured, repeated, and reinforced through workplace supervision (Alkhaldi et al., 2025; Nik Husain et al., 2016). Nevertheless, the sustainability of improvements depends on organizational reinforcement

mechanisms. Where management tolerates shortcuts or prioritizes service speed over protocol integrity, KAP gains erode quickly.

5. PERSONNEL HYGIENE IN MASS CATERING SYSTEMS

5.1. Hand Hygiene and Correct Technique

Hand hygiene is the most fundamental contamination barrier in food handling environments. Yet observational and microbiological evidence indicates substantial compliance variability across shifts and institutions (Dinçoğlu et al., 2025; Hartantyo et al., 2023). Effective practice requires appropriate timing, duration, and technique, including friction across palms, interdigital spaces, thumbs, fingernails, and wrists, followed by hygienic drying.

Critical handwashing moments include pre-shift start, post-restroom use, post-raw food contact, post-waste handling, post-face/hair contact, post-cleaning chemical handling, and before contact with ready to eat products. Compliance declines under service intensity unless supported by conveniently located sinks, uninterrupted soap and towel supply, behavioral prompts, and supervisory reinforcement. Thus, hand hygiene performance is a joint function of worker discipline and infrastructure design.

5.2. Personal Protective Equipment (PPE)

Protective clothing, hair restraints, gloves, and dedicated footwear reduce contamination transfer when used correctly. However, glove use is frequently misinterpreted as a substitute for hand hygiene, producing risk compensation behaviors that increase cross-contamination (Fariba et al., 2018; Tekinçay and Ademoğlu, 2025). Effective PPE governance requires task-specific change rules, contamination-awareness training, and direct observation during high-traffic preparation periods.

Uniform management is similarly important. Clean, kitchen dedicated garments, proper storage, and exclusion of external wear from production zones reduce environmental carry in contamination. Jewelry restrictions and nail hygiene policies remain essential for controlling both biological and physical hazards.

5.3. Staff Health Monitoring

Symptomatic workers, particularly those with vomiting, diarrhea, fever, jaundice, or infected skin lesions, present high transmission risk in food

environments. A robust health monitoring framework includes symptom self-reporting protocols, confidential supervisor notification channels, exclusion criteria, return to work rules, and nonpunitive sick leave culture (Ali et al., 2026; Teym et al., 2025).

Where economic or managerial pressures incentivize presenteeism, outbreak risk increases. Therefore, health surveillance is both a medical and policy issue. Institutions should align staffing plans and labor policies with preventive objectives so that workers are not forced to choose between income continuity and safe practice.

5.4. Food Safety Training for Staff

Training remains indispensable but must be behavior centered. Evidence supports blended approaches combining initial onboarding, short recurrent modules, demonstration-based instruction, peer mentoring, and immediate corrective feedback (Alkhaldi et al., 2025; Nik Husain et al., 2016). Programs focused solely on regulations without contextual application show weaker transfer to practice.

High turnover sectors require modular curricula aligned to role complexity (preparation, cooking, service, dishwashing, receiving). Microlearning formats, multilingual content, and visual process cues can improve retention in diverse labor contexts. Importantly, training outcomes should be evaluated through direct observations and process indicators rather than knowledge tests alone.

6. FOOD HYGIENE PRACTICES IN MASS CATERING

6.1. Cross-Contamination Prevention

Cross-contamination control is a cornerstone of safe mass food production. Effective strategies include spatial separation of raw and cooked workflows, color coded equipment, dedicated utensils, unidirectional process flow, and strict transition sanitation between tasks (Hartantyo et al., 2023; Özkan, 2021). Where complete spatial separation is infeasible, temporal separation and validated sanitation intervals become critical.

Storage hierarchy also matters: ready to eat foods should be protected on upper shelves, while raw meats are contained on lower levels to prevent drip transfer. Frequent touch points handles, faucet controls, trolley grips, and

cutting board edges require enhanced cleaning frequency because they function as contamination bridges across process zones.

6.2. Temperature Control and the Danger Zone

Temperature abuse remains one of the most persistent contributors to microbial growth. The danger zone concept (approximately 5 °C to 60 °C) remains operationally useful for staff training and monitoring design. In mass catering, risk concentrates during cooling, reheating, and prolonged hot holding. Control reliability depends on calibrated probes, standardized measurement points, real time recording, and escalation rules for deviations (Hartantyo et al., 2023; Wallace and Mortimore, 2016). Large volume food containers pose additional cooling risks due to thermal inertia; practical mitigation includes shallow pan distribution, blast chilling where available, and controlled airflow.

6.3. Safe Cooking and Storage

Safe cooking requires achievement of validated core temperatures appropriate to product class and hazard profile. Reheating should be rapid and verified, avoiding extended residence in moderate temperature bands. Storage systems should apply first-expiry/first-out principles, unambiguous labeling, and lot level traceability to support both quality control and incident investigation (Baş et al., 2006; Uçar et al., 2021).

The synthesis indicates that failures often occur at routine interfaces: unlabeled leftovers, ambiguous production time stamps, and undocumented reheating cycles. These are low visibility errors with high downstream consequence potential in mass service contexts.

6.4. Freezing and Thawing Practices

Frozen storage generally slows microbial activity, but thawing is a high risk transition stage. Safe thawing methods include refrigerated thawing, controlled running water thawing under defined limits, or microwave thawing when immediate cooking follows. Countertop thawing is consistently associated with elevated risk because product surfaces can enter growth permissive temperatures while cores remain frozen. Refreezing thawed products without validated control introduces quality degradation and possible safety compromise. Thus, thawing decisions should be integrated into production planning to minimize emergency shortcuts during service periods (Çalışkan Koç et al., 2025; Food Safety and Inspection Service, n.d.).

6.5. Witness Sample (Retention) Practices

Witness sample retention provides forensic support during suspected foodborne events. Standardized protocols should define sample mass, aseptic collection method, labeling elements (menu item, date/time, lot, collector), storage conditions, retention duration, and chain-of-custody (Council to Improve Foodborne Outbreak Response (CIFOR), 2020).

Retention systems are most useful when linked to wider verification architecture. If retained samples, environmental swabs, and process logs are integrated, facilities can perform faster root-cause triangulation and implement targeted corrective and preventive actions (Wellman et al., 2023).

7. KITCHEN AND EQUIPMENT HYGIENE

7.1. Cleaning versus Sanitization

Cleaning removes visible soils; sanitization reduces microbial loads to acceptable levels. Confusion between these functions remains common in practice and can lead to false confidence after visually satisfactory but microbiologically inadequate cleaning. Effective sanitation requires correct chemical selection, concentration control, contact time, rinsing logic, and compatibility with surface materials (Özkan, 2021; ŞAHİN and ÇETİN, 2017).

Standard operating procedures should distinguish routine, between batch, and deep clean cycles with explicit responsibilities and verification points. Inconsistent execution, particularly during shift transitions, is a recurrent failure mode in high throughput kitchens.

7.2. Surface and Equipment Sanitation

Food contact surfaces, small tools, and processing equipment can harbor residual biofilms if design or maintenance is inadequate. Damaged cutting boards, scratched plastics, inaccessible joints, and worn seals increase microbial persistence and reduce sanitization efficacy. Observational evidence from catering environments identifies preparation boards and high-use contact surfaces as recurrent hotspots (Dinçoğlu et al., 2025; Hartantyo et al., 2023).

Sanitation verification should combine visual checks with periodic microbiological or rapid hygiene tests, allowing trend analysis over time. Corrective action thresholds and retest policies should be predefined to prevent normalization of recurring nonconformities.

8. CURRENT CHALLENGES AND FUTURE PERSPECTIVES

High staff turnover remains a major barrier to consistent food safety performance, as it disrupts standardization, weakens tacit knowledge transfer, and increases early stage nonconformity risks among newly recruited personnel (Ali et al., 2026; Teym et al., 2025). Structured onboarding, competency based monitoring, and retention focused policies are therefore critical to maintaining process stability.

In addition, a persistent gap exists between training completion and behavioral change. Programs focused mainly on certification may improve knowledge without ensuring practical application in daily operations (Alkhaldi et al., 2025; Nik Husain et al., 2016). To address this, training should shift toward scenario-based learning, continuous reinforcement, and performance feedback linked to real operational indicators.

At the system level, the expansion of ready-to-eat foods and increasing supply chain complexity redistribute food safety risks. Greater reliance on suppliers and logistics systems elevates the importance of traceability, while ingredient diversity intensifies allergen management challenges (Ellahi et al., 2023). Accordingly, procurement and receiving processes must be treated as active control points.

Digitalization further transforms monitoring practices through real time data collection and automated alerts, improving detection capacity. However, its effectiveness depends on proper system design, user adoption, and clear response protocols; otherwise, it may lead to data overload without meaningful risk reduction (Verdouw et al., 2016).

Finally, emerging technologies such as IoT, AI, and blockchain offer enhanced monitoring, predictive analytics, and traceability capabilities (Ellahi et al., 2023). Despite their potential, implementation barriers particularly in small and medium enterprises necessitate supportive policies, workforce training, and evidence-based evaluation of their real-world impact on food safety outcomes

9. CONCLUSION AND RECOMMENDATIONS

This review shows that food safety in mass catering is a socio-technical outcome shaped by the interaction of people, processes, infrastructure, and governance. Key vulnerabilities persist in personnel hygiene (hand hygiene,

glove use, illness exclusion, task-transition sanitation) and process hygiene (cross-contamination control, temperature management, thawing, and sanitation verification). A central finding is that formal systems such as HACCP and SOPs are only effective when supported by a strong food safety culture and realistic implementation. Knowledge alone does not ensure behavior change; sustained performance depends on leadership, adequate facilities, continuous monitoring, and timely corrective actions.

From a management perspective, priorities should include behavior-based metrics, risk-based staffing, structured onboarding, and integrated verification systems. Policies should promote digital traceability and continuous monitoring. Research gaps remain in long-term effects of digital monitoring, cost-effectiveness of technologies, and culturally adapted interventions. Future studies should adopt mixed-method and implementation-focused designs.

Overall, improving food safety requires a shift from compliance-based approaches to adaptive prevention systems integrating training, culture, technology, and leadership.

REFERENCES

Ali, A., Tahir, A., Ahmed, N., Trafialek, J., Alohali, B. M., Mohamed Ahmed, I. A., Manzoor, M. F., and Madilo, F. K. (2026). Food Safety in the Catering Sector: Nonconformities, Challenges, and Strategic Interventions With Insights From South Asia and Africa. *Food Science & Nutrition*, *14*(1). https://doi.org/10.1002/fsn3.71400

Alkhaldi, S., Hod, R., Md Isa, Z., Idris, I. B., and Karim, N. (2025). The Impact of Food Safety Training Programs on Knowledge, Attitude, and Practice on Food Safety Among Migrant Workers – A Review. *Current Research in Nutrition and Food Science Journal*, *13*(2), 541–558. https://doi.org/10.12944/CRNFSJ.13.2.1

Baş, M., Ersun, A. Ş., and Kıvanç, G. (2006). Implementation of HACCP and prerequisite programs in food businesses in Turkey. *Food Control*, *17*(2), 118–126. https://doi.org/10.1016/j.foodcont.2004.09.010

Başaran, B. (2015). ISO 22000 Food Safety Management System. *Journal of Food and Health Science*, 9–26. https://doi.org/10.3153/JFHS16002

Çalışkan Koç, G., Özkan Karabacak, A., Süfer, Ö., Adal, S., Çelebi, Y., Delikanlı Kıyak, B., and Öztekin, S. (2025). Thawing frozen foods: A comparative review of traditional and innovative methods. *Comprehensive Reviews in Food Science and Food Safety*, *24*(2). https://doi.org/10.1111/1541-4337.70136

Council to Improve Foodborne Outbreak Response (CIFOR). (2020). GUIDELINES FOR FOODBORNE DISEASE OUTBREAK RESPONSE. In *GUIDELINES*

FOR FOODBORNE DISEASE OUTBREAK RESPONSE (3rd ed.). https://cifor.us/

Dinçoğlu, A. H., Çalışkan, Z., and Gürcan, G. (2025). Determination of the Hygiene Quality of Food Workers, Equipment, Surface Areas, and Ambient Air in Mass Catering Places. *Current Perspectives on Health Sciences*, *6*(1), 31–38. https://doi.org/10.58208/cphs.1576329

Ellahi, R. M., Wood, L. C., and Bekhit, A. E.-D. A. (2023). Blockchain-Based Frameworks for Food Traceability: A Systematic Review. *Foods*, *12*(16), 3026. https://doi.org/10.3390/foods12163026

Fariba, R., Gholamreza, J. K., Saharnaz, N., Ehsan, H., and Masoud, Y. (2018). Knowledge, attitude, and practice among food handlers of semi-industrial catering: a cross sectional study at one of the governmental organization in Tehran. *Journal of Environmental Health Science and Engineering*, *16*(2), 249–256. https://doi.org/10.1007/s40201-018-0312-8

Food Safety and Inspection Service. (n.d.). *Leftovers and Food Safety* . Retrieved 30 March 2026, from https://www.fsis.usda.gov/food-safety/safe-food-handling-and-preparation/food-safety-basics/leftovers-and-food-safety#_Thaw

Ghezzi, S., and Ayoun, B. (2013). Food safety in the US catering industry: empirical findings. *International Journal of Contemporary Hospitality Management*, *25*(3), 365–382. https://doi.org/10.1108/09596111311311026

Hartantyo, S. H. P., Selvaraj, R., Ho, J., Oh, J. Q., Er, J. C., Li, A., and Aung, K. T. (2023). Food Safety Controls during Bulk Food Preparation—An Observational Analysis. *Foods*, *12*(12), 2376. https://doi.org/10.3390/foods12122376

Kamboj, S., Gupta, N., Bandral, J. D., Gandotra, G., and Anjum, N. (2020). Food safety and hygiene: A review. *International Journal of Chemical Studies*, *8*(2), 358–368. https://doi.org/10.22271/chemi.2020.v8.i2f.8794

Ko, W.-H. (2013). The relationship among food safety knowledge, attitudes and self-reported HACCP practices in restaurant employees. *Food Control*, *29*(1), 192–197. https://doi.org/10.1016/j.foodcont.2012.05.076

Nik Husain, N. R., Wan Muda, W. M., Noor Jamil, N. I., Nik Hanafi, N. N., and Abdul Rahman, R. (2016). Effect of food safety training on food handlers' knowledge and practices. *British Food Journal*, *118*(4), 795–808. https://doi.org/10.1108/BFJ-08-2015-0294

Özkan, R. (2021). TOPLU BESLENME SİSTEMLERİNDE KULLANILAN GIDA KALİTE GÜVENCE SİSTEMLERİ. *Turkish Journal of Health Research*, *2*(3), 45–56.

Putri, M. S., and Susanna, D. (2021). Food safety knowledge, attitudes, and practices of food handlers at kitchen premises in the Port 'X' area, North Jakarta, Indonesia 2018. *Italian Journal of Food Safety*, *10*(4). https://doi.org/10.4081/ijfs.2021.9215

Roy, N., Aktarujjaman, M., Siddiky, A., Mollick, K., Rosby, R., Rogers, B. T., Hossain, E., Al Mamun, F., and Mamun, M. A. (2024). Food Safety Knowledge, Attitudes, Practices, and Associated Factors Among Adult Consumers in Bangladesh: Findings From a Nationwide Cross-Sectional Survey. *Health Science Reports*, *7*(12). https://doi.org/10.1002/hsr2.70307

ŞAHİN, B., and ÇETİN, S. A. (2017). Gıda Güvenliğinde Risk Faktörleri ve Hijyenin Önemi (Risk Factors and Hygiene Importance in Food Safety). *Journal of Tourism and Gastronomy Studies*, *5*(2), 310–321. https://doi.org/10.21325/jotags.2017.133

Teym, A., Adane, B., Assaye, B. T., and Zeleke, T. K. (2025). Food safety practice and associated factors among food handlers working in food and drinking establishments in Finote Selam, in resource-limited settings. *Frontiers in Sustainable Food Systems*, *9*. https://doi.org/10.3389/fsufs.2025.1536240

Uçar, A., Ülger, T. G., and Çakıroğlu, F. P. (2021). What changed between 2008–2020 about Employees' perception of hygiene in the catering industry in Ankara (Turkey)? *AIMS Public Health*, *8*(2), 275–284. https://doi.org/10.3934/publichealth.2021021

Verdouw, C. N., Wolfert, J., Beulens, A. J. M., and Rialland, A. (2016). Virtualization of food supply chains with the internet of things. *Journal of Food Engineering*, *176*, 128–136. https://doi.org/10.1016/j.jfoodeng.2015.11.009

Wallace, C. A., and Mortimore, S. E. (2016). HACCP. In *Handbook of Hygiene Control in the Food Industry* (pp. 25–42). Elsevier. https://doi.org/10.1016/B978-0-08-100155-4.00003-0

Wellman, A., Bazaco, M. C., Blessington, T., Pightling, A., Dwarka, A., Hintz, L., Wise, M. E., Gieraltowski, L., Conrad, A., Nguyen, T.-A., Hise, K., Viazis, S., and Beal, J. (2023). An Overview of Foodborne Sample-Initiated Retrospective Outbreak Investigations and Interagency Collaboration in the United States. *Journal of Food Protection*, *86*(6), 100089. https://doi.org/10.1016/j.jfp.2023.100089

YİĞİT TEKİNÇAY, M. S., and ADEMOĞLU, A. (2025). GIDA GÜVENLİĞİ VE PERSONEL HİJYENİNE YÖNELİK TUTUMUN DAVRANIŞ ÜZERİNDEKİ ETKİSİ: YEREL RESTORAN ÇALIŞANLARI ÜZERİNE BİR İNCELEME (THE IMPACT OF ATTITUDES TOWARD FOOD SAFETY AND PERSONNEL HYGIENE ON BEHAVIOR: A STUDY ON LOCAL RESTAURANT EMPLOYEES. *Journal of Gastronomy Hospitality and Travel (JOGHAT)*, *8*(4), 1749–1765. https://doi.org/10.33083/joghat.2025.591