

Chapter 15 / Capítulo 15

Primum non nocere: Software engineering
for medicine
ISBN: 978-9915-9876-8-2
DOI: 10.62486/978-9915-9876-8-2.Ch15
Pages: 246-261



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Chapter 15. Real-World Evidence and Economic Evaluation

Annier Jesús Fajardo Quesada  

Eduardo Antonio Hernández González 

René Herrero Pacheco 

Rolando Javier Álvarez Pérez 

1 Granma University of Medical Sciences. Granma, Cuba.

2 University of Medical Sciences of Pinar del Río. Pinar del Río, Cuba.

“An algorithm that works perfectly in the laboratory and fails in practice is not a working algorithm; it is a mirage. The truth—r evidence of artificial intelligence in medicine—lies not in the clean, carefully selected data of research, but in the noise, variability, and imperfection of real-world clinical practice.”

There is a very large gap between an algorithm's performance under controlled conditions and its effectiveness in real-world clinical practice. In the laboratory, the data are clean, the variables are welldefined, and the protocols are clear. In the real world, data are missing, measurements are imperfect, patients do not follow the rules, and doctors skip the protocols. An algorithm that showed an area under the curve of 0.95 in the validation study may decrease to 0.70 when implemented in a hospital with limited resources. This gap between efficacy (what the algorithm can achieve under ideal conditions) and effectiveness (what it actually achieves in practice) is among the greatest challenges in digital health, and it can be addressed only through studies that capture the complexity of the real world.

Real-world evidence (RWE) has become a central concept in the evaluation of health technologies. It refers to information derived from data collected in routine clinical practice, not in controlled trials. It includes data from electronic health records, patient registries, administrative databases, wearable devices, and other sources that reflect what actually happens when the technology is used under normal conditions. This evidence complements that from clinical trials, providing information on how interventions work in diverse populations, in varied contexts, and over time.

In this chapter, we explore the evaluation of medical software in the real world and its economic dimension. We begin with real-world validation, analyzing the differences from laboratory validation and the methods to address them. We will continue with economic evaluation and health technology assessment (HTA), which determine whether the software is worth its cost and whether it deserves to be funded by healthcare systems. We address how to demonstrate that the software is worth its cost, which is crucial for its adoption. Finally, we discuss economic evaluations specific to public and insurance systems, which have different criteria and processes. This is not an exhaustive guide to health economics but rather an introduction to the concepts that any developer must understand so that their software has a chance of being adopted.

15.1 VALIDATION IN THE REAL WORLD, NOT JUST IN THE LABORATORY

The history of artificial intelligence applied to medicine is littered with failed technologies that should never have been adopted clinically. Algorithms that published dazzling metrics in high-impact journals, with areas under the curve that bordered perfection and sensitivities that surpassed the expert human eye, only to collapse spectacularly when someone tried to use them in a real hospital, with real patients, amid the chaos of a night shift. This isn't about fraud or dishonest researchers; it's about an uncomfortable truth that the tech community is slow to accept: the lab lies—or at least tells only part of the story. Controlled conditions, curated data, and standardized protocols create an artificial world where algorithms may seem infallible, but that world does not exist outside the pages of scientific articles. Real-world validation is not an optional phase of development; it is the only filter that separates genuinely useful tools from digital mirages.

The distance between the laboratory and everyday clinical practice is not a gap; it is an abyss. In the laboratory, the data are clean, complete, and labeled by consensus experts and sourced from calibrated equipment and carefully selected patients. In the real world, data are the messy byproduct of urgent medical care: missing values because the patient was unconscious when blood pressure was attempted, uncertain diagnoses because the confirmatory test could not be performed, multiple conditions interacting in unpredictable ways, and incomplete records because staff were overwhelmed. An algorithm trained in the purity of the laboratory faces this chaos with the naivety of a tourist who expects everything to work just like at home and fails with the inevitability of a guarded prognosis.

15.1.1 THE SIX FRONTIERS OF FAILURE: WHY THE LAB DOES NOT PREDICT THE REAL WORLD

Data quality: a variable that is obsessively controlled in the laboratory but left to chance in

practice. In controlled studies, every measurement follows strict protocols: blood pressure is taken with the patient seated after five minutes of rest using a recently calibrated sphygmomanometer by specifically trained personnel. In the rush of an overwhelmed emergency department, the same blood pressure may be measured while the patient is lying in a hallway, immediately after being transported on a stretcher, using an automatic device of questionable calibration, by a nurse who is simultaneously attending to three other patients. The algorithm trained on laboratory data is expected to be consistent; it can be interpreted as a signal when it is noise or as noise when it is a signal. Robustness to these variations cannot be assumed; it must be demonstrated through explicit validation in environments where inconsistency is the norm.

Data integrity, or more precisely, its systematic incompleteness. Research studies design forms that capture all relevant variables for all participants; empty fields are exceptions that require justification. In real clinical practice, missing data is the rule: the lab test was not ordered because the physician considered it added no value in that specific context; the record was not completed because the computer system crashed; and the variable was not measured because the staff was occupied with another critical patient. An algorithm that requires a complete set of inputs to produce an output becomes a useless ornament at the most critical moment, precisely when data tend to be missing. To handle absence, feature engineering must incorporate strategies from the design phase, such as context-based intelligent imputation, probabilistic models that marginalize unobserved variables, or feature selection that prioritizes those available in 100% of real clinical cases.

The heterogeneity of the population—a diversity that controlled studies systematically eliminate through exclusion criteria that leave out precisely the most challenging patients. Frail elderly individuals with multiple comorbidities, patients with language or cultural barriers that hinder communication, people with mental illnesses that affect adherence, and immigrants with inaccessible medical histories—all are excluded from studies because of methodological complexity, yet they are precisely those who need care the most in real-world practice. An algorithm trained on a homogeneous population of young adults with a single, well-defined disease approaches real patients such as a child who has studied only arithmetic and suddenly encounters algebra. External validation in diverse populations is not a methodological luxury; it is the only way to know whether the algorithm has generalizable utility or is merely a tool for selected cases.

Variability in clinical practice—a reality that research protocols seek to eliminate and that real-world medicine celebrates as an expression of individualized clinical judgment. In controlled studies, all patients received the same standardized treatment, the same follow-up tests, and the same discharge criteria. In practice, each physician develops their own style, each department adapts protocols to its resources, and each hospital prioritizes them on the basis of its specific pressures. An algorithm that assumes uniform practice will fail when it encounters a physician who makes different decisions for valid reasons that the algorithm does not know. Adaptability to context—or at least the explicit characterization of how performance varies across different contexts—is essential for real-world utility.

The algorithm's own effect on practice—a form of feedback that does not exist in the laboratory, where the algorithm is a passive observer. When introduced into a real-world setting, the algorithm becomes an active agent that modifies the behavior of those who use it. Clinicians may place excessive trust in it, delegating critical judgment to the machine and losing the ability to detect the algorithm's errors when they occur. Alternatively, they may systematically distrust it, ignoring the recommendations and maintaining their previous practices. They may learn from the algorithm and improve their performance even when they are not using it, or they may become dependent and degrade their previous skills. These effects are an integral part of the real-world impact of the technology, but they are invisible in laboratory studies where the algorithm does not interact with humans making decisions. Evaluating them requires study designs that measure not only the model's technical performance but also how clinical behavior changes in its presence.

Time drift—the recognition that medicine is a constantly flowing river, not a static lake. Clinical guidelines are updated, new treatments emerge that alter outcomes, populations fall ill in different ways, and epidemics shift the prevalence of conditions. An algorithm trained on 2020 data may face a transformed clinical reality in 2025 where its predictions no longer apply. Validation cannot be a one-time event culminating in regulatory approval; it must be a continuous process of monitoring and adaptation. Unlike drugs, algorithms can and must be updated, retrained, and adjusted to changes in the environment. However, this requires a postimplementation surveillance infrastructure that is rarely planned with the same attention as the initial development.

15.1.2 DESIGNS FOR REALITY: HOW TO CAPTURE THE TRUTH WITHOUT LOSING CONTROL

To cross these six frontiers and obtain valid evidence on how algorithms actually perform, research methodology must evolve beyond classic laboratory designs. Observational studies using real-world data are the fundamental tool of this new approach. These studies analyze the algorithm's performance under completely natural conditions, without researcher intervention regarding who receives the test or how the result is recorded. They capture authentic variability, real-world noise, and the real-world limitations of clinical practice. They allow us to evaluate what happens when the algorithm encounters unselected patients with incomplete data in diverse contexts. However, this richness of reality entails known methodological vulnerabilities: selection biases regarding who accesses the algorithm, information biases regarding how results are documented, and confounding biases due to unmeasured factors that simultaneously affect the algorithm's indication and the outcome.

To mitigate these limitations without abandoning reality, researchers turn to hybrid designs that maintain some control while preserving pragmatism. Before-and-after studies, which compare the period prior to implementation with the period after implementation at the same center, control for hospital characteristics that remain constant over time, although they remain vulnerable to general temporal trends. Interrupted timeseries designs analyze the evolution of indicators before and after the intervention, looking for changes in level or slope that suggest a causal effect beyond the preexisting trend. Pragmatic trials represent perhaps the most sophisticated synthesis, maintaining randomization as a guarantee of causal inference but relaxing inclusion criteria to reflect the real-world population, allowing for flexibility in implementation according to the local context, and measuring outcomes that matter in daily practice beyond laboratory surrogate end points.

Pragmatic trials deserve particular attention because they reconcile scientific rigor with external validity. Unlike explanatory trials, which seek to demonstrate biological efficacy under ideal conditions, pragmatic trials evaluate effectiveness under real-world conditions of use. Their inclusion criteria are broad, reflecting the algorithm's actual indications; their protocols are flexible, allowing adaptation to local workflows; and their outcome measures are relevant to clinical and health policy decisions. They are more costly and complex than observational studies are, but they provide higher-quality evidence for implementation decision-making. A well-designed pragmatic trial can answer not only whether the algorithm works technically but also whether it improves important outcomes when used by real clinicians on unselected patients.

Continuous postimplementation monitoring completes the validation cycle. Once the algorithm is in use, automated systems must compare its predictions with observed outcomes, detecting signs of performance degradation before they cause harm. This surveillance requires a data infrastructure that enables longitudinal tracking, a statistical methodology to detect significant changes in the model's calibration or discrimination, and clear procedures for deciding when an algorithm needs recalibration, updating, or withdrawal. Artificial intelligence can even be used to monitor itself, applying anomaly detection techniques to the algorithm's own predictions to identify when the environment has changed enough to call its validity into question.

15.1.3 SHARED RESPONSIBILITIES: DEVELOPERS, CLINICIANS, AND THE VALIDATION PACT

For software developers, real-world validation must transcend the category of a postdevelopment activity to become an integral component of the product lifecycle. From the algorithm's conception, plans must be made regarding how external validation will be conducted, what real-world data sources will be needed, what methodological designs will be used, and what financial and human resources will be needed. This early planning avoids the temptation to launch a product "validated" solely on internal data or in academic competitions. Collaborating with clinicians from the design phase ensures that validation is conducted in relevant contexts, using outcome measures that are both clinically meaningful and employing procedures that do not unduly disrupt clinical practice. Methodological humility is essential: recognizing that the laboratory does not predict the real world, that development metrics do not guarantee clinical utility and that negative validation results are learning opportunities, not personal failures.

For clinicians, real-world validation is the only guarantee that the tools they use are trustworthy. An algorithm validated within their own hospital, with their specific patients, within their particular context of resources and practices, generates operational confidence that no publication in a foreign journal can replace. That is why it is crucial for clinicians to actively

participate in these processes, not as passive subjects of research but as coproducers of valid knowledge. They must see the results of local validation before adopting new technologies; they must report discrepancies between the algorithm's predictions and their clinical experience; they must contribute to identifying cases where the algorithm fails in unexpected ways. Validation is not the sole responsibility of those who write code; it is a shared responsibility of those who care for patients.

The future of real-world validation points toward increasingly close integration between research and daily clinical practice. Electronic health record systems, which are becoming increasingly universal and standardized, can be configured as platforms for continuous research where each use of the algorithm contributes to its cumulative evaluation. Adaptive “ ” designs allow the algorithm to be adjusted in real time on the basis of observed results, merging research with quality improvement. Federated learning methodology allows models to be trained and validated across multiple institutions without centralizing sensitive data, facilitating large-scale external validation studies while preserving privacy. All of this is technically achievable, but it requires sustained investment, genuine institutional collaboration, and the political will to prioritize evidence over speed to market.

We arrive at an inescapable conclusion: real-world validation is neither a regulatory formality nor a bureaucratic obstacle to commercial launch. It is the fundamental mechanism that separates technological promise from real health benefits and genuine innovation from a profitable mirage. An algorithm that works in the laboratory but not in practice is not a frustrated breakthrough; it is a design flaw that should never have left the controlled environment. An algorithm implemented without adequate real-world validation is an undeclared experiment on patients who did not consent to be test subjects. Building the bridge between the laboratory and real life is the responsibility of all the actors in the digital health ecosystem, and doing so well is the only way for artificial intelligence to finally fulfill its promise of improving human health in a demonstrable and sustainable manner.

15.2 WHEN CLINICAL EXCELLENCE IS NOT ENOUGH

After years of effort, we have developed an algorithm that provides almost-supernatural precision in which patients with heart failure will be readmitted within the next thirty days. Rigorously designed clinical studies have shown that physicians using the system can intervene proactively and reduce readmissions by twenty-five percent. Clinicians are enthusiastic, publications in high-impact journals are racking up citations, and international conferences are requesting our presence. Everything indicates that we have a resounding success on our hands. However, when we present this technological marvel to the procurement department of the public healthcare system, the response we receive baffles us: “Interesting. How much does it cost? How much does it actually save? What resources will another patient lose out on if we invest in this?” These questions reveal an uncomfortable truth: in the world of public health, clinical excellence is necessary but insufficient. For an innovation to reach patients, it must pass an additional filter that is as rigorous as the scientific one but based on a different logic: health technology assessment (HTA), a process that translates medical success into sustainable value for the system.

Health technology assessment is neither a bureaucratic whim nor a traba to innovation; it is the institutional response to a dilemma that healthcare systems have faced ever since medical technology began advancing faster than the resources available for it. In essence, HTA asks not only whether something works but also whether its effectiveness justifies its cost, whether it is sustainable on a population scale, whether it distributes its benefits equitably, and whether it aligns with the values and priorities that society, through its democratic representatives, has decided to prioritize. In most developed countries with public health systems, regulatory approval of a technology by the FDA or the EMA is only the first step; the second, which is decisive for patients' actual access, is a positive evaluation by the national HTA agency, which determines whether the health system should, can, and wants to pay for it.^(1,2,3)

15.2.1 THE GENEALOGY OF VALUE

The history of HTA dates back to the 1970s, when the healthcare systems of developed countries began to face an unsustainable tension: the technological explosion in medicine, with

its promises of unprecedented cures and improvements, clashed with the reality of finite budgets that were growing more slowly than expectations. The question “Does it work?” was no longer sufficient; the question “Is it worth it?” arose, and with it, the need for disciplines and processes that could answer it in a systematic, transparent, and reproducible manner was met. Since then, HTA has evolved from a marginal academic exercise into a central institution of health policy, with specialized agencies exerting decisive influence over which technologies reach citizens and which remain in the limbo of technical but not financial availability.

In Europe, this institutionalization takes on particularly powerful forms. The UK’s NICE (National Institute for Health and Care Excellence) has become a global benchmark, with the authority to recommend or reject technologies for the entire national health system on the basis of cost-effectiveness analyses that establish explicit thresholds of acceptability. The HAS (Haute Autorité de Santé) in France combines scientific evaluation with judgment regarding the “service médical rendu,” the actual added therapeutic value. IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) in Germany applies German methodological rigor to technology comparisons. These agencies, with their public procedures, detailed reports, and binding decisions, shape the environment in which any medical innovation seeking public funding must navigate. To ignore them is to set oneself up for commercial failure from the outset. (2,4)

For medical software developers, understanding this ecosystem is just as strategic as mastering machine learning architectures. An algorithm that is technically brilliant and clinically validated but fails to meet the cost-effectiveness threshold set by the NICE or its local equivalent may face an insurmountable barrier: regulatory approval exists, but public funding does not—which, in predominantly public systems, amounts to practical nonexistence for most patients. Therefore, incorporating the HTA perspective from the earliest stages of development is not a matter of business maturity; it is a condition for commercial survival. This means designing clinical studies with the aim not only of demonstrating biological efficacy but also of generating the specific data required by HTA economic models: information on implementation costs, impact on resource utilization, effects on quality of life measured with validated instruments, and appropriate comparators from the perspective of the healthcare decisionmaker. (3,4)

15.2.2 CORE: HOW VALUE IS MEASURED IN HTA

The beating heart of HTA is economic evaluation, a discipline that seeks to translate all dimensions of a technology—clinical and nonclinical—into a common metric that allows alternatives to be compared and resource allocation decisions to be made. The most widespread form of this translation is cost-effectiveness analysis, which compares the incremental costs of a new technology with the incremental benefits it generates, typically expressed in quality-adjusted life years (QALYs). A QALY represents a year of life lived in perfect health; years spent with disability or pain are weighted with lower values depending on the severity of the condition. The cost-per-QALY ratio tells us, in understandable terms, how much it costs to purchase health through the evaluated technology. (4)

Acceptability thresholds vary across countries and contexts, but they exist and are operationally decisive. In the United Kingdom, NICE traditionally uses a range of twenty to thirty thousand pounds per QALY, although it may adjust this for technologies that represent significant advances or benefit particularly vulnerable populations. In Spain, the Ministry of Health’s evaluation reports typically consider similar ranges in euros. If a technology falls below the threshold, the funding recommendation is favorable; if it falls above, it may be rejected or require price negotiation, restricted to subgroups where it is more cost-effective, or made conditional on the collection of additional evidence. These seemingly cold numbers determine the fate of innovations that have taken years of work and millions in investment. (3,4)

Calculating the cost-effectiveness of medical software presents specific challenges that models designed for drugs do not adequately capture. While a drug acts directly on the patient’s physiology with relatively predictable effects, medical software operates through complex mediations: it modifies clinicians’ behavior, alters organizational workflows, and generates ripple effects that spread throughout the system. Its impact on health outcomes is indirect, dependent on the implementation context, and mediated by factors such as user adherence, the quality of technical integration, and the system’s responsiveness to generated alerts. This makes economic models inherently more complex, requiring more structural assumptions and extensive sensitivity analyses. However, this complexity also presents opportunities: software that improves procedural

efficiency can generate economic benefits that go beyond direct clinical outcomes—arguments that HTA is increasingly able to appreciate.

15.2.3 FULL PICTURE: COSTS, SAVINGS, AND WHAT CAN'T BE QUANTIFIED

A rigorous economic evaluation of medical software must account for a range of costs that go far beyond the initial license price. Acquisition costs include not only user licenses but also technical installation, integration with existing hospital information systems, the potential need for additional hardware, and any infrastructure modifications that may be needed. Implementation costs encompass training for clinical and technical staff, managing organizational change, reduced productivity during the learning period, and resources dedicated to resolving initial issues. Maintenance costs include software updates, ongoing technical support, performance monitoring, and periodic recalibration of the algorithm. Finally, opportunity costs represent the value of alternative resources we forgo by investing in this software: what other interventions would we stop funding? What staff would we not hire? All these components must be inventoried and quantified honestly, not just the list price appearing in the contract.

On the benefits side, potential savings must be identified and quantified with equal rigor but also with methodological caution. Software that reduces hospital readmissions generates direct savings in length of stay and avoids procedures and untreated complications. A system that speeds up diagnosis saves costs from unnecessary diagnostic tests, days of work disability, and deferred treatments that become complicated. A tool that optimizes prescription reduces pharmaceutical costs and adverse events due to inappropriate medication. However, these savings do not materialize automatically; they depend on the healthcare organization's ability to reallocate the freed-up resources, on healthcare professionals adapting their practice to the new possibilities, and on the financing system not discouraging a reduction in activity (as occurs in fee-for-service systems). The analysis must distinguish between theoretical savings, potentially achievable savings, and savings actually realized in specific contexts.

Beyond quantifiable monetary costs and savings, HTA explicitly considers benefits that resist monetization but are central to the decision. Improvements in patients' quality of life, measured using validated instruments such as the EQ-5D or the SF-36, are incorporated into the calculation of QALYs and thus into the denominator of the cost-effectiveness ratio. The reduction in anxiety and uncertainty for patients and families, although difficult to capture in standardized surveys, can be assessed in qualitative analyses that complement quantitative models. Greater satisfaction and reduced burnout among healthcare professionals—organizational benefits that the software can generate, although traditionally underrepresented in HTA—are gaining recognition as determinants of the system's sustainability. Equity in access, the reduction in disparities among population groups, and the alleviation of the burden on informal caregivers—all these aspects are weighed in modern HTA, although their specific weighting varies across agencies and contexts.⁽⁵⁾

The perspective from which the economic analysis is conducted is a methodological choice with substantial consequences. The public health system perspective, which is common in evaluations for funding decisions, considers only the costs that fall directly on the public purse: payments to providers, personnel costs, pharmaceuticals, and funded devices. The broader social perspective also includes costs to patients and families: work time lost due to caregiving, transportation expenses to access treatments, and opportunity costs of informal care. The individual patient's perspective focuses on their subjective experience and the outcomes they value, even if the system does not pay for them directly. The choice between these perspectives can drastically alter the conclusions: software that reduces hospital costs but increases the burden on family caregivers may appear cost-effective from the system's perspective but not from the social perspective; an algorithm that improves quality of life but requires frequent travel may be valued differently by urban and rural patients.

15.2.4 PREPARING FOR SCRUTINY: EVIDENCE STRATEGY FROM DESIGN

For medical software developers, adequately preparing for HTA means a strategic investment in evidence generation that begins before the first line of code is written and continues throughout the entire product lifecycle. In the early phases, this involves designing clinical trials that not only demonstrate biological efficacy but also systematically collect data on real-world implementation

costs, healthcare resource utilization, and effects on quality of life measured using instruments accepted by HTA agencies. It means selecting appropriate comparators from the healthcare decision-maker's perspective: not just placebo or no intervention but also the therapeutic or organizational alternatives that the system currently uses or would consider using. This means planning subanalyses by risk groups, implementation contexts, and resource levels, which allow for identifying where the software is particularly cost-effective and where it is not.

The collection of information on the actual costs of technical integration and operational maintenance is particularly critical and frequently underestimated. HTA economic models are only as good as the data that feeds them, and developers who can provide estimates on the basis of real-world implementation experience, rather than theoretical projections, gain decisive credibility. This requires close collaboration with early implementation sites, rigorous documentation of all the resources consumed, and honesty about the setbacks and unforeseen costs that inevitably arise. An economic model that incorporates real implementation data—even if it is less favorable than initial projections—is more convincing than one based on unvalidated optimistic assumptions.

For clinicians involved in the development and evaluation of medical software, HTA may initially seem like an inappropriate intrusion of economic logic into the realm of healing. However, understanding that healthcare resources are finite and that every euro spent on one technology is a euro unavailable for another transforms this perception. HTA is not a cold accounting exercise but an institutionalized attempt to allocate resources as fairly and efficiently as possible, maximizing the population health generated with the available budget. Clinicians play an irreplaceable role in this process: they can contribute tacit knowledge about what actually works in practice, about what patients genuinely value beyond measurable endpoints, and about the hidden costs of implementation that abstract economic models may overlook. Their participation in HTA committees, in expert groups that inform evaluations, and in the generation of realistic evidence is essential for coverage decisions to reflect the complexity of medical practice.

The future of HTA applied to medical software lies in the development of specific methodologies that capture the unique characteristics of digital interventions. Traditional economic models, designed for drugs with predictable life cycles and direct biological effects, do not adapt well to technologies that evolve weekly, generate complex indirect effects, and whose value depends critically on the implementation context. New approaches are being developed that incorporate accelerated technological obsolescence, network effects and organizational externalities, as well as the importance of local adaptation. The involvement of patients and caregivers in evaluation processes is also advancing through public deliberation techniques and preference analysis so that HTA decisions reflect not only technical efficiency but also social values regarding justice, equity, and human dignity.

15.3 HOW TO DEMONSTRATE THAT THE SOFTWARE IS WORTH THE COST

The history of medical software development is littered with technical promises that never managed to become widespread healthcare realities. Algorithms that seemed magical in controlled demonstrations, that outperformed the best human specialists in academic competitions, that accumulated citations and recognition in high-impact journal publications but that at the end of the day remained on digital shelves, used only by those who created them and by a few early enthusiasts. They did not fail because of technical flaws; they failed because their creators never learned to answer the question that every payer, public or private, eventually asks: Why should I allocate my scarce resources to this rather than to the thousand other needs competing for them? In a world of limited budgets, where every euro spent on a technology is literally a euro unavailable for another potentially life-saving intervention, demonstrating clinical efficacy is merely the starting point. The real challenge is to build a comprehensive value e that convinces decisionmakers that this specific investment deserves priority over legitimate and competing alternatives.

Building this case is not an exercise in marketing or persuasive rhetoric; it is a rigorous process that combines solid scientific evidence, sophisticated economic analysis, and a deep understanding of the institutional and political context where decisions will be made. It requires the language of algorithmic precision and software architecture to be translated into the language of budgetary impact, systemic efficiency, and return on investment in population health. It requires segmenting heterogeneous audiences—from clinicians concerned with practical utility to managers obsessed with cash flow—and building specific arguments that resonate with each

group's priorities. In addition, methodological honesty about limitations and risks is needed because credibility, once lost with public procurement officials or hospital executive committees, is virtually irrecoverable.

15.3.1 KNOWING YOUR AUDIENCE

The essential starting point for any value-demonstration strategy is a comprehensive mapping of who the stakeholders are who will ultimately decide on the adoption and funding of the software and what decision-making factors influence each of them. This segmentation is not a minor tactical detail; it is the difference between a message that resonates and drives action and one that elicits indifference or rejection. A hospital's medical director, for example, operates under a logic of clinical efficiency and quality of care: they are concerned with reducing variability in practice, improving patient safety indicators, and positioning their institution as an innovative leader. The procurement manager of a public health system operates under imperatives of population-level cost-effectiveness and budgetary sustainability: they must demonstrate that every euro spent generates maximum health benefits for the covered population and that decisions can be defended in the face of audits and before courts of auditors. The frontline clinician, who will ultimately use or ignore the software, values immediate practical utility, seamless integration into their already overloaded workflow, and respect for their professional autonomy.

A message designed to convince the medical director of the reduction in adverse events may be completely irrelevant to the procurement manager, who needs to see the five-year budget impact analysis. An argument that resonates with clinicians regarding the reduction in administrative time may seem futile for medical directors, who seek to affect institutional quality indicators. Therefore, the value communication strategy must be multifaceted and adaptive, maintaining a core of common evidence but presenting it with differentiated emphasis depending on the audience. The supporting materials should include different versions: a financial executive summary for managers, a clinical impact analysis for medical directors, user testimonials for clinicians, and a systems-level overview for health policy makers. Consistency among these messages is essential; there can be no contradictions between what is promised to one group versus another, but the presentation must be tailored to maximize the perceived relevance for each stakeholder.

15.3.2 THE FIVE PILLARS OF THE VALUE PROPOSITION

The first fundamental pillar is the rigorous quantification of clinical benefits, expressed in terms that go beyond vague generalities and are grounded in specific, credible numbers relevant to decision-making. It is not enough to state that "the algorithm improves the diagnosis of sepsis"; one must be able to say, backed by published evidence or well-designed proprietary studies, that "the system detects sepsis in its early stages an hour and a half before conventional diagnosis, which is associated with a 15% reduction in 30-day mortality and a two-day decrease in the average ICU stay." These concrete numbers, especially when expressed in units that decisionmakers recognize as meaningful (lives saved, complications avoided, days of hospitalization reduced), have a persuasive power that qualitative statements cannot match. The evidence must come from sources that the audience respects: publications in indexed journals, independent external validations, or data from real-world implementations in similar institutions.

The second pillar requires the translation of these clinical benefits into economic terms that are understandable to those who manage budgets. Although a life saved is priceless in the abstract, healthcare systems need to know the incremental cost of saving it using this specific technology to compare it with alternatives. This translation requires economic models that estimate, on the basis of demonstrated clinical effects, the impact on resource utilization: avoided hospital stays, canceled unnecessary diagnostic procedures, complications that do not require treatment, and staff freed from administrative tasks. Each of these effects has an estimable unit cost within the local healthcare system, and aggregating them allows for the construction of a case for savings or, at the very least, a reasonable cost per unit of benefit. This is not an exact science—economic models incorporate uncertainties that must be made explicit—but it is an unavoidable necessity for dialog with administrators.

The third pillar is comparative analysis with legitimate alternatives because software is never evaluated in a vacuum. The relevant question is not whether the algorithm adds value in

the abstract but whether it adds more value than the other ways of addressing the same problem that the system could fund. These alternatives may include standard practices without additional technology, other competing software already available, nontechnological organizational interventions (such as manual screening protocols or rapid response teams), or even the reallocation of resources to other higher-priority areas. An incremental cost-effectiveness analysis comparing the software to the most relevant alternative from the decision-maker's perspective is the standard tool for this evaluation. If the software costs fifty thousand euros per QA-LY gained and the current alternative costs thirty thousand, the decision to adopt it will require additional arguments regarding benefits not captured in the formal economic analysis.⁽⁴⁾

The fourth pillar addresses the short-term budgetary impact, a practical consideration that often determines the feasibility of adoption regardless of long-term cost-effectiveness. Even if a software solution generates net savings over five years, its implementation may require significant upfront investments in licenses, technical integration, large-scale staff training, and workflow reorganization—costs that a rigid annual budget simply cannot absorb. A detailed budget impact analysis, projecting cash flows year by year over the implementation horizon, makes it possible to identify whether there are critical years of intensive investment and whether deferred financing mechanisms or staggered payment models can be designed to make adoption viable. This analysis demonstrates to decision makers that their actual budget constraints are understood and that practical solutions have been considered, not just theoretical arguments.

The fifth pillar incorporates benefits that resist direct monetary quantification but can tip the balance when formal economic analyses are very close. Improved healthcare staff satisfaction and retention are difficult to monetize but critical in the context of professional shortages. The reduction in administrative burden allows for more time for direct patient care. Greater equity in access to quality diagnostics across centers with different levels of specialization. The generation of data and knowledge that facilitates future innovations in the system. These benefits must be described qualitatively in a compelling manner, supported where possible by satisfaction surveys, staff turnover analyses, or equity studies. Although they do not directly enter into the cost-per-QALY calculations, they form part of the comprehensive value argument that decisionmakers consider, especially in systems where the sustainability of the professional model and equity are explicitly prioritized values.⁽⁴⁾

15.3.3 ANTICIPATION, TRANSPARENCY, AND LONG-TERM COMMITMENT

Healthcare procurement decisionmakers are skeptical of the profession and of accumulated experience. They have witnessed too many product presentations promising radical transformations that failed to deliver results, too many optimistic economic analyses that fell apart during actual implementation, and too many technologies abandoned after significant initial investments. Therefore, an effective value demonstration strategy actively anticipates the objections that will inevitably arise and has prepared solid, honest responses backed by evidence or planning.

Typical objections include doubts about generalizability: “Do your results from teaching hospitals apply to our context with more limited resources?” Response: present subanalyses at the resource level or propose a local pilot study before widespread adoption. Concerns about implementation costs: “What if the costs of integrating with our legacy systems exceed your estimates?” Response: offer integration budgets with pricecap guarantees or shared-risk models where the provider assumes part of the cost overrun. Skepticism about adoption by clinicians: “What if our doctors simply ignore the system?” Response: Present adoption data from similar implementations and offer training programs and organizational change initiatives with contractual adoption metrics. Every objection that is anticipated, addressed, and answered convincingly builds credibility; every objection that captures the presenter off guard erodes trust.

Transparency about limitations is a counterintuitive but powerful strategy. No software is perfect, no study is definitive, and no economic analysis is exact. Attempting to hide these universal truths breeds suspicion that more serious problems are being concealed, and when they are inevitably discovered, the presenter's credibility is destroyed. It is far more effective to proactively acknowledge limitations: “Our algorithm is less accurate in patients with advanced kidney disease, a subgroup representing eight percent of the target population; we are working on a version specifically for this group, but in the meantime, we recommend enhanced clinical monitoring in these cases,” and to explain why, despite these limitations, the software deserves consideration. This honesty builds trust and allows for a constructive discussion about risk

management rather than a defensive confrontation.

The value proposition must include an explicit long-term vision that extends beyond the initial commercial transaction. The adoption of medical software is not a one-time event but the beginning of a relationship that will last for years, with implications for maintenance, updates, performance monitoring, and adaptation to changes in the clinical and technological environment. Demonstrating that one has a robust plan for this evolution, guaranteed technical support commitments, shared development roadmaps, user feedback mechanisms that drive continuous improvements, and assurances that the software will remain compatible with emerging standards is especially important for public systems that cannot afford to be trapped with obsolete technologies or defunct vendors. This long-term vision transforms the vendor's perception from a transactional supplier to a strategic partner.

15.3.4 PARTNERS, REAL-TIME INSIGHTS, AND THE FUTURE OF VALUE

Clinicians who have used the software are the most powerful allies in building the value proposition. A hospital director may be skeptical of a sales representative's claims, even if they are accompanied by impressive PowerPoint slides, but will listen with genuine interest to their own doctors if they endorse the technology on the basis of their first-hand experience. That is why investing in ensuring that early adopters are not only satisfied but also equipped to articulate their experience is an essential strategy. Their testimonials about specific cases where the software made a difference, their real-world usage data demonstrating sustained adoption, and their stories of patients who benefited—which can be shared (while respecting confidentiality)—are invaluable assets that no economic analysis can replace. The case for technical and economic value must be combined with the case for human value, as told by those on the front lines of care.

The future of demonstrating value in medical software points toward greater methodological standardization and closer integration of real-time evidence. As healthcare systems become fully digitized, it will be technically possible to continuously monitor the impact of implemented software, automatically compare outcomes between exposed and unexposed patients, detect signs of performance degradation, and adjust funding decisions on the basis of current data rather than historical studies. Value-based payment models, where the manufacturer receives revenue contingent on actual clinical outcomes achieved, are gaining ground as a mechanism for aligning incentives. In these models, demonstrating value is not—a presale exercise but an ongoing process: the software is funded as long as it demonstrates that it works, and funding ceases if it stops doing so. This logic, although technically demanding in terms of measurement infrastructure and complex contractual agreements, represents perhaps the purest way to ensure that payment is made only for value actually generated.

15.4 ECONOMIC EVALUATION FOR PUBLIC SYSTEMS AND INSURANCE

The difference between a tax-funded public health system and a private insurer competing for customers is not merely an accounting matter of who signs the checks. It is a profound divergence in the ultimate objectives they pursue, in the incentives driving decisionmakers, and consequently, in the criteria they use to assess whether a technology merits adoption. A public system, accountable to an entire population and funded with resources derived from legitimate fiscal coercion, seeks to maximize population health with available resources, prioritizing equity in access and distributive justice. A private insurer, which must attract customers in a competitive market and generate returns for its shareholders, seeks to offer a product perceived as attractive while maintaining the company's financial sustainability. These different objective worlds generate different questions, different decision-making processes, and different evaluation cultures that medical software developers ignore at their own commercial and, ultimately, ethical risk.

Understanding these differences is not an academic exercise in comparative health policy; it is a strategic necessity for anyone who want their technological innovation to move beyond the prototype phase and actually reach patients. Software designed and evaluated solely from the perspective of a public system can fail spectacularly in the private market, and vice versa. The population-level cost-effectiveness metric that convinces a European HTA agency may be irrelevant to a U.S. insurer concerned with retaining high-spending customers. Demonstrating a reduction in health inequalities that excites a public system director may not resonate with a private insurance sales director. Consciously navigating between these logics, adapting the

message and evidence to each context without compromising the product's integrity, is among the most sophisticated skills a digital health innovator must develop.

15.4.1 AT THE HEART OF THE STATE: THE LOGIC OF PUBLIC SYSTEMS

Public health systems, which are predominant in Europe and are expanding in many Latin American countries, operate under the logic of maximizing population health subject to budgetary constraints and principles of equity. The decision to fund a new technology, including medical software, typically goes through a formal evaluation process by an HTA agency, such as the NICE in the United Kingdom, the HAS in France, or the IQWiG in Geria, which publicly analyzes the evidence and issues binding or highly influential recommendations. This process is deliberately transparent, with explicit criteria, published methods, and opportunities for manufacturers to provide information and answer questions. The central question guiding the evaluation is whether the technology provides sufficient health value relative to its cost, compared to available alternatives, considering the system's budgetary constraints and equity obligations.(2, 3)

The dominant metric in these contexts is the cost per quality-adjusted life year (QALY), which allows for the comparison of very different technologies on a common scale: how much it costs, in terms of system resources, to gain one year of life in perfect health through this specific intervention. There are implicit or explicit thresholds, typically between twenty thousand and thirty thousand pounds or euros per QALY in most European systems, which function as rough filters: technologies below the threshold have a high probability of a favorable recommendation, while those above require additional justification or price negotiation. For medical software, calculating this ratio involves economic models that project the algorithm's clinical effects on mortality and morbidity, convert them into QALY gains, and relate them to all incremental system costs attributable to the software.(4)

However, the final decision in public systems is rarely based solely on the calculated cost-effectiveness. HTA agencies explicitly consider additional factors that capture dimensions of value not reflected in the economic ratio. Equity is crucial: software that disproportionately benefits disadvantaged groups, reduces health disparities, or improves access in underserved areas may receive a favorable recommendation even if its cost per QALY exceeds the standard threshold. The severity of the condition being treated matters: technologies for conditions with high mortality or disability, especially if no effective alternatives exist, are given greater flexibility. The total budgetary impact—how much it would cost to implement the technology at a population-wide scale and with full—can block adoptions that are cost-effective individually but unsustainable when aggregated. The quality of available evidence determines the confidence in effectiveness estimates. In addition, social value, defined through public consultations and stakeholder engagement, can tip the balance when uncertainty exists.(4)

The decision-making process in public systems is inherently long and complex, involving multiple sequential stages that can span months or years. The technical evaluation by the HTA agency generates a preliminary report on which the manufacturer can comment. A formal recommendation is submitted to public procurement officials, who initiate price negotiations on the basis of demonstrated value and international price comparisons. Final inclusion in the list of covered benefits requires contractual agreements specifying conditions for implementation, monitoring, and potential revisions. Throughout this process, the manufacturer has structured opportunities to provide additional information, address methodological questions, and propose risk-sharing agreements or outcome-based payments. This protracted process requires dedicated teams, strategic patience, and comprehensive documentation that cannot be improvised at the last minute.

15.4.2 IN THE HEALTHCARE MARKET: THE LOGIC OF PRIVATE INSURERS

Private insurers, which predominate in the U.S. system but are also present as a complement in Europe and expanding in Latin America, operate under a radically different logic that combines cost concerns with commercial imperatives of competitiveness. Here, there is no centralized agency that evaluates technologies for the entire system; each insurer makes independent decisions on the basis of its own criteria, which are often opaque and vary from one entity to another. The central question is not only about population health value but also about the product's appeal to consumers and companies purchasing insurance and its impact on the

insurer's financial sustainability.

Cost-effectiveness remains relevant, but it is evaluated with a shorter time horizon and a different perspective. Insurers are particularly interested in the impact on their operating costs over one or two years, not just in long-term health value. Software that reduces hospital admissions is valuable because it lowers payments to providers; a tool that improves the management of chronic diseases is attractive because it reduces the use of costly services. However, the satisfaction perceived by policyholders also matters greatly: software that improves the patient experience, facilitates access to care, and provides a sense of control and personalization can be a powerful marketing tool that differentiates the insurer's product in a competitive market, even if it does not generate measurable direct savings. In this context, "effectiveness" includes customer retention and the acquisition of new policyholders—business metrics foreign to the vocabulary of public systems.

The decision-making process at insurers is faster and less formal than that at public HTA agencies but also more opaque and unpredictable. Decisions are made by internal medical committees, often without public guidelines or explicit thresholds, on the basis of available evidence, direct negotiations with manufacturers, and strategic considerations regarding market positioning. For the software developer, this means that simply submitting a standardized evidence dossier is not enough; it is necessary to build business relationships with each relevant insurer, understand their specific priorities—are they focused on aggressively reducing costs or on differentiating themselves through perceived quality? Do they prioritize technologies for chronic or acute care ()—and adapt the value proposition to each context? With respect to what convinces United Healthcare may not work with Aetna or Cigna, segmentation and personalization are essential.

Payment models in the private sector also differ substantially. While public systems typically negotiate prices on the basis of demonstrated value using international benchmarks, private insurers operate according to a more commercial logic: prices based on expected usage volume, exclusivity discounts, bundling agreements with other products, and, increasingly, outcome-based payment models in which the manufacturer assumes part of the risk that the software will not deliver the promised benefits. This commercial flexibility can be an opportunity for developers who can offer risk-sharing structures, but it also entails contractual complexity and the need for outcome monitoring systems that many manufacturers have not yet developed.

15.4.3 BETWEEN TWO WORLDS: MIXED SYSTEMS AND LATIN AMERICAN COMPLEXITY

Latin American health systems present particularly complex configurations where public provision, private insurance, and social security systems with hybrid characteristics coexist, often with permeable boundaries. In Mexico, INSSTE and IMSS operate under a social security framework for formal workers, while the public health system serves the uninsured population; in Brazil, the public SUS coexists with private health plans that cover a significant portion of the population; in Chile, FONASA and ISAPRE represent the public–private duality with complex interactions. Each of these components has its own funding rules, technology assessment processes, and adoption criteria.⁽⁵⁾

In some Latin American countries, there are HTA agencies that evaluate technologies for the public component of the system—CONITEC in Brazil and INVIMA in Colombia, with expanding evaluative functions—but private insurers and employer-sponsored plans may make independent decisions on the basis of commercial criteria. In others, funding decisions are decentralized at the provincial or state level, creating a patchwork of regulations that manufacturers must navigate. Local evidence is of paramount importance here: a study conducted in the United States or Europe, however rigorous, has limited validity in contexts where epidemiology, personnel costs, and clinical practices differ substantially. Generating evidence with local patients, conducting cost analyses using country-specific data, and making comparisons with standard practices within that specific system are essential investments for credibility. ^(2,3,6)

Transparency emerges as a common value across all these systemic complexities. Both public system evaluators and medical committees of private insurers in Latin America as well as in Europe or North America, value manufacturers being open to their methods, their primary data, and the limitations of their analyses. Hiding inconvenient information, presenting economic analyses biased by unwarranted optimism, or denying access to underlying data is the quickest way to destroy credibility and close doors. Conversely, being explicit about how results were obtained, acknowledging uncertainties and wide confidence intervals, and offering independent

audits of the data builds the trust that facilitates evaluation and, eventually, adoption.

Price negotiation is another cross-cutting element, although it takes different forms in each context. The list price is rarely the final price; in public systems, negotiation is formal and value-based, while in private insurers, it is more commercial and volume-based, but in both cases, it is inevitable. The developer must have a clear pricing strategy, with a maximum price that reflects the product's demonstrated value and a minimum price below which the business' sustainability is called into question. Within this range, the flexibility to offer discounts for volume commitments, staggered payments aligned with implementation, or shared-risk models where the final price depends on the results obtained can be decisive in closing deals.

15.4.4 STRATEGY, LOCAL EVIDENCE, AND THE FUTURE OF EVALUATION

For medical software developers, navigating this complexity requires a strategic approach that begins long before the product is ready for market. It is essential to identify, from the design phase, which markets are priorities and what types of payers predominate in them to guide the generation of evidence and the construction of the value proposition. Clinical studies must be designed with the specific questions that evaluators from each system will consider: equity analyses for European public systems, patient satisfaction studies for U.S. insurers, and implementation data in resource-limited settings for Latin American markets. Proactive relationships must be built with key stakeholders—HTA agencies, insurance procurement directors, scientific societies that influence practice guidelines, and patient associations advocating for coverage—before their approval is urgently needed.

Patience and persistence are essential virtues. Evaluation processes in public systems are deliberately designed to be lengthy to ensure democratic deliberation; processes at private insurers can be unpredictable because of staff turnover and changes in corporate strategy. The developer must be prepared to adapt to changes in evaluation criteria, to new evidence emerging from competitors, and to budget crises that shift priorities without losing sight of the long-term vision or the product's integrity. They must be prepared to generate ongoing evidence because economic evaluation is no longer a one-time event prior to commercialization but a process that continues postimplementation, especially in outcome-based payment models.

The role of clinicians in these ecosystems can be decisive in multiple ways. Their testimony regarding the software's real-world utility in practice, their participation in local studies that generate contextualized evidence, and their endorsement before scientific societies that issue usage recommendations—all contribute to the product's credibility. However, they can also help developers understand the particularities of the local system: the informal actors who influence decisions, the unwritten processes that are sometimes more important than official protocols, and the political sensitivities that determine which technologies are prioritized. Their on-the-ground knowledge is an asset that cannot be replaced by international consultancies.

The future of economic evaluation for medical software points toward greater harmonization between countries and between the public and private sectors. Initiatives such as EUnetHTA in Europe seek to coordinate evaluations and reduce duplication of effort, although implementation is slow and complex. Value-based payment agreements, where the final price depends on the health outcomes actually achieved in real-world practice, are gaining ground in both sectors because they align the incentives of manufacturers and payers. Dynamic evaluations, which are automatically updated as digital health systems generate continuous real-world evidence, are technically feasible and increasingly in demand by payers who do not want to be trapped in decisions on the basis of outdated evidence. All of these findings suggest a future where health technology assessment will be more efficient, more transparent, and more genuinely aligned with the value that technologies generate in practice.

We arrive at a conclusion that transcends business strategy: economic evaluation, in all its institutional forms, is not an end in itself but a means to ensure that healthcare resources—whether derived from taxes, insurance premiums, or direct payments—are allocated in the most beneficial way.

REFERENCES

- US Food and Drug Administration. Artificial intelligence in software as a medical device [Internet]. Silver Spring (MD): FDA; updated Jan 6, 2025 [cited May 14, 2025]. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>
- Oortwijn W, Sampietro-Colom L, Habens F, Cranovsky R, Dehne M, Granados A, et al. The new definition of health technology assessment: a milestone in international collaboration. *Int J Technol Assess Health Care*. 2021;37(1):e54. <https://doi.org/10.1017/S0266462321000416>
- National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual. PMG36 [Internet]. London: NICE; 2022 Jan 31 [updated 2025 Jul; cited 2025 Mar May 14]. Available from: <https://www.nice.org.uk/process/pmg36>
- Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1-503. <https://doi.org/10.3310/hta19140>
- EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208. [https://doi.org/10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9)
- Colombia. Decree 4725 of December 21, 2005, regulating the system of health registrations, marketing authorization, and health surveillance of medical devices for human use. *Official Gazette*. Dec. 26, 2005;(46.134). Available at: <https://www.invima.gov.co/productos-vigilados/dispositivos-medicos/dispositivos-medicos-equipos-biomedicos>