ISSN: 2181-3337 | SCIENTISTS.UZ VOLUME 5 ISSUE 10 | OCTOBER 2026

FROM VISIBILITY TO CLINICAL WARRANT: RETHINKING PROGRESS IN MEDICAL PHYSICS

Dustnazar Shermatov, Satorov Dalerjon
State Educational Institution "Avicenna Tajik State Medical University"

DOI: 10.67620/sai.510d0007

ABSTRACT
Historical accounts of medical physics often organize change around discoveries and increasingly sophisticated instruments. This article develops a different interpretation: the discipline advances when new forms of visibility acquire defensible warrants for clinical action. A bounded critical documentary comparison examines computed tomography, professional qualification, and external validation of imaging artificial intelligence using published technical accounts, international professional guidance, and a cross-institutional empirical study. Hounsfield's descriptions demonstrate that reconstructed images depend on measurement design and computational inference; training recommendations locate responsibility within qualified clinical practice; and pneumonia-screening research shows why internal predictive performance cannot establish transportability. These sources support a distinction among physical validity, inferential adequacy, and clinical warrant. An original evidence–warrant atlas makes that distinction explicit without assigning numerical scores to heterogeneous evidence. The contribution is an analytical framework, not a comprehensive empirical history or a validated assessment instrument. It redirects evaluation from technological capability alone toward the conditions under which a particular output justifies a particular decision. Limitations include selective source coverage, reliance on one artificial-intelligence case, and the absence of patient-level reanalysis.

ACCEPTED

07.11.26

Keywords: *medical physics, clinical warrant, computed tomography, external validation, professional qualification.*

CC BY

Introduction

An image can become clearer without becoming sufficient grounds for a clinical decision. Computed tomography makes this distinction concrete: reconstruction creates cross-sectional representations from transmission measurements, while the interpretation of those representations still depends on the conditions under which evidence is produced (Hounsfield, 1973, 1979). Contemporary imaging artificial intelligence introduces a related problem when successful internal discrimination does not persist in an external population (Zech et al., 2018). These cases motivate a conceptual question about progress: what must be justified between generating an output and authorizing its clinical use?

This article argues that medical physics develops through the construction of clinical warrants: explicit reasons for treating a physically generated or computationally derived output as appropriate evidence for a specified clinical purpose. A warrant is not an additional property of the image. It is a justified relationship among an output, its production conditions, the population being examined, and the decision contemplated.

The research question is therefore not which invention followed which discovery, but how the evidentiary obligations attached to medical images change when measurement becomes reconstruction and reconstruction becomes prediction. Three documentary anchors permit a focused answer: Hounsfield's technical accounts of CT, international recommendations on medical physicists' qualifications, and Zech et al.'s (2018) investigation of cross-hospital pneumonia prediction. The resulting argument concerns continuity in responsibility, not equivalence among technologies.

Literature Review: From Discovery to Evidentiary Responsibility

Medical imaging joins physical measurement, computational processing, and clinical interpretation. Hounsfield (1973) describes the instrumental and computational components of CT; international guidance identifies the educational and clinical preparation required of medical physicists (International Atomic Energy Agency [IAEA], 2013). Reading these sources together suggests that technical capability and professional responsibility should be examined within the same evaluative account. Organizing analysis principally around instruments can obscure the distinct evidence needed to support interpretation and implementation.

Hounsfield's (1973) system description supports a less direct relationship between physical signal and medical knowledge. The instrument acquires transmission measurements from multiple angles and calculates absorption values for cross-sectional display. Reconstruction therefore changes what can be represented, but also introduces a computational operation between observation and representation. Hounsfield's (1979) Nobel lecture makes the experimental contingencies unusually visible: a preserved specimen initially yielded exaggerated tissue differences, prompting checks with fresh specimens. The image's apparent anatomical persuasiveness did not settle whether the observed contrast represented the intended biological condition.

The IAEA (2013) approaches reliability from another direction. Its recommendations distinguish academic education, supervised clinical preparation, and professional recognition. This is not historical evidence that every institution has implemented such requirements. It is normative evidence that sophisticated equipment does not make expertise self-executing.

Zech et al. (2018) extend the issue into predictive imaging. Their cross-sectional study examined whether pneumonia-screening models generalized across three hospital systems. Models could exploit information identifying acquisition sites, while differences in disease prevalence made site identity predictive. The conceptual challenge is not simply that an algorithm may be inaccurate. An apparently successful internal evaluation may reward a relationship that does not survive relocation. Together, these sources suggest that the relevant unit of assessment is the situated imaging practice, not the isolated machine or model.

Methodology

This study uses a critical documentary comparison rather than a systematic review, meta-analysis, or original clinical experiment. Four published sources form a purposive corpus: Hounsfield's (1973) technical system description, his Nobel lecture (Hounsfield, 1979), IAEA (2013) qualification recommendations, and Zech et al.'s (2018) cross-institutional pneumonia study. They were selected to examine reconstruction, professional responsibility, and predictive transportability, respectively. The comparison develops a conceptual argument across contrasting evidence types rather than estimating their historical prevalence or their combined clinical effects.

The comparison asks four questions of each source: What output is produced or governed? Which operation makes it evidentially meaningful? What limitation prevents an unrestricted inference? What responsibility follows for clinical implementation? Statements are separated into documented observations, source recommendations, and the present authorial interpretation. Table 1 records these distinctions; Figure 2 translates them into an evidence–warrant atlas. Neither display constitutes a numerical coding scheme.

ISSN: 2181-3337 | SCIENTISTS.UZ

The quantitative component is deliberately limited. Figure 1 redraws two published AUC estimates and their reported 95% confidence intervals for the same jointly trained model evaluated internally and externally (Zech et al., 2018). No patient records were accessed, models retrained, or statistical tests recomputed. The difference between point estimates is descriptive, and the source's significance assessment is not represented as a new finding.

Source selection was governed by relevance to a specific evidentiary transition: from measurement to representation, from technical capability to accountable service, or from internal prediction to external use. Statements about technical operations, professional recommendations, and empirical performance were kept distinct. Broad priority claims, universal modality rankings, and forecasts of clinical benefit were excluded because the selected corpus does not establish them. This strategy supports focused interpretation at the expense of historical completeness.

Table 1

Evidence Types, Interpretive Limits, and Responsibilities

Source / evidence type	What it supports	What it does not establish
Hounsfield (1973, 1979) / technical and retrospective accounts	CT reconstructs cross-sections; specimen conditions can alter apparent contrast.	Universal diagnostic superiority or improved survival from precision alone.
IAEA (2013) / professional recommendations	Education and supervised clinical preparation are qualification requirements.	Universal implementation or guaranteed safety through credentials.
Zech et al. (2018) / cross-institutional empirical study	Internal and external discrimination can differ; site information can confound predictions.	A universal external-performance penalty or clinical-outcome benefit.

Note. Original analytical synthesis. Evidence categories are not interchangeable; no common quantitative scale is implied.

Analysis: Three Changes in What Must Be Justified

Reconstruction and the limits of anatomical appearance

CT alters the evidentiary structure of imaging rather than simply sharpening a projection. Hounsfield (1973) describes calculating cross-sectional absorption values from numerous transmission readings. The resulting representation depends on how measurements are acquired and mathematically combined. The defensible claim is that reconstruction recovers information unavailable in the same form from a conventional projection; it is not that a computed section provides direct, assumption-free access to tissue.

The preserved-brain episode in Hounsfield's (1979) lecture is particularly instructive. Formalin enhanced the readings and exaggerated the apparent tissue differences. Subsequent checking with fresh specimens retained evidence of anatomical differentiation while qualifying its magnitude. Here experimental correction did not oppose innovation. It established which aspect of the promising result could reasonably travel from laboratory preparation toward biological interpretation.

This example grounds the first distinction in the proposed framework. Physical validity concerns whether a system measures and represents its target properties adequately under specified conditions. Inferential adequacy concerns whether those properties support the interpretation assigned to them. A convincing image can fail the second condition even when the measurement itself is technically successful.

Qualification and the location of responsibility

Professional recommendations identify calibration, verification, technical supervision, and other responsibilities of clinically qualified medical physicists (IAEA, 2013). The analytical significance extends beyond occupational classification. Where clinical outputs depend on calibrated instruments and

ISSN: 2181-3337 | SCIENTISTS.UZ

computational procedures, institutions must establish who is competent and responsible for evaluating those procedures and recognizing their limits.

IAEA (2013) recommendations make education and clinical training integral to qualification. Their relevance here is structural: acquiring equipment and acquiring a dependable service are different achievements. The recommendations should not be converted into an unsupported assertion that formal certification alone guarantees safe outcomes. They instead identify responsibilities that equipment purchase cannot discharge.

Clinical warrant thus includes an institutional dimension. A technically appropriate procedure may lack a sufficient implementation basis if the organization cannot establish who evaluates its performance, recognizes its limitations, and responds when conditions change. This article proposes that such responsibility belongs inside the analysis of innovation, not in a closing paragraph about implementation.

Prediction and the portability of evidence

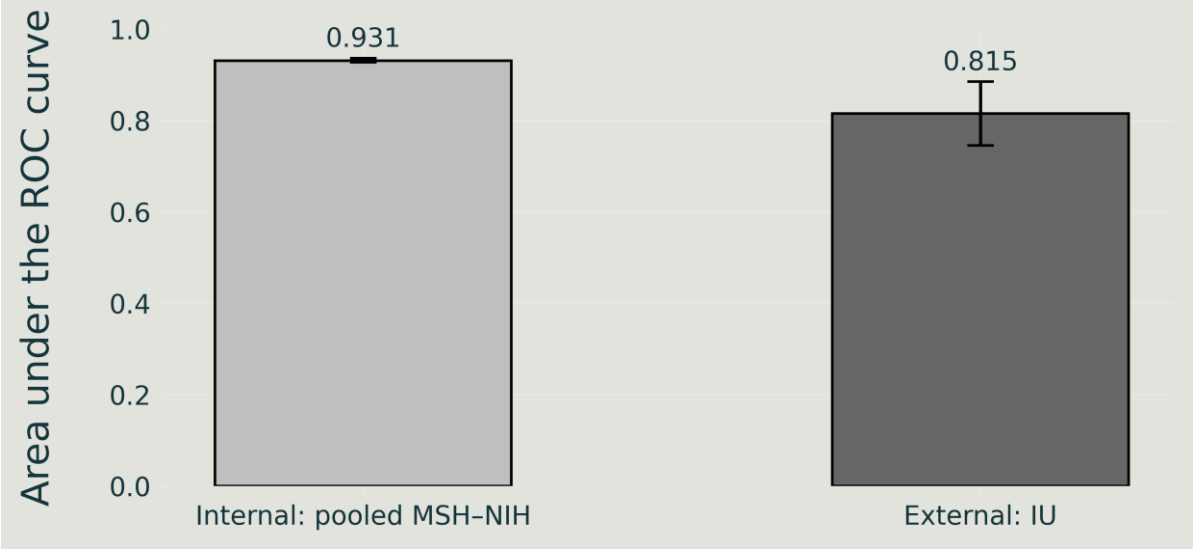
Zech et al. (2018) provide an empirical counterexample to treating improved internal performance as sufficient evidence of wider readiness. Their jointly trained model achieved an internal AUC of 0.931 (95% CI [0.927, 0.936]) on the pooled Mount Sinai Hospital–National Institutes of Health test set, compared with 0.815 (95% CI [0.745, 0.885]) on the external Indiana University set. Figure 1 displays the estimates without suppressing uncertainty or truncating the axis.

The point-estimate difference is 0.116 AUC units. It is not an 11.6% reduction in diagnostic accuracy: AUC measures discrimination across thresholds rather than the fraction of correctly classified examinations. Nor does the comparison isolate one cause of degradation. The study's additional experiments support concern about site-associated information and prevalence differences, but the observed contrast should not be treated as a universal penalty for external deployment.

The deeper result is that the internal evaluation and the deployment question are different questions. A pooled test set can preserve relationships among site, acquisition characteristics, and labels that a new institution does not share. Consequently, improved internal discrimination can coexist with a weaker basis for extrapolation. The evidence does not establish that imaging AI inevitably fails; it establishes that success must be stated with its evaluation setting attached.

Figure 1

Internal and External Discrimination of the Jointly Trained Pneumonia Model



Note. Original redraw of estimates reported by Zech et al. (2018). Error bars indicate reported 95% confidence intervals. MSH = Mount Sinai Hospital; NIH = National Institutes of Health Clinical Center; IU = Indiana University Network for Patient Care. Both estimates concern the jointly trained MSH–NIH model; they are not results from different models. No patient-level reanalysis was performed.

ISSN: 2181-3337 | SCIENTISTS.UZ

Discussion: An Evidence–Warrant Account of Progress

The cases resist a simple chronology in which measurement problems are solved first, professional issues later, and algorithmic issues last. Each introduces a different location at which evidential confidence may become unwarranted. Specimen preparation can change what contrast means; inadequate clinical preparation can weaken implementation; institutional differences can make predictive performance less portable. These are not interchangeable failures, but all interrupt the passage from output to justified action.

Figure 2 is placed here because its purpose is comparative interpretation rather than historical decoration. Its evidence–warrant atlas juxtaposes documented support, the inference that must be restrained, and the warrant still required. Unlike a flowchart, it does not imply that passing through a sequence guarantees readiness. Unlike a radar plot, it avoids manufacturing comparable numerical dimensions from a technical report, professional guidance, and a clinical AI study. The design contribution is the explicit visual separation of what sources establish from what an evaluator must still justify; no claim is made that the matrix form itself is unprecedented.

The framework has two practical implications. First, technology assessment should specify the decision for which adequacy is claimed. Better visualization, stronger discrimination, and demonstrable benefit to patients occupy different evidentiary positions. Hounsfield's (1979) discussion of radiotherapy planning already recognized the distinction: greater planning precision did not by itself establish improved longer-term survival.

Second, education should connect physics with the evaluation of inference. Teaching reconstruction solely as an algorithm misses the question of what its outputs authorize; teaching AI solely through performance metrics misses the conditions under which those metrics remain informative. The IAEA's (2013) emphasis on academic and clinical preparation supplies an institutional basis for this proposal, although the curricular recommendation developed here has not been experimentally evaluated.

For future research, the atlas can generate concrete questions rather than readiness scores. A comparative implementation study could examine whether evaluation reports document acquisition conditions, interpretation limits, intended populations, and accountable personnel. Such work would need a prespecified coding protocol and independent assessment of agreement. The present article establishes neither the reliability of that protocol nor the effect of using it on clinical outcomes.

Figure 2

Evidence–Warrant Atlas of Medical Imaging Practices

Published evidence	Inference not established	Additional warrant required
CT reconstruction Hounsfield (1973, 1979) Cross-sectional representation; specimen preparation can alter apparent contrast.	Anatomical plausibility alone establishes biological interpretation or improved patient outcomes.	Check acquisition and specimen conditions; distinguish planning precision from demonstrated benefit.
Professional qualification IAEA (2013) Academic education and structured clinical training are recommended.	Equipment availability or credentials alone guarantee a dependable clinical service.	Assign responsibility for calibration, performance assessment, and response to changing conditions.
Pneumonia prediction Zech et al. (2018) The same model has different internal and external discrimination.	Internal AUC establishes transportability or patient benefit at another institution.	Evaluate the intended population and setting; investigate site-associated predictive signals.

ISSN: 2181-3337 | SCIENTISTS.UZ

Note. Original qualitative analytical design based on Hounsfield (1973, 1979), IAEA (2013), and Zech et al. (2018). The left column summarizes documented support; the middle and right columns express the authors' interpretive synthesis and proposed evaluation obligations. Cell size, position, and shading encode no magnitude, causal strength, readiness score, or temporal ordering.

Limitations

The documentary corpus is purposive and small. It cannot establish the frequency of validation failures, global workforce adequacy, or the complete development of medical physics. The CT cases support a focused interpretation of reconstruction rather than an exhaustive account of imaging history. One AI study illustrates transportability concerns but cannot characterize every algorithm, modality, or contemporary deployment. The proposed framework has not been validated as an assessment instrument.

The analysis compares different kinds of evidence. A technical account documents instrument design, professional guidance states recommendations, and an empirical study estimates model performance. Their convergence supports a conceptual argument, not a pooled causal estimate. Moreover, institutional accountability is developed here as an interpretive implication of the sources, not as an independently tested explanation of patient outcomes. Prospective evaluation would require operational definitions, a prespecified protocol, and evidence that the resulting judgments can be applied consistently.

Conclusion

Medical physics becomes clinically consequential when physical capability is joined to defensible interpretation and accountable use. CT demonstrates the productive role of reconstruction while exposing the need to test what apparent contrast signifies. Qualification recommendations locate reliability within trained clinical practice. Cross-institutional AI evaluation shows why success within one evidentiary setting does not automatically authorize use in another.

The proposed evidence–warrant framework replaces neither technical performance assessment nor clinical evaluation. It specifies their relationship: the output must be credible, its interpretation must fit the intended claim, and its use must be justified in the relevant setting. Read this way, the history of medical physics is not merely a history of seeing more. It is a history of learning what a particular form of seeing permits medicine to conclude.

REFERENCES

1. Hounsfield, G. N. (1973). Computerized transverse axial scanning (tomography): Part 1. Description of system. *British Journal of Radiology, 46*(552), 1016–1022. https://doi.org/10.1259/0007-1285-46-552-1016
2. Hounsfield, G. N. (1979, December 8). *Computed medical imaging* [Nobel lecture]. Nobel Foundation. https://www.nobelprize.org/uploads/2018/06/hounsfield-lecture.pdf
3. International Atomic Energy Agency. (2013). *Roles and responsibilities, and education and training requirements for clinically qualified medical physicists* (IAEA Human Health Series No. 25). https://www.iaea.org/publications/10437/roles-and-responsibilities-and-education-and-training-requirements-for-clinically-qualified-medical-physicists
4. Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine, 15*(11), Article e1002683. https://doi.org/10.1371/journal.pmed.1002683

ISSN: 2181-3337 | SCIENTISTS.UZ