Universal Library of Business and Economics Monograph

ISSN: 3064-9951 | Volume 3, Issue 3

Open Access | PP: 78-113

DOI: https://doi.org/10.70315/uloap.ulbec.2026.0303006

An Integrated Financial Analytics Framework for Health Plans - Risk Pool Estimation, Reserve Modeling, and Strategic Forecasting in Medicare Advantage and Exchange Lines of Business

Gitanshu Yadav

Abstract

U.S. health plans operating in Medicare Advantage and the Affordable Care Act Exchange manage financial risk through three functions that are usually kept apart: IPA-level risk-pool estimation, IBNR reserve modeling, and state-level forecasting. This fragmentation causes the same underlying quantities to be estimated inconsistently, producing reserve inadequacy, forecast variance, and weak decision support. This monograph develops an integrated financial analytics framework that connects the three functions within a single operational architecture built on one organizing discipline: each shared quantity is estimated once and flows between layers through three explicit cross-layer feedback loops. The framework specifies a per-IPA financial liability exposure calculation with a leading, concurrent, and lagging early-warning hierarchy; a dynamic IBNR model with an interpretive reporting layer carrying audience-tailored templates; and a state-level, scenario-based Exchange forecast with regulatory drivers embedded rather than adjusted for. As far as the author is aware, no publicly disclosed methodology connects these three layers through such feedback loops. The framework is presented as a replicable design, buildable on the data and platforms a plan already possesses, and validated qualitatively within a large U.S. managed care administrator. It offers finance leadership a blueprint for consolidating fragmented analytics into a coherent system that improves reserve adequacy, forecasting accuracy, and risk-sharing transparency.

Keywords: *Medicare Advantage, ACA Exchange, Risk Pool, IBNR Reserve, Medical Loss Ratio, Financial Forecasting, Health Plan Finance, Integrated Analytics.*

PREFACE

This monograph sets out a framework I developed in the practice of health-plan financial analysis. Working across Medicare Advantage and the Affordable Care Act Exchange, I found that the hardest problems were rarely inside any single calculation. They arose in the spaces between the calculations: where a risk-pool settlement drew on one view of unpaid claims and the reserve drew on another, or where a forecast never received the reserve development that would have corrected it. The functions were each competent, and the organization still reconciled the same numbers more than once.

The framework presented here is my attempt to close those spaces. It organizes risk-pool estimation, reserve modeling with cross-functional reporting, and state-level forecasting into three connected layers, and it wires them together with explicit feedback loops so that a signal produced in one becomes an input to another. The discipline at its centre is simple to state and demanding to practise: estimate each shared quantity once, and let it flow.

The intended reader is the practitioner who lives with these problems: the finance, actuarial, contracting, and operations professional in a U.S. managed care organization. The book is written to be used. Its methods are presented as a design that a plan can build on the data and platforms it already has, and its claims are stated with the restraint that honest practice requires. I hope it proves useful to colleagues working to make health-plan finance more coherent than the tools they inherited.

Citation: Gitanshu Yadav, "An Integrated Financial Analytics Framework for Health Plans - Risk Pool Estimation, Reserve Modeling, and Strategic Forecasting in Medicare Advantage and Exchange Lines of Business", Universal Library of Business and Economics, 2026; 3(3): 78-113. DOI: https://doi.org/10.70315/uloap.ulbec.2026.0303006.

Table of Contents

List of Abbreviations

Abbreviation	Expansion
ACA	Affordable Care Act
AER	Administrative Expense Ratio
ASOP	Actuarial Standard of Practice
CMS	Centers for Medicare & Medicaid Services
CMS-HCC	CMS Hierarchical Condition Category (risk-adjustment model)
FFS	Fee-for-Service
IBNR	Incurred But Not Reported (reserve)
IPA	Independent Physician Association
MA	Medicare Advantage
MAPD	Medicare Advantage Prescription Drug (plan)
MLR	Medical Loss Ratio
NAIC	National Association of Insurance Commissioners
PMPM	Per Member Per Month
RADV	Risk Adjustment Data Validation
RAF	Risk Adjustment Factor
SAP	Statutory Accounting Principles
SEP	Special Enrollment Period

INTRODUCTION

Relevance of the Topic

The financial management of U.S. health plans has grown structurally more complex, and the analytics used to manage it have not kept pace with that complexity. Two lines of business have expanded in parallel and now dominate the managed care landscape. Medicare Advantage pays plans through risk-adjusted capitation and delegates much of the financial risk to provider organizations, with Medicare spending an estimated 22 percent more per Medicare Advantage enrollee than it would under fee-for-service (Medicare Payment Advisory Commission, 2024). The Affordable Care Act Exchange market has become a large and volatile individual market in its own right, drawing 21.4 million plan selections in the 2024 open enrollment period, a 31 percent increase over the prior year (Centers for Medicare & Medicaid Services, 2024b). Both lines operate under continually evolving CMS regulation, from risk-adjustment model revisions to the medical-loss-ratio rule that requires insurers to spend at least 80 or 85 percent of premium on medical care (Centers for Medicare & Medicaid Services, 2026).

Against this complexity, financial analytics in most plans remain fragmented. Risk-pool estimation, reserve modeling, and forecasting are performed by separate functions, using separate data, and connected only loosely. The fragmentation is not a cosmetic inconvenience. It produces reserve inadequacy when the reserve does not see the enrollment shift the forecast already detected, forecast variance when the forecast does not receive the reserve development the actuary already observed, and weak decision support when financial signals do not reach the operating functions that could act on them. As the two lines of business grow and the regulatory environment tightens, the cost of managing them through disconnected analytics grows with them. What is needed is not a better tool for any single function but a unified, architectural approach that addresses the complexity of managed care finance at a structural level rather than through isolated tactical improvements.

The stakes of getting this right are concentrated in the same quantities that the fragmented functions each estimate separately. A risk pool is settled on an estimate of unpaid claims; a reserve is an estimate of the same unpaid claims; a forecast rests on the enrollment and utilization assumptions that also condition the reserve. When these estimates are produced independently, the plan does not merely duplicate work, it holds inconsistent versions of figures that determine settlement, solvency, and pricing at once. In a business where Medicare Advantage revenue is large and sensitive to coding and selection, and where the medical-loss-ratio rule converts efficiency into a binding financial threshold, the difference between a connected and a disconnected set of estimates is the difference between anticipating a financial problem and discovering it after it has settled. The intended reader of this monograph is the practitioner who lives with that difference: the finance, actuarial, contracting, and operations professional in a U.S. managed care organization who must make risk-pool, reserve, and forecasting decisions that depend on one another but are usually produced apart.

Aim and Objectives of the Research

The aim of this monograph is to develop and substantiate an integrated financial analytics framework that connects IPA-level risk-pool estimation, IBNR reserve modeling with cross-functional reporting, and state-level Exchange forecasting within a single coherent architecture. The framework is intended as a replicable methodology for a working health plan, not as an abstract model.

Seven objectives structure the work. The first is to formalize the theoretical and regulatory foundations on which health-plan financial analytics rest. The second is to review existing industry and actuarial practice and to identify the specific gaps that recur across the three functions. The third is to design the integrated architecture that closes those gaps. The fourth, fifth, and sixth are to develop the methodology of each of the three operational layers in turn. The seventh is to specify the cross-layer feedback loops that connect the layers, to validate the framework through implementation experience, and to draw out the managerial recommendations that follow. These objectives map directly onto the chapters that follow.

Scientific Novelty and Practical Significance

The scientific novelty of the work is architectural. Each individual component of the framework has analogues in industry and in the literature: plans estimate risk pools, actuaries model IBNR reserves, and planning teams forecast the Exchange line. What is new is the integration of all three into a single operational architecture with explicit cross-layer feedback loops. As far as the author is aware, no publicly disclosed methodology connects IPA-level risk-pool estimation, dynamic IBNR reserving with cross-functional reporting, and state-level Exchange forecasting through the three feedback loops specified here. The claim is stated with deliberate restraint: the absence of a public methodology does not prove that no proprietary system exists, but the integrated architecture, and the discipline of estimating each shared quantity once and letting it flow between layers, is a contribution that the publicly available literature does not contain.

The practical significance is a replicable blueprint. For a U.S. health plan seeking to consolidate fragmented financial analytics into a coherent system, the framework offers a concrete design: specific data-integration rules, a per-IPA exposure calculation with early-warning indicators, a dynamic reserve model with an interpretive reporting layer, a state-level scenario-based forecast, and the feedback loops that connect them. The framework is presented as an engineered design, buildable on the data and platforms a plan already possesses, with the intent of improving reserve adequacy, forecasting accuracy, and risk-sharing transparency across the Medicare Advantage and Exchange lines of business.

Structure of the Monograph

The monograph proceeds in seven chapters. Chapter 1 establishes the theoretical and regulatory foundations of U.S. health-plan financial management. Chapter 2 reviews existing practice and identifies the gaps that motivate an integrated approach. Chapter 3 specifies the framework's architecture and its three cross-layer feedback loops. Chapters 4, 5, and 6 develop the methodology of the three operational layers in turn: IPA-level risk-pool estimation, IBNR reserve modeling with cross-functional reporting, and state-level Exchange forecasting. Chapter 7 turns to implementation, validation, and managerial implications, tracing a single regulatory event through the whole system before setting out a roadmap, an honest account of the framework's limitations, and its forward directions. The progression moves from foundations through architectural design and layer-by-layer methodology to implementation, so that the reader arrives at the operational detail of each layer already understanding the architecture it serves.

CHAPTER 1. THEORETICAL AND REGULATORY FOUNDATIONS OF U.S. HEALTH PLAN FINANCIAL MANAGEMENT

A health plan operating in the Medicare Advantage and Affordable Care Act Exchange markets carries financial risk of a particular kind. Revenue is fixed in advance through administratively set payment formulas, while the cost of care is uncertain, delayed in its reporting, and sensitive to enrollment, utilization, and regulatory change. The financial analytics that a plan builds on top of this structure inherit its logic. Estimating a risk pool, setting a reserve, and forecasting a line of business are not three unrelated exercises. They are three readings of the same underlying question: how much of the premium already collected, or the capitation already paid, will ultimately be consumed by claims that have not yet been settled. This chapter establishes the economic and regulatory baseline that the rest of the monograph builds on, and shows why a plan that manages these questions in separate silos leaves financial signal unused.

The Economics of Managed Care and Capitated Payment Models

Managed care in the United States developed as a response to the cost trajectory of open-ended fee-for-service (FFS) reimbursement. Under FFS, a payer reimburses each service after it is rendered, which places the financial consequence of utilization entirely on the payer and gives providers no direct stake in the total cost of a covered population. Capitation reverses that arrangement. A capitated contract pays a provider organization a fixed amount per member per month (PMPM) in exchange for accepting responsibility for a defined scope of care, so that the provider, not only the payer, holds a share of the utilization risk. The design intent, long recognised in the payment-incentive literature, is to align the provider's financial interest with efficient care rather than with service volume (Robinson, 2001; Cattel et al., 2018), an alignment that reshaped the organizational structure of managed care itself (Robinson & Casalino, 1996; Robinson, 1999).

In the Medicare Advantage program, this arrangement is frequently intermediated by an Independent Physician Association (IPA), a legal entity that contracts with a health plan on behalf of a network of physicians. The plan delegates a portion of financial risk to the IPA through a capitation payment, and the two parties agree on how any surplus or deficit against that capitation will be shared. The mechanics of this risk-sharing pool are central to everything that follows. At the end of a settlement period, the capitation revenue attributed to an IPA is compared against the claims incurred by that IPA's attributed members. When claims run below capitation, a surplus is available for distribution; when they run above, the deficit must be absorbed according to the contract's loss-sharing terms. The settlement is therefore a direct function of two quantities that are themselves estimates: the risk-adjusted revenue owed to the IPA, and the full cost of care incurred but not yet fully paid. Figure 1.1 traces this mechanic. Risk-adjusted capitation revenue on one side and incurred claims, including the portion not yet reported, on the other are brought to a common basis and compared; the resulting surplus or deficit is then divided between the plan and the IPA according to the contract's terms. The figure makes visible why the two inputs, revenue and cost, must be estimated on the same population and period: the settlement is the difference between them, so any inconsistency in how they are measured falls directly into the result.

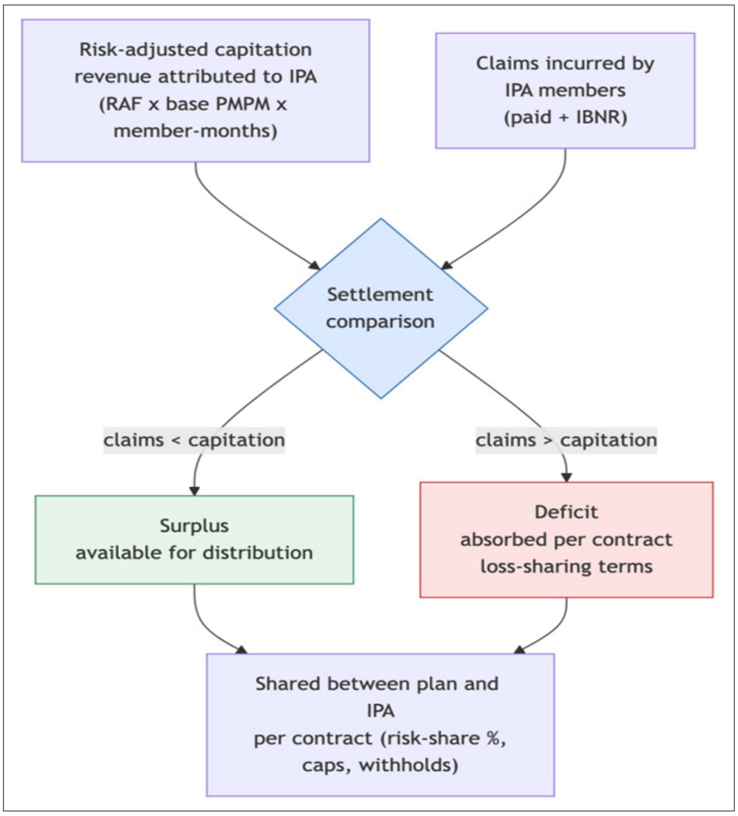

Figure 1.1. The IPA risk-pool settlement mechanic: risk-adjusted capitation revenue is compared against incurred claims (paid plus IBNR); the resulting surplus or deficit is shared between the plan and the IPA according to the contract's loss-sharing terms (author's construction).

The settlement itself is rarely a single clean calculation. A risk-sharing pool typically defines a target, often expressed as a share of capitation or premium available for medical claims, and specifies how results above and below that target are divided between the plan and the IPA. Some contracts cap the provider's downside, so that the plan reabsorbs losses beyond a threshold; others share surplus asymmetrically, or hold back a portion pending run-out of late claims. Each of these terms changes the financial meaning of the same raw experience, which is why a settlement cannot be read from paid claims alone. It must be read through the contract, against a reserve, and with an explicit view of how much cost remains unreported. A plan that estimates the pool without holding these three elements together, contract terms, incurred cost, and reserve, produces a settlement figure that is precise in appearance and unreliable in substance.

Capitation does not remove risk from the system; it redistributes it and changes who must measure it. A plan that pays capitation still bears residual exposure, because a provider organization in financial distress cannot always honour the losses assigned to it, and because the plan remains accountable to regulators for the adequacy of its own reserves. Newhouse and McGuire, assessing the program's performance, note that the financial results of Medicare Advantage depend heavily on how payment is set relative to the risk actually enrolled (Newhouse & McGuire, 2014). Rice and Smith, surveying capitation and risk adjustment internationally, make the same structural point from the design side: a capitation payment is only as sound as the risk-adjustment method that scales it to the population it is meant to cover (Rice & Smith, 2001). The remainder of this chapter follows that thread from the payment formula into the reserve and the forecast.

Table 1.1. Allocation of financial risk under fee-for-service versus capitated payment (author's synthesis of the payment-incentive literature: Robinson, 2001; Cattel et al., 2018; Newhouse & McGuire, 2014).

Dimension	Fee-for-service	Capitation (IPA risk-sharing)
Who holds utilization risk	Payer	Shared between payer and provider organization
Provider incentive	Service volume	Cost-efficient management of a population
Revenue basis	Retrospective, per service	Prospective, risk-adjusted PMPM
Principal estimation task	Claims payment accuracy	Risk-pool settlement and reserve adequacy
Residual payer exposure	Full cost of services	Provider default and reserve adequacy

Risk Adjustment in Medicare Advantage

Because capitation is paid prospectively, the amount must reflect the expected cost of each member, not an average across the whole market. Medicare Advantage accomplishes this through the CMS Hierarchical Condition Category model (CMS-HCC), which maps a member's documented diagnoses into condition categories and combines them with demographic factors to produce a Risk Adjustment Factor (RAF) for that member (Centers for Medicare & Medicaid Services, 2024c). A RAF of 1.0 represents an average-risk beneficiary; a member with multiple documented chronic conditions carries a higher RAF and draws proportionally higher capitation. For the purposes of this monograph, the RAF is treated as a foundational input rather than as a subject in its own right: it is the coefficient that scales revenue, and its role here is to propagate into the risk pool and the reserve, as shown in Figure 1.1.

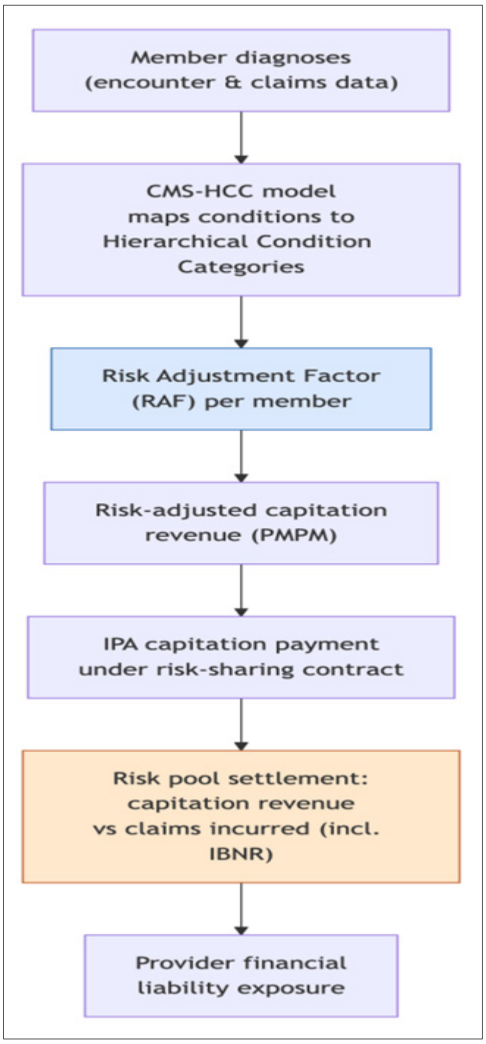

Figure 1.2. Propagation of the Risk Adjustment Factor from member diagnoses through risk-adjusted capitation into IPA-level risk-pool settlement and provider liability exposure (author's construction).

Two features of risk adjustment matter for financial analytics. The first is that the RAF depends on documentation and coding, not only on underlying health. Because higher recorded diagnoses raise revenue, the model creates a well-documented incentive toward more complete, and in some analyses more intense, coding. Geruso and Layton demonstrate that risk scores rise faster in Medicare Advantage than the underlying health of the population would predict, a pattern they attribute to coding behaviour rather than to morbidity (Geruso & Layton, 2017; Layton et al., 2015), and machine-learning approaches to risk adjustment have been explored precisely to reduce such distortions (Kan et al., 2019; McWilliams et al., 2025). Curto and colleagues, examining coding-intensity variation across plans, show that this behaviour is uneven, so that two plans covering similar populations can report materially different risk scores (Curto et al., 2025). For a finance team, this means that RAF-driven revenue carries a quality dimension: revenue supported by weakly documented diagnoses is more exposed to later audit adjustment than revenue supported by well-substantiated ones.

The second feature is that risk scores translate into payment through a formula that regulators revise. Payments to Medicare Advantage plans were expected to increase on average by 3.70 percent, or over $16 billion, from 2024 to 2025 under the applicable CMS rate announcement (Centers for Medicare & Medicaid Services, 2024a). A change of this magnitude, applied through the interaction of the risk score and the county benchmark, moves the revenue side of every risk pool at once. Jacobs and Kronick, comparing risk-based payments with alternative benchmarks, show that the relationship between what a plan is paid and what its members actually cost is neither fixed nor self-correcting (Jacobs & Kronick, 2018). Recent work on machine-learning payment formulas signals that the methodology itself remains under active revision (Andriola et al., 2024). The financial consequence is that a risk pool cannot be settled on the assumption that revenue is stable; the revenue assumption is itself a moving, regulation-driven quantity that any reserve built on top of it must accommodate.

Reserve Estimation Principles: IBNR and Statutory Accounting

The cost side of the settlement is dominated by claims that have been incurred but whose final value is not yet known. The Incurred But Not Reported (IBNR) reserve is the estimated liability for care that members have already received but that has not yet been reported to, or fully paid by, the plan. IBNR exists because of lag: a service rendered in one month may be billed weeks later, adjudicated later still, and adjusted after that. At any financial close, a portion of the true cost of the period is invisible in the paid-claims data and must be estimated. Figure 1.3 shows why. Between the moment care is incurred and the moment its claim is finally paid sits a chain of steps, billing, adjudication, payment, and later adjustment, each of which takes time. The IBNR reserve is the estimate of the cost that has entered this pipeline but not yet emerged as a paid claim, which is exactly the cost a settlement or a financial close would otherwise miss.

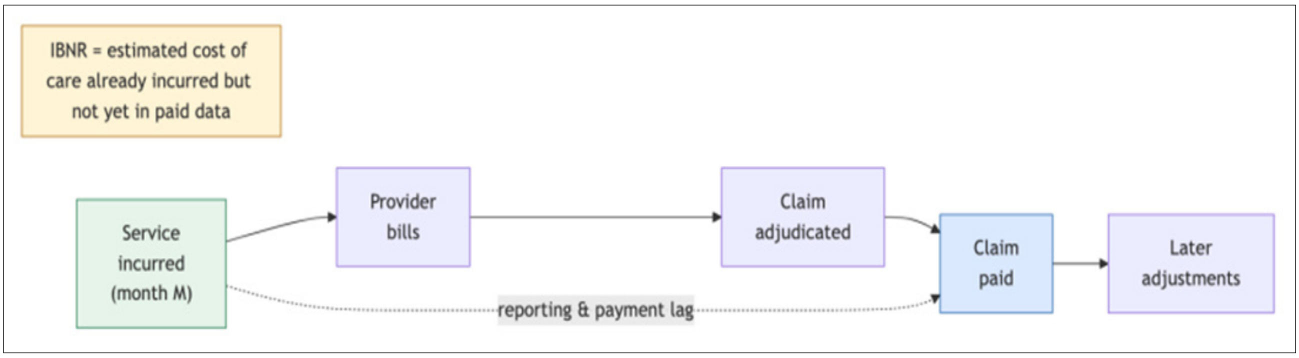

Figure 1.3. Why IBNR exists: the lag between care incurred and claim paid. The reserve estimates the cost of care already incurred but not yet visible in the paid-claims data (author's construction).

The standard estimation approach uses the historical pattern of how claims for a given incurred period accumulate as they are paid. If past experience shows that a given fraction of an incurred month's ultimate cost is typically paid within one month, a larger fraction within two, and so on, those fractions form completion factors that can be applied in reverse to recent months whose claims are still developing. This is the completion-factor, or chain-ladder, family of methods. Mack's distribution-free treatment of the chain-ladder estimator remains the reference point for quantifying the uncertainty around such a projection, rather than treating the point estimate as exact (Mack, 1993, 2008). Where recent periods are too immature for their own paid data to be trustworthy, the Bornhuetter-Ferguson method blends the emerging experience with an a priori expected loss, dampening the volatility that a pure chain-ladder projection would show on the least mature months (Alai et al., 2009). More recent work extends these ideas into stochastic and machine-learning formulations that model the full distribution of the reserve rather than a single figure (Sriram & Shi, 2020; Chukhrova & Johannssen, 2021; Schneider & Schwab, 2024), and Bayesian and solvency-oriented formulations extend the same reasoning to the reserve's full technical provision (England et al., 2012; Dreksler et al., 2015); but the underlying logic that a monograph on financial analytics must respect is unchanged: the reserve is an estimate with a distribution, and its central value depends on assumptions about development, seasonality, and enrollment.

Three adjustments complicate the calculation in a capitated Medicare Advantage and Exchange environment. The first is offsets: amounts such as capitation already paid, pharmacy rebates, or reinsurance recoveries reduce the net liability and must be netted against gross incurred cost rather than ignored. The second is seasonality. Utilization is not uniform across the year, so a completion pattern estimated on annual averages will misstate the reserve for months that sit on a seasonal peak or trough unless the pattern is allowed to vary. The third, and the most consequential in these markets, is enrollment change. When membership grows, shrinks, or shifts in risk composition between periods, the historical development pattern no longer applies cleanly to the current inventory of incurred claims, because the population generating those claims is not the population that generated the historical pattern. A static reserve model, calibrated once and left unchanged, absorbs these movements silently and produces reserve error precisely when enrollment is most volatile. This is the analytical weakness that a dynamic reserve model, developed in a later chapter, is built to address, and it is why reserve modeling cannot be treated as a periodic accounting entry divorced from the plan's enrollment and forecasting view.

Statutory accounting places this estimate under formal discipline. U.S. health insurers report on Statutory Accounting Principles (SAP), a conservative framework oriented toward solvency and policyholder protection rather than toward the going-concern presentation of general-purpose accounting. Under this framework the adequacy of claim reserves, including IBNR, is subject to actuarial opinion. The Actuarial Standards Board sets the practice expectations: guidance on estimating incurred health claims governs how the liability is calculated (Actuarial Standards Board, 2017), and guidance on statements of actuarial opinion governs how its adequacy is attested (Actuarial Standards Board, 2021). Commercial reference frameworks such as the Milliman Health Cost Guidelines are widely used to benchmark the assumptions that feed these calculations (Milliman, n.d.). The practical implication for the integrated framework is that the IBNR reserve is not an internal management figure that a plan may set at will. It is a regulated, opinion-bearing quantity, and any analytics layer that produces or informs it inherits that accountability.

The ACA Exchange Market and its Financial Drivers

The Affordable Care Act Exchange market operates on a different payment logic from Medicare Advantage but raises the same estimation questions. Here the plan sells individual coverage at a community-rated premium, receives federal subsidy support on behalf of eligible enrollees, and is subject to a set of rules that constrain how premium may be spent. The central regulatory constraint is the Medical Loss Ratio (MLR). Under the Affordable Care Act, insurers are required to spend at least 80 or 85 percent of premium dollars on medical care and quality improvement, depending on market segment, and to rebate the difference when they fall short (Centers for Medicare & Medicaid Services, 2026). The MLR converts what would otherwise be an internal efficiency metric into a binding financial parameter: a plan that spends too little of its premium on care must return money, while a plan that spends too much erodes its margin. Karaca-Mandic and colleagues, evaluating the MLR as a regulatory target, show that the ratio interacts with plan size and cost structure in ways that make it an imperfect but consequential lever (Karaca-Mandic et al., 2013), and the first decade of the reformed individual market has been analysed extensively for how these rules shape insurer behaviour (George, 2016; Handel & Kolstad, 2022; Geruso & Layton, 2020).

The Exchange also carries risk-transfer machinery that directly shapes what a plan can expect to keep. Risk adjustment moves funds between insurers in the same state market so that a plan enrolling higher-risk members is compensated from plans enrolling lower-risk members, which means a carrier's net revenue depends not only on its own enrollment but on the risk profile of its competitors. Permanent and temporary reinsurance mechanisms, where they operate, further compress the tail of high-cost claims. Oyeka and Wehby, studying state reinsurance programs, find that these arrangements measurably affect premiums and insurer participation, so they are structural inputs to a forecast rather than afterthoughts (Oyeka & Wehby, 2023). Adverse selection compounds the challenge, because enrollment through Special Enrollment Periods (SEP) and mid-year churn can draw a costlier population than the open-enrollment cohort a premium was priced against. Fung and colleagues document exactly this pattern of selection within the individual market (Fung et al., 2018). For financial analytics, the lesson is that an Exchange forecast built on aggregate enrollment and average cost will miss the drivers that actually determine the result, and must instead resolve enrollment, risk transfer, and selection at a level granular enough to see them.

Two further ratios complete the operating picture. The Administrative Expense Ratio (AER) captures the share of premium consumed by non-medical operating costs, and the PMPM cost, already introduced in the capitation context, serves as the common denominator that makes medical and pharmacy costs comparable across states, products, and time. These three quantities, MLR, AER, and PMPM, are the working vocabulary of Exchange financial planning, and they are all forward-looking: a plan sets premium for a coverage year on the basis of projected values, then manages the year against them. Under the Affordable Care Act, ratemaking shifted away from retrospective adjustment toward continuous, assumption-driven projection bounded by the medical-loss-ratio constraint, so that pricing and financial planning became a single ongoing exercise rather than an annual event (Karaca-Mandic et al., 2013). Layton and colleagues, analysing health-plan payment in the marketplaces, describe how the risk-adjustment transfers, reinsurance, and rating rules of the Exchange interact to determine what any single insurer ultimately retains (Layton et al., 2017). Rate-setting is therefore an exercise in projecting enrollment, cost, and regulatory effect simultaneously, which is precisely the task that the forecasting layer developed later in this monograph is designed to support.

The Regulatory and Reporting Environment for U.S. Health Plans

The financial analytics discussed so far do not exist in a neutral space; they are shaped by the reporting obligations a health plan must meet. A Medicare Advantage and Exchange carrier reports simultaneously to several audiences, each with its own definitions and cadence. Statutory financial statements filed with state insurance departments, on forms coordinated through the National Association of Insurance Commissioners, carry the reserve figures and require actuarial certification of their adequacy. The Centers for Medicare & Medicaid Services impose separate reporting tied to bids, risk-adjustment data submission, and medical-loss-ratio attestation. Guidance specific to risk contracts, such as the Healthcare Financial Management Association's statement on accounting and reporting for risk contracts, addresses how risk-sharing arrangements should be reflected in a provider organization's financial statements (Healthcare Financial Management Association, 1997). Table 1.2 summarises how these obligations map onto the analytical layers developed in later chapters.

Table 1.2. Principal statutory and regulatory reporting obligations for a Medicare Advantage and Exchange carrier, and the analytical layer each obligation constrains (author's synthesis of ASB, n.d.-a, n.d.-b; CMS, n.d.-a, n.d.-b; HFMA, n.d.).

Obligation	Primary authority	Core requirement	Analytical layer constrained
Statutory financial statements	State departments / NAIC framework	Certified reserve adequacy under SAP	IBNR reserve modeling
Risk-adjustment data submission	CMS	Accurate, auditable diagnosis data	IPA risk-pool estimation
Medical loss ratio attestation	CMS	80 or 85 percent premium spend, with rebate	Exchange forecasting
Actuarial opinion on liabilities	ASB practice standards	Independent attestation of reserves	IBNR reserve modeling
Risk-contract accounting	HFMA guidance	Consistent treatment of risk-sharing results	IPA risk-pool estimation

The regulatory environment does more than impose paperwork. It fixes the definitions on which internal analytics must agree, and it sets the points at which an estimate becomes an attested figure with external consequences. Born and colleagues, studying how insurers manage underwriting profitability, show that reserve and reporting practices are themselves managerial choices with financial effects, not merely mechanical compliance (Born et al., 2021). This observation motivates the argument of the next chapter. When risk-pool estimation, reserve modeling, and Exchange forecasting each answer to overlapping regulatory definitions but are produced by separate teams working from separate data, the plan reconciles the same underlying quantities more than once, and inconsistencies between those reconciliations become financial risk in their own right. A consolidated theoretical and regulatory baseline, of the kind assembled in this chapter, is what makes the case for an integrated rather than a siloed approach both visible and measurable.

CHAPTER 2. REVIEW OF EXISTING PRACTICES AND IDENTIFIED GAPS IN FINANCIAL ANALYTICS ACROSS U.S. HEALTH PLANS

The foundations in Chapter 1 describe three estimation problems that share a common structure. Industry practice, however, has developed a separate toolset for each, and those toolsets rarely meet. This chapter reviews how large payers and the actuarial profession actually estimate IPA-level risk pools, model IBNR reserves, and forecast the Exchange line, and it identifies the specific limitations that recur across all three. The review draws on public payer disclosures, regulatory and industry reports, and the peer-reviewed literature. It ends by naming the single structural problem, fragmentation, that connects the individual gaps and motivates the integrated framework of Chapter 3.

IPA-Level Risk Pool Estimation in Industry Practice

Public payer disclosures confirm that risk pools and provider risk-sharing are financially material, but they present the mechanics at a level of aggregation that hides per-IPA detail. The large Medicare Advantage organizations, including UnitedHealth Group, Humana, and Centene, describe capitation, risk-sharing arrangements, and medical-cost reserves in their annual filings, yet they report these at the segment or consolidated level rather than by individual provider group (UnitedHealth Group, 2026; Humana, 2026; Centene Corporation, 2025). The scale of the underlying activity is documented in regulatory reports. MedPAC records that by 2024 the average Medicare beneficiary could choose among 43 plans, a proliferation that reflects how many distinct provider arrangements sit beneath the program (Medicare Payment Advisory Commission, 2024). A payer administering risk pools across that many plans and provider groups is reconciling a large number of separate settlement calculations, each governed by its own contract terms.

The financial stakes of getting risk adjustment right at this level are visible in the aggregate. MedPAC estimates that Medicare spends approximately 22 percent more for MA enrollees than it would for comparable beneficiaries in fee-for-service Medicare, a difference driven substantially by coding and favourable selection (Medicare Payment Advisory Commission, 2024), a cost concentration that the wider literature on treating high-need populations has long documented (Reschovsky et al., 2011). Guidance on accounting for risk contracts exists to bring consistency to how these arrangements are reflected in financial statements (Healthcare Financial Management Association, 1997). What the public methodologies do not provide is a per-IPA view that links each provider group's contract terms to its own claims experience and risk-adjusted revenue in a single, forward-looking exposure estimate. The common limitations of current practice are therefore threefold: results are reported at too high a level of aggregation to manage individual provider groups, contract terms are weakly integrated with claims experience, and the orientation is retrospective, describing settlements after they have occurred rather than projecting exposure early enough to act.

The structure of the disclosures themselves explains the limitation. A payer's annual filing reports a consolidated or segment-level medical care ratio, aggregate reserves for claims payable, and a narrative description of capitated and risk-sharing arrangements. These figures answer the questions an investor or a solvency regulator asks: is the book of business profitable in aggregate, and are the reserves adequate at the entity level. They are not designed to answer the question a finance team managing a provider network must ask, which is whether a particular IPA, under its particular contract, is trending toward a loss position that the plan will have to absorb. The public methodology stops at the level where external accountability ends, and internal management need begins precisely there. A plan that relied only on the disclosed, aggregate view would see a risk pool problem after it had already flowed into consolidated results, when the opportunity to renegotiate a contract or adjust a reserve assumption had passed. This is the retrospective character of current practice stated concretely: the information exists, but at a grain and a cadence that make it a record rather than a warning.

Table 2.1. What large-payer public disclosures and regulatory guidance address, versus what per-IPA risk-pool management requires (author's synthesis of UnitedHealth Group, 2026; Humana, 2026; Centene Corporation, 2025; HFMA, n.d.; MedPAC, 2024).

Element	Addressed in public disclosures / guidance	Gap for per-IPA management
Capitation and risk-sharing existence	Yes, at segment or consolidated level	No per-IPA breakdown
Reserve adequacy	Yes, aggregate medical-cost reserves	Not attributed to individual pools
Contract-to-claims linkage	Described in principle	Not operationalized per contract
Orientation	Retrospective settlement	No forward-looking exposure signal
Early-warning capability	Not addressed	Required for proactive engagement

IBNR Reserve Modeling in Industry Practice

Reserve estimation is the most methodologically mature of the three areas. The standard actuarial toolkit, the chain-ladder method and its distribution-free error analysis (Mack, 1993), the Bornhuetter-Ferguson blend for immature periods (Alai et al., 2009), and completion-factor techniques, is well established and codified in actuarial practice standards (Actuarial Standards Board, 2017), and a growing body of work extends these estimators to stochastic, individual-claim, and double-chain-ladder formulations (Carvalho & Carvalho, 2019; Agbeko et al., 2014; De Felice & Moriconi, 2019; Delong et al., 2021; Veprauskaitė & Adams, 2017). These methods are effective at their designed task: converting a history of claim development into a central reserve estimate with a measurable uncertainty band.

The mechanics of the completion-factor approach make its assumptions explicit. For each incurred month, the method estimates the fraction of ultimate cost already paid, then grosses up the paid figure to an estimated ultimate. For mature months, where most claims have run out, the completion factor is close to one and the estimate is stable. For the most recent months, the completion factor is small and uncertain, so a large multiplier is applied to a small and noisy paid amount, which is why the Bornhuetter-Ferguson blend substitutes an a priori expected loss for the least mature periods. Credibility weighting between recent experience and prior expectation is the actuary's lever for managing that instability.

The limitation is not in the methods but in how they are typically deployed. A completion pattern estimated on stable historical data assumes that the population and the payment environment generating current claims resemble those that generated the history. In a capitated Medicare Advantage and Exchange environment that assumption is frequently violated. A mid-year enrollment surge changes the mix of members whose claims are still developing, so the historical completion pattern understates or overstates the true liability depending on the direction of the shift. A benefit or network change alters utilization timing, which distorts the development pattern itself. Because these movements enter the reserve silently, a quarter-end refresh discovers them only after they have already affected the settlement of the risk pools that depend on the same period's cost. The reserve is technically sound and operationally late. Enrollment shifts, benefit changes, and regulatory recalibration move the underlying population, so a reserve model calibrated once and refreshed only at quarter-end absorbs those movements silently. The consequence is reserve error that is largest exactly when enrollment is most volatile, which is also when accurate reserves matter most for risk-pool settlement. A second limitation is organizational: reserve outputs are commonly produced as an actuarial deliverable and handed to finance as a number, without a structured path back into operational planning. The reserve is treated as an endpoint rather than as a signal that should inform forecasting and provider engagement. This separation is the gap that Layer 2 of the framework is designed to close.

Forecasting and Management Reporting in the Exchange Market

Conventional Exchange forecasting operates at an aggregated level, projecting membership and cost for a product line or a national book and adjusting after the fact for regulatory change. The market that this approach is asked to predict is large and volatile. Analysis of the individual market's first decade documents enrollment growth from 10.6 million to 17.4 million, alongside a fall in the uninsured share of the population to 9.2 percent by 2019, a market reshaped repeatedly by policy (Handel & Kolstad, 2022; Obama, 2016). Insurer consolidation and consumer behaviour further move premiums and enrollment in ways a national forecast cannot see (Dafny et al., 2012; Ho et al., 2017), and the reform's effects have been traced through state hospital sectors and household financial stability alike (Duggan et al., 2022; Dodini, 2022). Entry and exit of insurers move premiums materially; the population-weighted second-lowest-price silver premium would have been an estimated 5.4 percent lower in 2011 had a single large insurer participated in every market it was positioned to enter (Dafny et al., 2015). Competition and segmentation effects of this size are lost in a national average.

The reason state-level grain matters is that the Exchange is not one market but a collection of state markets with different rules, different competitive structures, and different risk pools. Rating areas, the presence or absence of a state reinsurance program, the number of participating insurers, and the local health of the population all vary across state lines, and each moves the cost and premium a plan will realise. A national projection collapses this variation into a single average that corresponds to no state the plan actually operates in. When a plan prices and staffs to that average, it over-resources states where its position is strong and under-resources states where cost is running ahead of premium, and it discovers the mismatch only when state-level results diverge from the consolidated forecast.

The limitations of aggregated forecasting follow directly. State-level dynamics, where enrollment, competition, and regulation actually operate, are averaged away, so a forecast can be accurate in aggregate while wrong in every state that matters for a specific plan. Regulatory drivers such as rate rules and risk-transfer parameters are commonly treated as ex-post adjustments layered onto a baseline rather than as structural inputs to the projection. Scenario flexibility is limited, because a static annual budget cannot be re-run quickly as conditions change. The result is a forecast that supports high-level narrative but not the state-by-state pricing and resource-allocation decisions that determine an Exchange plan's financial outcome. Layer 3 of the framework addresses this by moving the unit of projection to the state and embedding regulatory drivers directly.

Cross-Functional Communication of Financial Results

The fourth area is not an estimation method but a reporting practice, and it is where analytical accuracy most often fails to become financial value. In common practice, financial reporting is a downstream activity: results are produced by finance and actuarial teams and distributed to other functions with limited tailoring to what each audience must decide. A reserve estimate delivered as a single figure, or a variance report written in actuarial terms, does not tell a sales, operations, or clinical leader what to do differently. The cost of this gap is real. When financial signals do not reach the functions that can act on them, commission structures drift out of alignment with financial results, operational responses to emerging cost trends are delayed, and strategic decisions are made without the financial context that would have changed them.

Each audience needs a different translation of the same underlying result. A sales organization compensated through a commission structure needs to know how financial performance maps to incentive payments, so that what it is paid to do and what the plan needs it to do stay aligned. An operations team needs variance against plan expressed in operational drivers it can influence, not in reserve terminology. Clinical leadership needs cost trends framed with the clinical context that explains them. A board needs the strategic summary and the scenario range, not the calculation. A single undifferentiated report serves none of these audiences well, because each must translate it privately before acting, and private translation is where financial meaning is lost or distorted.

The gap here is one of connection rather than of content. The estimation layers can be accurate and the reserves well governed, yet if the output is not translated into the specific decisions of each function, the analytics do not change behaviour. What is missing in typical practice is a defined interpretive layer, owned by finance, that converts actuarial and financial output into audience-specific decisions and routes early-warning signals to the functions that can respond. Establishing that layer, and wiring it as an explicit feedback path rather than a one-directional report, is the contribution developed in Layer 2 and formalised as the third cross-layer loop in Chapter 3.

Synthesis of Gaps and the Case for an Integrated Framework

Across the four areas, the specific limitations rhyme. Risk-pool estimation is aggregated and retrospective. Reserve modeling is static and organizationally isolated. Exchange forecasting is national in grain and treats regulation as an afterthought. Cross-functional communication is one-directional and untailored. These are not four independent problems; they are four expressions of a single structural condition. Each function optimises its own output against its own data, and the connections between them, where the same quantities are estimated more than once, are left unmanaged.

The literature on Medicare Advantage risk adjustment shows why the connections carry financial weight. Independent estimates converge on the conclusion that MA risk scores and payments run consistently above what beneficiary health alone would justify. Enrollees in private Medicare plans generate risk scores an estimated 6 to 16 percent higher than comparable fee-for-service beneficiaries (Geruso & Layton, 2017); the mean coding inflation rate across MA contracts has been estimated at 8.4 percent, with about 68.1 percent of enrollees in contracts coding more intensely than Medicare's own adjustment assumes (Curto et al., 2025); and CMS applies a statutory minimum coding-pattern adjustment of 5.90 percent for CY2025 in partial response (Centers for Medicare & Medicaid Services, 2024a). Figure 2.1 places these published estimates alongside MedPAC's 22 percent MA-versus-FFS spending difference. The measures are distinct and not directly comparable, but they point in one direction: the revenue side of every risk pool is systematically sensitive to coding and selection, which means an error or a blind spot in one layer propagates into the others.

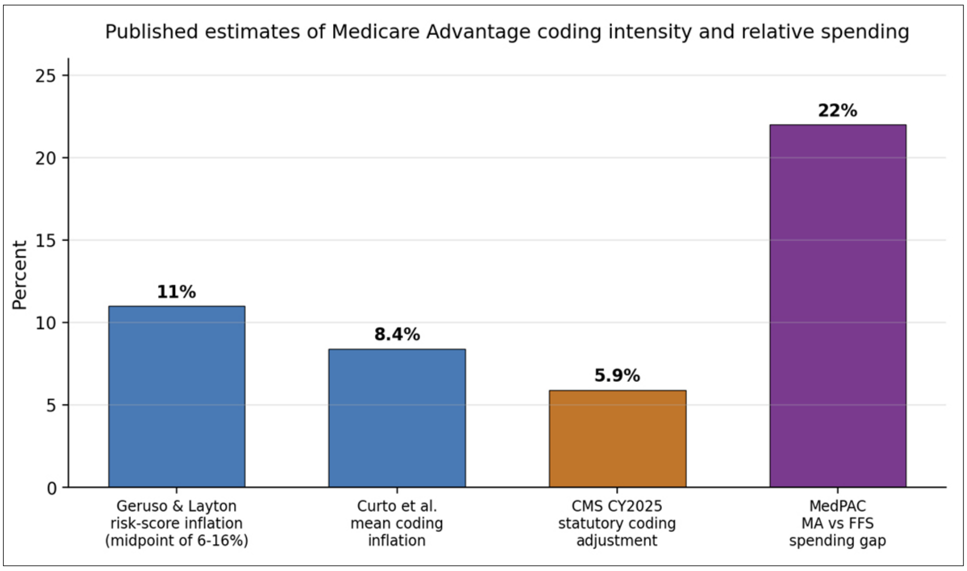

Figure 2.1. Published estimates of Medicare Advantage coding intensity and relative spending. Bars show four distinct, separately published measures (they are not components of one another): risk-score inflation (Geruso & Layton, 2017; midpoint of the 6–16% range), mean contract coding inflation (Curto et al., 2025), the CMS CY2025 statutory coding-pattern adjustment (CMS, 2024a), and the MedPAC estimate of MA-versus-FFS spending (MedPAC, 2024).

The mechanism by which fragmentation becomes cost is concrete. The IBNR reserve and the risk-pool settlement both depend on the same incurred-but-unpaid claims for a period. The Exchange forecast and the reserve both depend on the same enrollment and utilization assumptions. When these are produced by separate teams from separate extracts, the plan holds more than one version of the same quantity: the reserve team's view of unpaid claims may differ from the number used in a risk-pool settlement, and the forecast's enrollment assumption may differ from the one behind the reserve. Reconciling those versions consumes effort, and to the extent they are not reconciled, the plan is making risk-pool, reserve, and pricing decisions on quietly inconsistent numbers. The financial consequence is not a single large error but a persistent tax of small ones: settlements that must be reopened, reserves that swing at quarter-end, and forecasts that miss because they never saw the reserve development that a connected system would have fed back to them.

This is the case for an integrated framework. When three functions answer overlapping regulatory definitions from separate data and hand their outputs across organizational walls, the plan reconciles the same underlying quantities repeatedly and inconsistently, and the inconsistencies become financial risk. The research hypothesis of this monograph follows: a unified operational architecture, one that connects risk-pool estimation, reserve modeling with cross-functional reporting, and state-level forecasting through explicit cross-layer feedback loops, can reduce fragmentation and improve financial outcomes across all three functional areas at once. Table 2.2 summarises the gaps and the layer that addresses each.

Table 2.2. Synthesis of identified gaps and the framework layer that addresses each (author's analysis).

Functional area	Core limitation in practice	Framework response
IPA risk-pool estimation	Aggregated, retrospective, weak contract-claims linkage	Layer 1: per-IPA forward exposure + early warning
IBNR reserve modeling	Static, isolated from operational planning	Layer 2: dynamic reserve feeding forecast and reporting
Exchange forecasting	National grain, regulation as afterthought	Layer 3: state-level projection with embedded drivers
Cross-functional communication	One-directional, untailored	Loop 3: finance as interpretive, bidirectional layer
Connections between all four	Unmanaged; quantities re-estimated inconsistently	Cross-layer feedback loops (Chapter 3)

Chapter outcome: an evidence-based justification for the integrated framework. The gaps are real, documented, and structurally linked; the following chapter specifies the architecture designed to close them.

CHAPTER 3. THE INTEGRATED FINANCIAL ANALYTICS FRAMEWORK: ARCHITECTURE AND METHODOLOGY

The gaps established in Chapter 2 share one cause and therefore admit one kind of solution. If fragmentation is the problem, then the response cannot be a better tool for any single function; it must be an architecture that connects the functions so that the quantities they share are estimated once and flow between them. This chapter specifies that architecture. It presents the conceptual model, the operational blueprint, the data-integration logic that makes a single controlled view possible, and the three cross-layer feedback loops that are the framework's central contribution. The methodology is presented as a design: a structure engineered to be implemented on the data and platforms a mid-sized health plan already possesses, rather than a claim about measured outcomes, which Chapter 7 addresses separately.

Conceptual Model

The framework organizes health-plan financial analytics into three operational layers and the feedback loops that connect them. Layer 1 estimates IPA-level risk-pool exposure. Layer 2 produces the IBNR reserve and translates it into cross-functional reporting. Layer 3 forecasts the Exchange line at the state level. Each layer corresponds to one of the estimation problems of Chapter 1 and closes one of the gaps of Chapter 2. What makes the arrangement a framework rather than a set of three tools is the explicit wiring between them, developed in Section 3.4.

The organizing discipline of the framework can be stated as a single rule: each shared quantity is estimated once and then reused, never re-derived independently by a second function. The unpaid-claims estimate that sits inside the IBNR reserve is the same estimate that enters a risk-pool settlement. The enrollment and utilization assumptions behind the state forecast are the same assumptions that condition the reserve. In fragmented practice these quantities are computed several times from several extracts, and the versions disagree; in the framework they are computed once, in the layer best positioned to compute them, and passed to the others. This rule is what the three-layer structure operationalizes, and it is why the connections between layers carry as much of the design as the layers themselves.

Four design principles govern the architecture. Granularity requires that each layer operate at the unit where financial decisions are actually made: the individual IPA for risk pools, the incurred period for reserves, and the state for forecasting. Regulatory embedding requires that CMS and statutory rules enter the calculations as structural inputs rather than as post-hoc adjustments, so that a rule change propagates through the model the way it propagates through the business. Cross-functional accessibility requires that each layer's output be expressible in the decision language of the function that consumes it, which is the discipline that turns analytics into action. Replicability requires that the whole be buildable on standard tools and documented assumptions, so that the framework is a method others can adopt rather than a proprietary artifact. These principles reflect established directions in payment and reserving practice, where the value of a design lies in aligning measurement with the decision it supports (Robinson, 2001; Cattel et al., 2018).

Architectural Blueprint

Figure 3.1 shows the framework as a flow from data inputs through the three processing layers to their output channels. Inputs divide into internal data, which the plan already holds, and external data, which the regulatory and market environment supplies. The three layers process these inputs into their respective outputs: per-IPA exposure dashboards and settlement reserves from Layer 1, reserve estimates and audience-tailored reports from Layer 2, and multi-year state forecasts and scenarios from Layer 3.

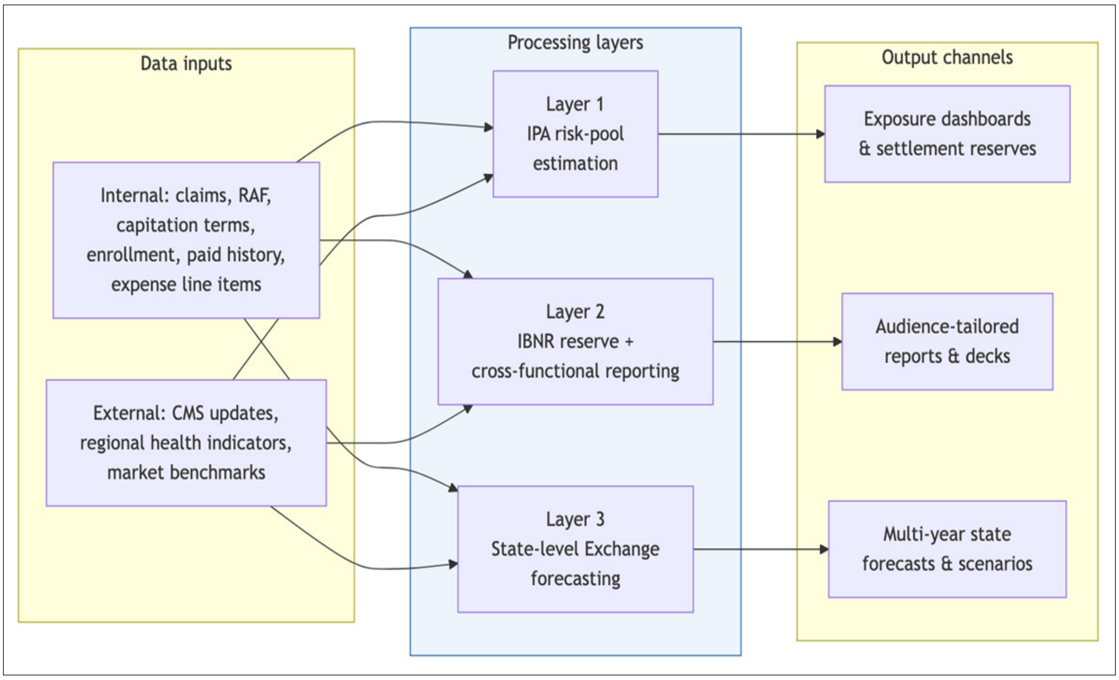

Figure 3.1. Architectural blueprint of the integrated framework: data inputs feed three processing layers, each producing decision-ready output channels (author's construction).

The output channels are defined by the decisions they serve, which is what keeps the framework from producing analytics that no one uses. Layer 1's per-IPA exposure dashboard exists to support three concrete decisions: how much settlement reserve to hold for each pool, which IPAs to flag for contract renegotiation, and where to engage a provider network before a loss position hardens. Layer 2's reporting channel exists to move a governed reserve estimate into the operating decisions of sales, operations, clinical leadership, and the board, each in its own terms. Layer 3's forecast channel exists to support pricing, multi-year planning, and resource allocation at the state level. Defining each output by its decision, rather than by the calculation that produces it, is the cross-functional-accessibility principle made concrete, and it is the standard against which each layer's methodology in the following chapters is judged.

The operational implementation described here uses a spreadsheet-based platform. The choice is deliberate and carries both an advantage and a limitation worth stating plainly. The advantage is reach: a spreadsheet implementation runs on infrastructure every finance team already operates, requires no procurement or engineering programme, and keeps every assumption visible and auditable in the same document that produces the result, which serves the replicability principle directly. The limitation is scale: a spreadsheet platform is bounded in data volume and in concurrent use, and a plan that outgrows those bounds will migrate the same logic to a database and reporting stack. The framework is defined by its logic and its feedback structure, not by the spreadsheet, so the migration changes the substrate without changing the method. Presenting the architecture at the level of logic rather than implementation is what makes it portable across the platforms a plan may use as it grows.

The architecture also assigns responsibility. Layer 1 is owned jointly by finance and provider contracting, because its outputs drive settlement and renegotiation. Layer 2 is owned by finance in concert with the actuarial function, because it produces a reserve that carries an actuarial opinion (Actuarial Standards Board, 2021). Layer 3 is owned by finance with strategic planning, because its outputs feed pricing and resource allocation. The feedback loops are, by design, the shared responsibility that prevents each function from optimising in isolation.

Ownership in this scheme is not the same as isolation. In conventional practice, the fact that the actuarial function owns the reserve and the planning function owns the forecast is precisely what keeps the two apart, because ownership is read as a boundary. The framework treats each layer's owner as accountable for producing a signal and for delivering it to the next layer, so that the boundary becomes an interface rather than a wall. Finance sits at the centre of all three ownership pairings deliberately: it is the function positioned to see risk pools, reserves, and forecasts together, and the framework formalises that position by making finance the interpretive layer through which the loops run. This mirrors the way the payment and reserving literature locates value not in any single calculation but in the alignment of measurement with the decision it must serve (McGuire et al., 2020; Born et al., 2021).

Data Inputs and Integration Logic

A framework that connects three layers is only as reliable as the data model beneath it. The internal inputs are the claims data that record utilization, the member RAF scores that scale revenue, the capitation contract terms that define each risk-sharing arrangement, the enrollment records that determine exposure, the historical paid-claims data that drive completion factors, and the expense line items that populate the forecast. The external inputs are CMS regulatory updates, regional health indicators, and market benchmarks. Figure 3.2 shows how these heterogeneous sources are reconciled.

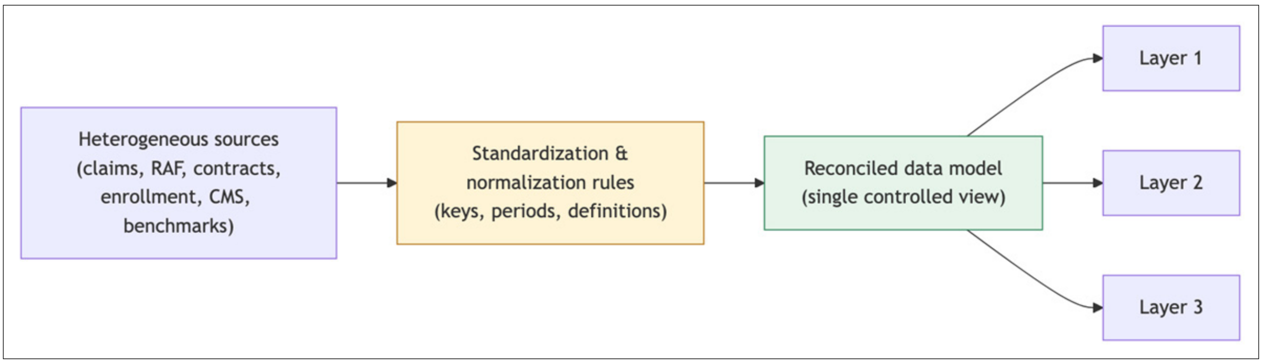

Figure 3.2. Data-integration logic: heterogeneous internal and external sources are standardized and reconciled into a single controlled view before entering the three layers (author's construction).

The integration logic rests on standardization and normalization rules applied before any layer runs. Sources must agree on member and provider keys, on the definition of an incurred period, and on the treatment of adjustments such as offsets and pharmacy rebates, so that a claim counted in the reserve is the same claim counted in the risk pool. In practice these rules resolve a set of recurring mismatches. Member identifiers differ between the enrollment system and the claims system and must be crosswalked before attribution is reliable. Provider identifiers must roll up correctly to the IPA that bears the risk, or a claim will be attributed to the wrong pool. Incurred and paid dates must be distinguished consistently, because the reserve depends on incurred-period cost while the cash view depends on paid-period cost. RAF scores must be aligned to the payment year they govern, since a score drives revenue in a specific period and misalignment distorts every downstream calculation. Each rule is individually mundane; their absence is what produces the quietly inconsistent numbers that Chapter 2 identified as the mechanism of fragmentation. The purpose is not a larger dashboard but a single controlled financial view, in which finance does not forecast one enrollment trend while the reserve assumes another and provider operations tracks a third attribution. This reconciliation is what allows a quantity to be estimated once and reused, and it is the technical precondition for the feedback loops. Where a plan pays capitation and holds IBNR against the same population, the reserve and the settlement must draw on a reconciled view of unpaid claims or the two will disagree by construction (Actuarial Standards Board, 2017; Milliman, n.d.).

Cross-Layer Feedback Loops

The feedback loops are the element of the framework that, as far as the author is aware, is not present in publicly disclosed health-plan methodologies. Each loop connects two layers so that a signal produced in one becomes an input to another, converting three parallel processes into a single system that learns from itself within a planning cycle. Figure 3.3 shows the three loops.

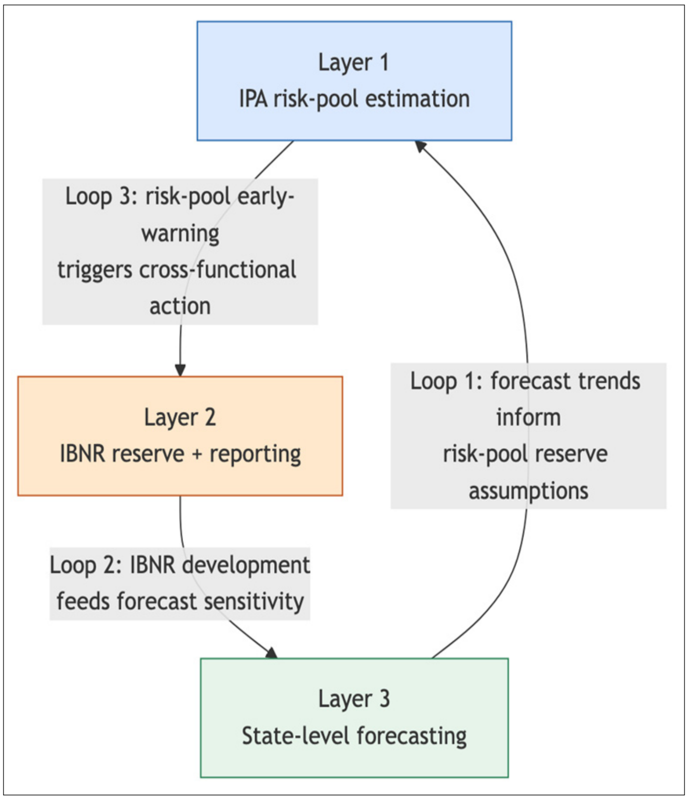

Figure 3.3. The three cross-layer feedback loops that connect the operational layers into a single system (author's construction).

Loop 1 runs from forecasting to risk-pool reserve assumptions. The state-level Exchange forecast produces trend information, on utilization, cost, and enrollment, that is also relevant to the Medicare Advantage risk pools, because the two lines of business draw on overlapping provider networks and shared cost environments. When the forecast detects a cost trend building in a region, that signal informs the reserve sensitivity applied to the risk pools exposed to the same providers, so the risk-pool estimate anticipates the trend rather than discovering it at settlement. Without the loop, the forecast's foresight stays trapped in Layer 3.

Loop 2 runs from IBNR dynamics to forecast sensitivity. The reserve model, by construction, is the first place a change in claim development becomes visible, because it watches the most recent incurred periods as they mature. When development accelerates or slows relative to the historical pattern, that movement is a leading indicator for the forecast, which can then re-run its scenarios against the new development rather than against a stale assumption. The reserve stops being an endpoint and becomes a sensor feeding the forecast.

Loop 3 runs from risk-pool early-warning to cross-functional communication. When Layer 1 flags an IPA trending toward a loss-sharing position, that signal is routed, through the interpretive reporting layer of Layer 2, to the functions that can act: provider contracting to open a renegotiation, operations to investigate the utilization driver, and finance to adjust the settlement reserve. The early-warning signal becomes an operational trigger rather than a line in a retrospective report. This loop is where the communication gap of Chapter 2 is closed, not by better reports alone but by wiring the signal to the decision.

The loops require governance to function, not merely wiring. Each loop has a cadence: Loop 2, watching claim development, operates at the monthly close when new paid data arrives; Loop 1, carrying forecast trends into reserve assumptions, operates on the forecast's rolling cycle; Loop 3, routing early-warning signals, operates on demand whenever a threshold is crossed. Each loop also needs a defined owner for the hand-off, so that a signal produced in one layer is received and acted on in another rather than falling into the gap between two teams. The framework specifies, for each loop, what signal passes, on what cadence, and to whom, so that the connection is a governed process rather than an informal courtesy between colleagues. Without that specification the loops would exist on the diagram and not in the operation, which is the failure mode the design is built to prevent.

Taken together, the loops make the framework more than the sum of its layers. A change anywhere, a CMS rate adjustment, a development shift, an emerging IPA loss, enters the system at one layer and propagates to the others through the loops, so that the plan's risk-pool, reserve, and forecasting views move together instead of drifting apart. Chapter 7 traces exactly this propagation through a worked example.

Differentiation from Existing Practices

The framework's originality is architectural rather than component-level. Each individual component has analogues in practice: payers estimate risk pools, actuaries model IBNR, and planning teams forecast the Exchange. What differs is the integration. Table 3.1 compares the framework with standard industry practice along the four dimensions that Chapter 2 identified as gaps.

Table 3.1. The integrated framework versus standard industry practice (author's analysis, drawing on the gap review of Chapter 2).

Dimension	Standard practice	Integrated framework
Granularity	Aggregate / national	Per-IPA, per-incurred-period, per-state
Regulatory embedding	Post-hoc adjustment	Structural input to each layer
Cross-functional integration	One-directional reporting	Interpretive layer + bidirectional loop
Connection between functions	Unmanaged, re-estimated	Three explicit feedback loops

The distinction matters because it locates the contribution correctly. A reader could object that each layer restates known practice, and the objection would be correct at the component level: risk-pool estimation, IBNR reserving, and Exchange forecasting are all established disciplines with substantial literatures, several of which Chapters 4 through 6 draw on. The contribution is not a new estimator inside any layer. It is the architecture that makes the three layers one system, the data discipline that lets a shared quantity be computed once, and the three feedback loops that turn parallel processes into a self-informing whole. An architecture can be original even when its parts are familiar, in the same way that the value of an integrated design in payment or reserving lies in the alignment it creates rather than in any single new formula (Rice & Smith, 2001; Cattel et al., 2018).

The originality claim is stated with deliberate restraint. Individual payers and vendors operate sophisticated analytics, and the absence of a public methodology combining all three layers with these explicit feedback loops does not prove that no proprietary system does so privately. The defensible claim, and the one this monograph makes, is that no publicly disclosed methodology known to the author connects IPA-level risk-pool estimation, dynamic IBNR reserving with cross-functional reporting, and state-level Exchange forecasting into a single operational architecture with the three cross-layer feedback loops specified here. That architecture, and the discipline of estimating each shared quantity once and letting it flow, is the contribution the following three chapters develop layer by layer.

Chapter outcome: a fully specified architecture, its data foundation, and its feedback structure, prepared for the layer-by-layer methodological elaboration of Chapters 4 through 6.

CHAPTER 4. LAYER 1: IPA-LEVEL RISK POOL ESTIMATION AND FINANCIAL LIABILITY EXPOSURE

Layer 1 is where the framework first touches money at risk. It takes the claims, risk scores, and contract terms introduced in Chapter 1 and produces, for each Independent Physician Association, a single estimate of the plan's financial liability exposure, together with an early-warning signal when that exposure trends toward a loss-sharing position. This chapter specifies the layer as a methodology: its scope, its data inputs, the calculation that turns three data sources into one exposure figure, the early-warning design that makes the figure actionable, the benchmarks that give it context, and the outputs through which it feeds the rest of the framework. The design is presented as engineered logic; illustrative figures are marked as such, since the plan's own operational values are not disclosed here.

Scope and Objectives of Layer 1

The purpose of Layer 1 is to answer, for every IPA in the network and early enough to act, a single question: is this provider group's cost running ahead of its risk-adjusted revenue in a way that will create a loss the plan must share or absorb. The estimate supports four concrete decisions. It sets the settlement reserve held against each pool, so that the plan's books anticipate the eventual settlement rather than being surprised by it. It produces settlement projections that let finance forecast pool payouts before the period closes. It flags IPAs whose economics warrant contract renegotiation, giving contracting a quantified basis for the conversation. And it triggers provider engagement early, while a utilization trend can still be influenced, rather than after it has hardened into a settled loss.

The four decisions are interdependent, which is a further reason to produce them from one estimate rather than from four analyses. The settlement reserve and the settlement projection are the same quantity viewed at two moments, one for the balance sheet and one for the cash forecast, so a difference between them signals an inconsistency rather than two legitimate views. The renegotiation flag and the provider-engagement trigger both depend on the same early-warning read of the pool, differing only in the response they call for: a structural change to the contract in one case, an operational intervention within the current contract in the other. Producing all four from a single per-IPA exposure estimate guarantees that they agree with one another, which a set of separately maintained analyses cannot. This is the estimate-once discipline of Chapter 3 applied to the pool: one figure, four uses.

These objectives distinguish Layer 1 from the retrospective settlement described in Chapter 2. A settlement calculated after a period closes is an accounting record; an exposure estimate maintained continuously is a management instrument. The difference is not the arithmetic, which is similar, but the timing and the connection: Layer 1 is built to run forward and to hand its signals to the reserve model through Loop 1 and to the communication layer through Loop 3. The scale of the stakes is not hypothetical. Because Medicare pays managed care organizations an estimated 22 percent more for their enrollees than fee-for-service Medicare would cost (Medicare Payment Advisory Commission, 2024), the risk-adjusted revenue that flows into each pool is both large and sensitive to coding and selection (Geruso & Layton, 2017; Curto et al., 2025), and predicting the high-cost members who drive pool losses is itself an active analytics problem (Maisog et al., 2019), which makes an accurate, timely per-IPA view of the offsetting cost a material control rather than a reporting nicety.

Data Inputs and Integration Logic

Layer 1 integrates three data sources that are usually held apart. The first is claims data, which records the cost that IPA-attributed members have actually incurred, including the incurred-but-not-reported portion carried from Layer 2. The second is member RAF scores, which scale the capitation revenue attributed to the pool. The third is the provider contract terms, which define how any surplus or deficit is shared. Bringing these together is the integration task, and it is where most of the methodological difficulty lies, because the three sources speak different languages. Claims are organized by member and service date; RAF scores are organized by member and payment year; contracts are organized by provider group and effective period. The reconciliation rules of Chapter 3 apply here in their most demanding form.

The central reconciliation is between the capitation payment schedule and the utilization experience. The plan pays each IPA a risk-adjusted capitation on a monthly schedule; the IPA's members generate claims on their own timeline. Bringing the two to a common member-month basis is the reconciliation that risk adjustment and capitation design both presuppose but rarely enforce operationally (Rice & Smith, 2001); to estimate exposure, the two must be brought to that common basis, so that the revenue attributed to a pool and the cost incurred by the same members over the same period can be compared without timing artefacts. A further complication is that IPA contracts are heterogeneous: risk-sharing percentages, stop-loss thresholds, withhold arrangements, and carve-outs differ from one agreement to the next. The framework handles this by standardizing each contract into a common set of parameters, so that a single calculation engine can process structurally different agreements. The standardized parameter set captures the terms that determine financial outcome: the risk-sharing percentage that divides surplus and deficit, the stop-loss attachment point that caps the provider's downside, the withhold that holds back a portion of capitation pending performance, the services carved in or out of the pool, and the settlement period and run-out window that fix when the pool closes. Representing every contract in these terms means that a heterogeneous network becomes a single table of parameters against which one engine computes exposure, and it means that when a contract changes, only its parameters change, not the calculation. The standardization does not erase the differences; it represents them in a uniform way, which is what allows the plan to compare exposure across a network of contracts that were never written to be comparable (Healthcare Financial Management Association, 1997). This uniform representation is also what makes the renegotiation flag actionable, because it lets contracting see immediately which parameter, the risk share, the attachment point, or the withhold, is driving a given IPA's exposure.

Calculation Logic for per-IPA Financial Liability Exposure

The exposure calculation converts the three reconciled inputs into one figure. Figure 4.1 shows the flow. For a given IPA and period, the plan first assembles the gross risk-pool position: the risk-adjusted capitation revenue attributed to the pool, computed as the members' RAF-adjusted base rate applied over their member-months, set against the claims those members incurred, including the IBNR estimate for care not yet reported. The gross position is then passed through the contract terms, which apply the risk-sharing percentage, any stop-loss caps, and any withholds, to produce the net financial liability exposure the plan carries for that pool.

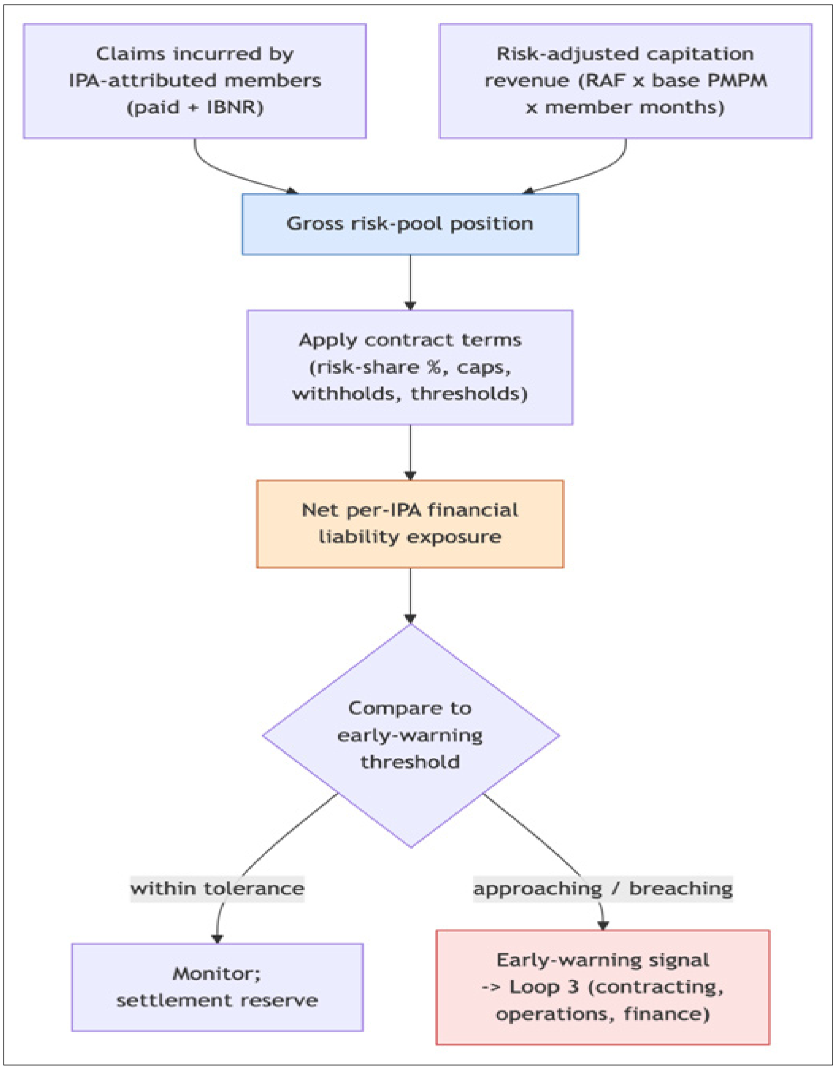

Figure 4.1. Per-IPA financial liability exposure calculation: risk-adjusted capitation revenue and incurred claims (including IBNR) form the gross pool position, which contract terms convert into net exposure and an early-warning comparison (author's construction).

Two features of the calculation deserve emphasis. The first is the treatment of IBNR within the pool. A risk-pool position computed on paid claims alone understates cost, because the most recent months are incomplete; the exposure estimate must therefore incorporate the IBNR reserve for the pool's members, estimated with the completion-factor and Bornhuetter-Ferguson methods whose uncertainty the actuarial literature quantifies (Mack, 1993; Alai et al., 2009), which is the point at which Layer 1 depends on Layer 2. This dependency is not a weakness but the intended design: it is the connection that Loop 1 and Loop 2 formalise, ensuring that the reserve used in the settlement and the reserve carried on the books are the same estimate. The second feature is sensitivity. The exposure figure depends on assumptions, chiefly the completion factors behind the IBNR component and the RAF scores behind the revenue, and a responsible estimate reports not a single number but a range. To illustrate, if a pool shows a net position near breakeven at a central IBNR assumption but tips into a shared loss under a modestly higher completion-factor assumption, that sensitivity is itself the signal: the pool is fragile even if its point estimate looks safe. Reporting the sensitivity, not only the point estimate, is what makes the exposure figure honest.

The arithmetic can be stated as a sequence, with the values left as placeholders because the plan's own figures are not disclosed here. For a pool in a period, revenue is the sum over members of the RAF-adjusted base rate times member-months; cost is paid claims plus the IBNR estimate for the same members and period; the gross position is revenue minus cost. The contract then divides the gross position: a surplus is shared at the risk-sharing percentage up to any cap, and a deficit is absorbed by the plan to the extent the contract assigns it, bounded by the stop-loss attachment. The plan's net exposure is the portion of any deficit it must carry after those terms apply. Each step is elementary; the methodological work is in ensuring that revenue and cost are computed on the same reconciled population and period, so that the subtraction is meaningful. An exposure figure that subtracts a cost measured on one member set from a revenue measured on another is precise and wrong, which is exactly the failure the integration logic of Section 4.2 is built to prevent.

Early-Warning Indicator Design

An exposure figure becomes a management instrument only when it is attached to a threshold and a response. The early-warning design classifies indicators into a hierarchy by how early they speak. Leading indicators move before the loss materialises: an accelerating utilization trend, a rising RAF-to-cost divergence, or an authorization pattern that presages higher claims. Concurrent indicators move as the loss develops: the pool's paid-plus-IBNR position crossing a share of capitation. Lagging indicators confirm what has already happened: a settled deficit. The value of the hierarchy is that it directs attention to the leading indicators, where action is still possible, rather than to the lagging ones, where only accounting remains. Table 4.1 sets out the hierarchy with representative triggers and the function each escalates to.

Table 4.1. Early-warning signal hierarchy for IPA risk pools: indicator class, representative trigger, and escalation path (author's design; thresholds are illustrative and set by each plan).

Signal class	What it measures	Representative trigger	Escalates to
Leading	Emerging utilization / coding-cost divergence	Utilization trend above expected for two consecutive months	Operations, clinical
Concurrent	Current pool position vs capitation	Paid + IBNR exceeds a set share of capitation	Finance (reserve), contracting
Lagging	Realised settlement result	Settled deficit beyond contract share	Contracting (renegotiation)

The distinction among the three classes is not academic; it determines who can still change the outcome. A leading indicator reaches operations and clinical leadership while a utilization trend is still forming, when a care-management intervention or an authorization review can bend the trajectory. A concurrent indicator reaches finance and contracting when the pool position has moved but the period has not yet settled, when a reserve adjustment protects the books and a contracting conversation can begin. A lagging indicator reaches contracting after settlement, when the only remaining lever is the next agreement; that the gap between risk-based payment and realised cost is neither fixed nor self-correcting is exactly why the earlier signals matter (Jacobs & Kronick, 2018). A system that surfaces only lagging indicators, which is what a retrospective settlement process does, structurally confines the plan to the one lever that acts too late. Building the hierarchy so that leading indicators are watched most closely is therefore the design choice that converts monitoring into prevention. Figure 4.2 places the three classes on the timeline of a developing loss, showing how the window for action narrows from prevention, through mitigation, to reaction as a pool moves from a leading signal toward a settled deficit.

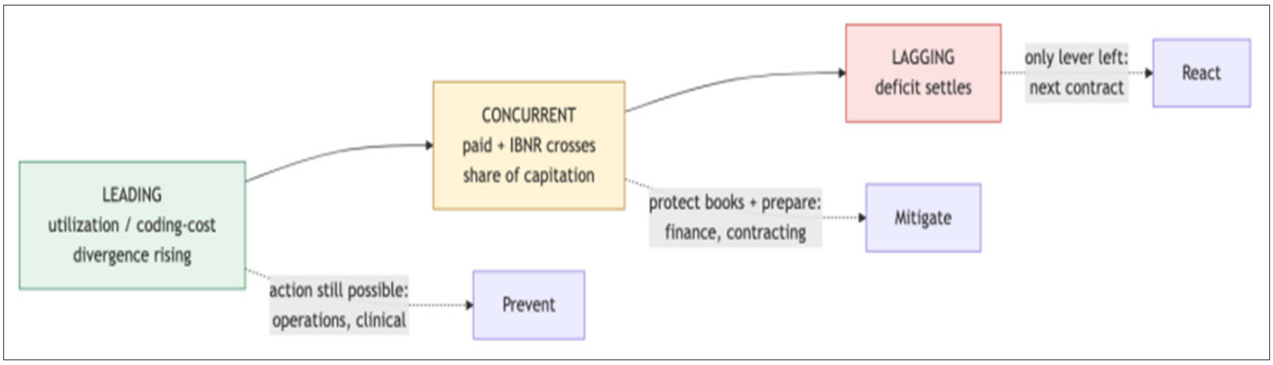

Figure 4.2. The early-warning signal hierarchy on the timeline of a developing loss: leading indicators allow prevention, concurrent indicators allow mitigation, and lagging indicators leave only reaction (author's construction).

The threshold logic is deliberately parameterised rather than fixed, because the right trigger differs by contract and by the plan's risk appetite. What the framework fixes is not the numeric threshold but the discipline: every pool has defined trigger conditions, and every trigger has a named escalation path and owner, so that a breach produces an action rather than a note. This is the operational content of Loop 3. When a leading indicator fires, the signal routes to operations and clinical leadership to investigate the driver; when a concurrent indicator fires, it routes to finance to adjust the settlement reserve and to contracting to prepare; when a lagging indicator fires, it informs the next renegotiation. The escalation protocol turns the exposure estimate into a sequence of decisions with owners, which is what the retrospective practice of Chapter 2 lacked.

Customized Cost and Utilization Benchmarks per IPA

An exposure figure is easier to interpret against a benchmark than in isolation, and Layer 1 constructs a customized benchmark for each IPA rather than applying a single network-wide average. The internal component of the benchmark is built from the plan's own historical experience for comparable populations, adjusted for the risk profile of the specific pool, so that an IPA serving a sicker population is not judged against one serving a healthier one. The external component draws on published comparables to place the pool in the wider market. Regional cost benchmarks such as those in the Milliman Health Cost Guidelines and total-cost-of-care analyses, which document material regional variation in commercial cost, provide the external anchor; one published comparison finds regional differences in total cost of care ranging from 4.6 percent in the Northeast to 10.1 percent in the West (Milliman, 2025). Blending the internal and external components produces a benchmark that reflects both the plan's own experience and the market a given IPA operates in. Table 4.2 lists the inputs to the blended benchmark.

Table 4.2. Inputs to the per-IPA blended cost and utilization benchmark (author's design; external anchors per Milliman, 2025).

Component	Source	Adjustment	Purpose
Internal experience	Plan's own historical claims	Risk-profile normalization	Peer-comparable baseline
External comparable	Published cost benchmarks (e.g., Milliman)	Regional alignment	Market context
Utilization pattern	Plan's authorization and claims data	Seasonality	Detects trend vs level
Recalibration cadence	Periodic review	New experience	Keeps benchmark current

A benchmark is only as good as its maintenance. Because populations and costs move, the benchmark is recalibrated on a defined cycle, validated against realised experience each period, and adjusted when it drifts from the pools it is meant to describe. The validation step is not optional bookkeeping; a benchmark that silently ages will flag the wrong pools, and the credibility of the entire early-warning system depends on the benchmark tracking reality. Validation proceeds by comparing what the benchmark predicted for a pool against what the pool actually realised once the period matured. Where the two agree, the benchmark is confirmed; where they diverge systematically, the divergence is itself information, pointing either to a benchmark that has drifted or to a pool whose economics have genuinely changed. Distinguishing those two cases is the recalibration judgement, and it is why recalibration is a reviewed step rather than an automatic refresh: a benchmark that is re-fit to every deviation would chase noise and lose its ability to flag real change, while one that never moves would slowly describe a market that no longer exists. The framework fixes the cadence and the comparison but leaves the recalibration decision to informed review, which is the same discipline the reserving standards apply to actuarial assumptions.

Outputs and Operational Use

Layer 1 produces three outputs, each defined by a decision. The per-IPA exposure dashboard presents, for each pool, its net exposure, its position against benchmark, and its early-warning status, so that finance and contracting can see the whole network at a glance and drill into any pool. The settlement reserve recommendation translates the exposure estimates into the reserve the plan should hold against pending settlements, which feeds directly into Layer 2 and the plan's financial statements. The contract-renegotiation flags identify, with supporting analytics, the IPAs whose economics warrant a changed agreement, giving contracting a quantified case rather than an impression, in the spirit of aligning provider payment with the outcomes it is meant to produce (Cattel et al., 2018; Robinson & Casalino, 1996), while whether administrative and capture costs belong inside the risk calculation remains a live design question (Douven et al., 2022).

The dashboard is the layer's operational face, and its design follows the cross-functional-accessibility principle. Rather than presenting a table of exposures that each viewer must interpret, it presents each pool with the three elements a decision-maker needs together: the net exposure figure, its position against the pool's own benchmark, and its early-warning status expressed as a state rather than a raw number. A finance director scanning the network sees which pools are safe, which are fragile, and which have breached, without reading a calculation. A contracting lead drilling into a flagged pool sees the parameter driving the exposure. The dashboard is not a report that describes the past; it is an instrument that presents the current state of every pool in the terms each function acts on, which is what allows a network of many IPAs to be managed by exception rather than one settlement at a time.

These outputs are also the layer's connections to the rest of the framework. The settlement reserve recommendation is the substance of Loop 1's contribution in reverse, aligning the pool reserve with the book reserve. The early-warning status is the trigger for Loop 3, carrying the signal into cross-functional action through the reporting layer of Chapter 5. In the author's operational experience within a large U.S. managed care administrator, sharing a structured, contract-based view of pool economics with provider partners changed the character of the conversation, from a dispute over outcomes to a shared reading of the same analytics, though the framework here is presented as a replicable design rather than as a measured result. That design, a per-IPA exposure estimate that is calculated once, benchmarked, monitored against early-warning thresholds, and wired into the reserve and the communication layers, is the documented methodology of Layer 1.

Chapter outcome: a fully specified Layer 1 methodology, from data integration through exposure calculation, early-warning design, benchmarking, and output, connected forward to Layers 2 and 3 through the framework's feedback loops.

CHAPTER 5. LAYER 2: IBNR RESERVE MODELING AND CROSS-FUNCTIONAL FINANCIAL REPORTING

Layer 2 does two things that practice usually keeps apart. It produces the Incurred But Not Reported reserve for the Medicare Advantage line, and it translates that reserve, and the financial results around it, into the specific decisions of the functions that must act on them. Chapter 2 identified both a technical gap, static reserves that miss enrollment volatility, and an organizational one, reserves handed to finance as a number with no path back into operations. Layer 2 is designed to close both. This chapter specifies the dynamic reserve model, its data inputs, its calculation and assumption framework, and the interpretive reporting layer through which the reserve becomes cross-functional action and the third feedback loop is realised. As in Chapter 4, the methodology is presented as design, with illustrative figures marked as such.

Scope and Objectives of Layer 2

The purpose of Layer 2 is to hold an accurate, timely reserve for the Medicare Advantage line and to make that reserve useful beyond the finance and actuarial functions that produce it. It supports three decisions. It sets the reserve that establishes Medicare Advantage reserve adequacy, the figure that carries an actuarial opinion and appears in the plan's statutory statements (Actuarial Standards Board, 2021). It produces variance reporting that explains why results differ from plan, in terms each operating function can act on. And it drives audience-tailored communication, so that the same underlying result reaches sales, operations, clinical leadership, and the board in the form each needs to make a decision.

The two objectives, an accurate reserve and useful communication, are usually treated as belonging to different departments. The framework treats them as one layer because they depend on the same estimate and fail together when they are separated. A reserve that is accurate but uncommunicated changes no behaviour; a report that is timely but built on a stale reserve misleads. Placing the reserve model and the reporting layer in a single design forces the estimate and its interpretation to stay consistent, and it is what allows the reserve to become the sensor that feeds Loop 2 and the signal that travels Loop 3. The reserve is not the end of Layer 2's work but the middle of it.

The reserve also carries weight beyond the Medicare Advantage line, which is a further reason to treat it as a shared asset rather than a departmental output. Reserve and reporting practices are themselves managerial choices with financial effects, not neutral compliance exercises, and the way a plan sets and communicates its reserves shapes its reported profitability and its capital position (Born et al., 2021). A reserve that is understood across the organization is one whose movements can be anticipated and managed; a reserve that lives only in the actuarial function is one the rest of the plan reacts to after the fact. The design intent of Layer 2 is to make the reserve legible enough that operations, sales, and leadership can act on its signals before those signals become settled results.

Dynamic IBNR Model: Data Inputs

The reserve model is dynamic in the specific sense that its inputs are refreshed and its assumptions re-examined as new data arrives, rather than being set once and held. Figure 5.1 shows the flow. Three input streams feed the model. The first is historical paid-claims data, adjusted for offsets: amounts such as capitation already paid, pharmacy rebates, and reinsurance recoveries that reduce the net liability and must be netted against gross incurred cost rather than ignored. The second is the claim development pattern, the record of how claims for each incurred period accumulate as they are paid, from which completion factors are estimated. The third is enrollment, whose trends, growth, churn, and shifts in risk mix, determine both the exposure and the applicability of the historical development pattern.

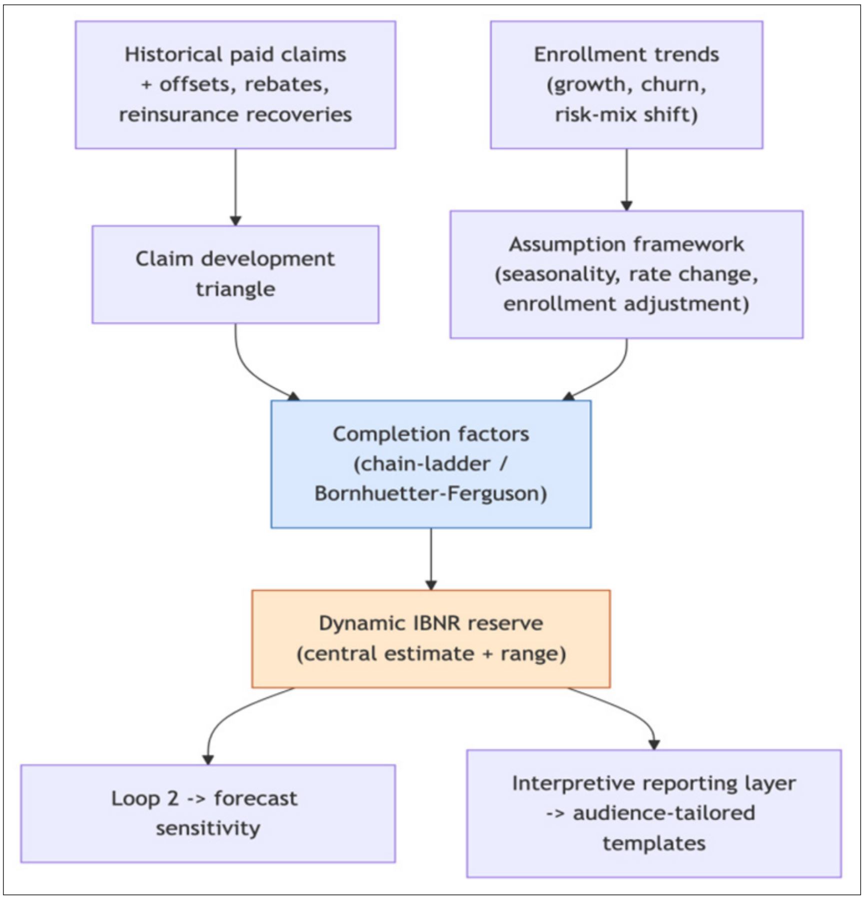

Figure 5.1. Dynamic IBNR model: paid claims with offsets and enrollment trends feed a development triangle and an assumption framework, producing completion factors and a reserve that feeds Loop 2 and the reporting layer (author's construction).

The integration of these three streams is what makes the model dynamic rather than static, and it is the specific advance over the practice reviewed in Chapter 2. A static model treats the completion pattern as fixed and applies it regardless of how enrollment has moved. The dynamic model treats the completion pattern as conditional on enrollment: when membership grows, shrinks, or changes in risk composition, the model adjusts the pattern applied to the affected incurred periods rather than assuming the past applies unchanged. This is the same multi-dimensional integration the author developed in practice, combining historical paid claims, offsets, claim development, and enrollment trends into a single adaptive estimate rather than a fixed table, and it is why the reserve tracks reality through a period of volatility instead of discovering the volatility at quarter-end.

The treatment of offsets deserves particular care, because an offset applied inconsistently distorts the net reserve in a way that is hard to detect. Capitation already paid to a provider group reduces the plan's remaining liability for that group's members and must be recognised, or the reserve double-counts cost the plan has already funded. Pharmacy rebates arrive on a lag and reduce net drug cost, so a reserve that ignores them overstates the liability. Reinsurance recoveries cap the plan's exposure to high-cost claims and must be netted where they apply. Each offset is individually straightforward; the discipline is to apply them on the same reconciled basis as the gross claims, so that the net reserve and the risk-pool exposure that draws on it agree. This is the reserving counterpart to the integration discipline of Chapter 3, and industry cost-benchmarking frameworks presuppose exactly this netting when they express cost on a comparable basis (Milliman, n.d.).

Reserve Calculation Logic and Assumption Framework

The calculation follows the completion-factor logic introduced in Chapter 1, applied with explicit assumptions. For each incurred month, the model estimates the cumulative completion factor, the fraction of the month's ultimate cost already paid, and grosses up the paid-to-date figure to an estimated ultimate. Figure 5.2 shows the characteristic shape of the completion pattern: mature months are nearly complete and stable, while the most recent months are far from complete and carry a large, uncertain gross-up. For those least mature months, where a small and noisy paid amount would otherwise be multiplied by a large factor, the model blends the emerging experience with an a priori expected loss in the manner of the Bornhuetter-Ferguson method, damping the volatility that a pure chain-ladder projection would show (Mack, 1993; Alai et al., 2009).

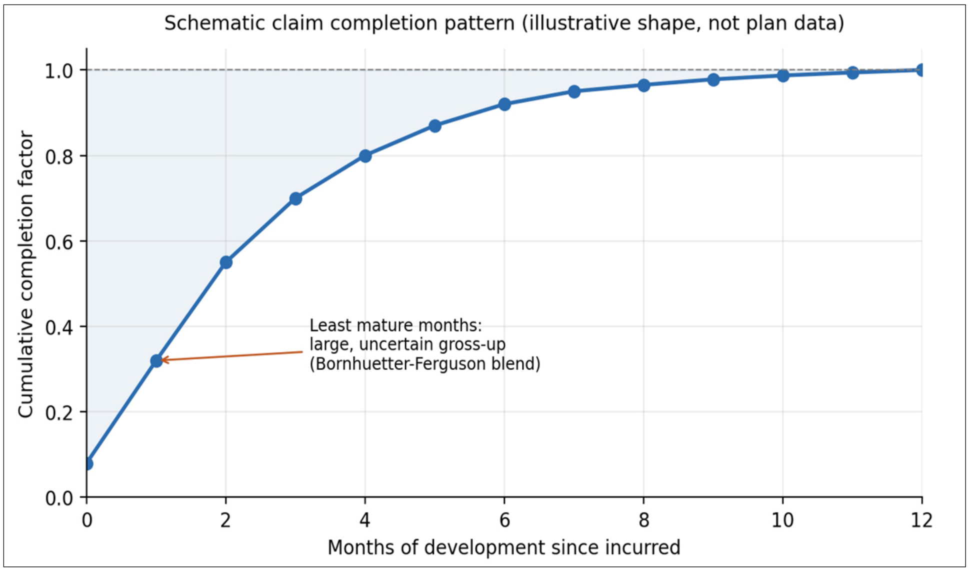

Figure 5.2. Schematic claim completion pattern (illustrative shape, not plan data). The least mature months carry the largest and most uncertain gross-up, which is where the Bornhuetter-Ferguson blend and assumption governance matter most.

Three assumptions carry most of the reserve's sensitivity, and the framework makes each explicit rather than implicit. Seasonality is treated as a systematic adjustment to the completion pattern, so that a month sitting on a utilization peak is not evaluated against an annual-average pattern. Enrollment shifts are treated as a reason to adjust the pattern applied to recent periods, because a surge changes the mix of members whose claims are still developing. Rate changes, whether in benefits or in provider contracts, are treated as structural breaks that can distort the development pattern itself. Because the reserve is an estimate with a distribution rather than a single number, the model reports a central estimate with a range, and modern stochastic and machine-learning reserving methods extend exactly this idea of modelling the whole distribution rather than a point (Sriram & Shi, 2020; Chukhrova & Johannssen, 2021; Schneider & Schwab, 2024). Assumption governance completes the framework: each assumption is documented, its sensitivity is tested, and material changes are reviewed rather than made silently, which is the discipline the actuarial opinion presupposes (Actuarial Standards Board, 2017).

Reporting the reserve as a range rather than a point is not statistical decoration; it changes what the plan can decide. A central estimate tells finance what to book, but the width of the range tells management how much confidence to place in that figure and how much capital to hold against being wrong. A reserve whose range is narrow, because the underlying periods are mature and enrollment is stable, supports a firmer settlement and a tighter forecast. A reserve whose range is wide, because recent months are immature or enrollment has moved, is a signal to hold more reserve, to widen the forecast scenarios, and to watch the affected risk pools more closely. The range is therefore an input to Loop 1 and Loop 2 in its own right: a widening range in the reserve is itself the leading indicator that the forecast and the risk-pool estimates should treat the period as uncertain. A model that reported only a point estimate would discard exactly the information the connected system most needs.

Translation of Actuarial Outputs for Non-Financial Stakeholders

A reserve estimate is expressed in actuarial and financial terms that a sales, operations, or clinical leader cannot act on directly. Layer 2 therefore includes an explicit interpretive layer, owned by finance, that sits between the actuarial output and the operating functions. Its role is to convert a reserve movement or a variance into the operational meaning that the receiving function needs: not that the reserve rose by a certain amount, but that a cost trend in a specific area is developing and what it implies for the decision that function owns.

Positioning finance as this interpretive layer is a deliberate design choice, and it is where Layer 2 differs most from conventional reporting. In common practice the actuarial function produces the reserve and other functions are left to interpret it privately, which is where financial meaning is lost. The framework assigns the interpretation to finance because finance is the function positioned to see the reserve, the risk pools, and the forecast together, and it makes that interpretation a defined step rather than an informal courtesy. This interpretive layer is also the carrier of Loop 3: when Layer 1 flags an IPA early-warning signal, it is finance, in this interpretive role, that translates the signal into the specific action each function should take and routes it accordingly. The translation is not a softening of the numbers for a lay audience; it is the conversion of a financial estimate into an operational decision, which is what turns analytics into behaviour.

The interpretive layer also protects the integrity of the numbers as they travel. When each function interprets a raw reserve privately, the same figure acquires several unofficial meanings, and decisions are made on versions that no one reconciled. A defined interpretive layer replaces those private readings with one authoritative translation per audience, so that sales, operations, clinical leadership, and the board are acting on consistent readings of the same estimate rather than on divergent ones. This is the reporting-side expression of the estimate-once discipline: the quantity is computed once in Layer 2 and translated once for each audience, never re-derived. Regulatory metrics reinforce why the consistency matters, since a ratio such as the medical loss ratio interacts with cost structure in ways that make its interpretation consequential rather than mechanical (Karaca-Mandic et al., 2013), and a plan whose functions read that interpretation differently will act at cross purposes.

Tailored Reporting Templates

The interpretive layer delivers its output through reporting templates built for specific audiences, each carrying the content that audience's decisions require. Table 5.1 sets out the principal templates and the decision each serves.

Table 5.1. Audience-tailored reporting templates in Layer 2: the content each audience receives and the decision it supports (author's design).

Audience	Template content	Decision supported
Sales	MAPD commission framework: financial results linked to commission calculation for agents and agencies	Incentive alignment; accurate, timely commission payment
Operations	Variance report in operational drivers	Performance monitoring; corrective action
Clinical leadership	Cost-trend analysis with clinical context	Utilization and care-management response
Board / executive	Strategic summary with scenario commentary	Strategy, capital, and pricing direction

The sales template is the most developed, because commission is where financial results and operational behaviour meet most directly. The Medicare Advantage Prescription Drug (MAPD) commission framework links the plan's financial outcomes to the commissions paid to external agents and agencies, so that incentive payments reflect performance and contractual terms and are calculated accurately and paid on time. Presenting this framework to the sales organization, and maintaining it in close coordination with sales operations and compliance, keeps the commission structure aligned with financial results and regulatory requirements rather than drifting away from them, which Chapter 2 identified as a specific cost of poor communication. The operations, clinical, and board templates apply the same principle in their own registers: each translates the same underlying reserve and result into the content and the decision proper to its audience. The templates are not four summaries of one report; they are four different reports built from one governed estimate, which is what prevents the private, inconsistent re-interpretation that undifferentiated reporting invites.

The design of the commission template illustrates why tailoring is a financial control and not a presentational nicety. A commission structure that is not kept aligned with financial results rewards behaviour the plan's economics no longer support, and the misalignment is invisible until it appears in margin. By linking commission calculation directly to the financial framework that Layer 2 maintains, and by keeping that linkage current as results move, the plan ensures that what it pays its distribution channel to do stays consistent with what its financial position can sustain. Aligning payment with the outcomes it is meant to produce is the same principle that governs value-based provider payment, applied here to the plan's own sales incentives (Cattel et al., 2018). The operations template applies the principle to internal performance, the clinical template to care management, and the board template to strategy, but the commission case makes the stakes concrete: a reporting layer that fails to tailor is not merely less useful, it is a source of financial drift.

Outputs and Operational Use

Layer 2 produces three outputs. The reserve estimate report carries the central estimate, its range, and its assumption basis into the statutory and management accounts. The variance analysis package explains the difference between actual and expected results in terms that connect to operational drivers. The audience-specific presentation decks and dashboards deliver the tailored templates to each function on its own cadence.

These outputs are the layer's connections to the rest of the framework. The reserve feeds Layer 1, supplying the IBNR component of every risk pool's exposure, and it feeds Layer 3 through Loop 2, because a change in claim development is a leading indicator for the forecast. The audience-tailored reporting is the vehicle of Loop 3, carrying early-warning signals from Layer 1 into cross-functional action. In the author's operational experience within a large U.S. managed care administrator, presenting reserve findings to senior leadership in this tailored form, rather than as an actuarial deliverable, changed reserve adequacy and risk exposure from a technical report into an input to strategic planning, though the framework here is presented as a replicable design rather than as a measured outcome. That design, a dynamic reserve that is estimated once with governed assumptions and then translated, through an interpretive finance layer, into the specific decisions of every function, is the documented methodology of Layer 2.

Chapter outcome: a fully specified Layer 2 methodology, from the dynamic reserve model through the interpretive reporting layer, connected to Layers 1 and 3 through the reserve dependency and the second and third feedback loops.

CHAPTER 6. LAYER 3: STATE-LEVEL FORECASTING AND MANAGEMENT REPORTING FOR THE EXCHANGE MARKET

Layer 3 forecasts the ACA Exchange line at the level where its economics actually operate: the individual state. Chapter 2 showed that conventional Exchange forecasting works at a national grain, averaging away the state-level enrollment, competition, and regulation that determine a plan's result, and treating regulatory change as an adjustment layered onto a baseline rather than as a structural input. Layer 3 is designed to correct both. This chapter specifies the methodology: state-level member-month projection, line-item expense decomposition, scenario-based modeling of the key ratios, and the direct embedding of regulatory drivers, together with the outputs through which the forecast supports pricing and planning and feeds the risk-pool layer through the first feedback loop.

Scope and Objectives of Layer 3

The purpose of Layer 3 is to produce a forecast granular enough to price and plan by state and flexible enough to be re-run as conditions change. It supports four decisions. It informs pricing strategy, since a premium is set by state and must reflect that state's cost and competitive structure. It supports multi-year planning, because Exchange participation is a multi-year commitment whose economics evolve. It guides resource allocation across states, directing capital and attention to where they earn a return. And it supports regulatory compliance monitoring, keeping the medical loss ratio and related constraints in view before they bind. The market these decisions govern is large and moving: the 2024 open enrollment period drew over 21.4 million plan selections across the Marketplaces, a 31 percent increase over the prior year, with new-consumer enrollment up 41 percent (Centers for Medicare & Medicaid Services, 2024b). Figure 6.1 shows the composition and the growth.

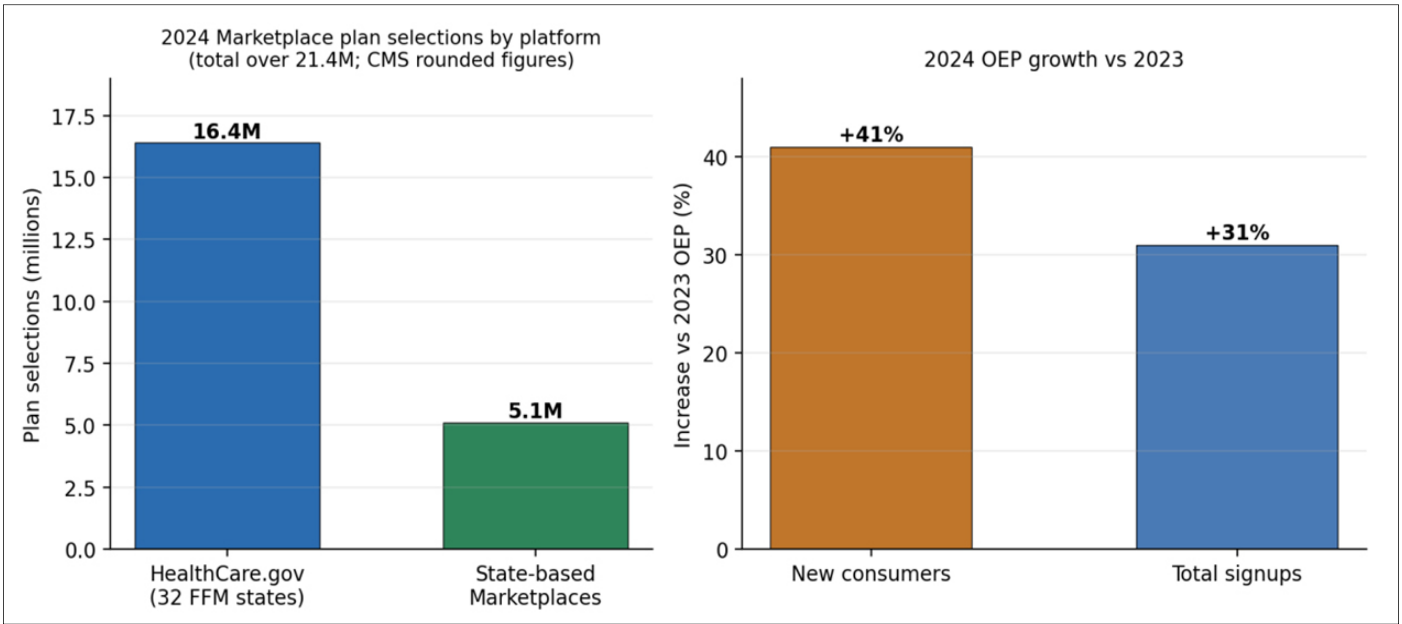

Figure 6.1. 2024 Marketplace plan selections by platform and year-over-year growth. Of over 21.4 million total plan selections, 16.4 million were through HealthCare.gov and 5.1 million through the 19 state-based Marketplaces (CMS rounded figures, each reported to one decimal, so the components round to slightly above the stated total); total signups rose 31 percent and new-consumer enrollment 41 percent over the 2023 period (Centers for Medicare & Medicaid Services, 2024b).

The four decisions are connected by their shared dependence on the state-level forecast, which is why the framework produces them from one engine rather than from separate analyses. Pricing sets the premium a plan will live with for a coverage year; planning commits capital across a multi-year horizon; allocation directs where that capital and the organization's attention go; compliance monitoring keeps the whole within regulatory bounds. A pricing decision made on a national cost assumption, a planning decision made on an aggregate growth rate, and an allocation decision made on last year's footprint will each be individually defensible and collectively inconsistent, because none of them sees the state-level reality the others depend on. Producing all four from a common state-level forecast is the estimate-once discipline of Chapter 3 applied to the Exchange line.

The scale and the growth rate together make the case for the layer. A market that adds millions of enrollees in a single year, unevenly across states, cannot be planned from a national average, because the average conceals precisely the state-level movements that determine where a plan gains or loses. The individual market has been reshaped repeatedly across its first decade, growing from 10.6 million to 17.4 million enrollees as policy and subsidy conditions changed

(Handel & Kolstad, 2022), and that volatility falls unevenly across states. The objective of Layer 3 is to give each pricing, planning, and allocation decision the state-specific forecast it requires, and to keep that forecast current enough to act on.

State-Level Member-Month Projections

The foundation of the forecast is a projection of member-months by state, because member-months are the unit in which both revenue and cost accrue. The projection begins from current enrollment and applies the forces that move it: growth from new enrollment, churn as members enter and leave, and shifts in the risk composition of the enrolled population. Regional health indicators enter here as structural inputs, conditioning the expected cost of the projected membership rather than being applied as an afterthought. The projection also treats the enrollment calendar explicitly. Open enrollment produces the annual cohort against which premiums were priced, while Special Enrollment Periods (SEP) and mid-year churn add members on a different schedule and, often, with a different risk profile.

Distinguishing these enrollment channels matters financially because they do not carry the same cost. Adverse selection operates within the individual market, so members entering off-cycle can be costlier than the open-enrollment cohort a premium was set against (Fung et al., 2018). A projection that treated all enrollment as one undifferentiated stream would misstate both the volume and the cost of membership, and would do so differently in each state. The mechanics of the projection make the channels visible rather than blending them. Open-enrollment membership is projected from the priced cohort and its expected persistence, since a share of each year's enrollees does not renew. Churn is projected as a rate that varies by state and by subsidy status, because the propensity to enter and leave the market is not uniform. Special Enrollment Period additions are projected separately, with their own volume and their own cost assumption, so that off-cycle enrollment does not inherit the risk profile of the open-enrollment cohort. Risk-mix shift is tracked as the composition of the membership changes, because a pool that grows through healthier or sicker entrants moves its own expected cost. Each of these forces operates at a different rate in each state, which is why a single national churn or growth assumption misstates the base everywhere.

By projecting member-months at the state level and by channel, Layer 3 produces the exposure base that the expense decomposition and the scenario engine then act on. The projection is the point at which the forecast becomes state-specific, and every downstream calculation inherits that granularity.

Line-Item Expense Decomposition

A forecast that projects only a total cost cannot explain a variance or locate a cost driver, so Layer 3 decomposes expense into its line items and projects each on its own logic. Medical costs are projected from utilization and unit-cost trends, the largest and most variable component. Pharmacy costs are projected separately, because their drivers, drug mix and rebate timing, differ from medical drivers, and because rebates arrive on a lag that a combined projection would blur. Administrative expenses are split into their fixed and variable components, since the two respond differently to enrollment change: fixed costs dilute as membership grows, while variable costs scale with it. Operational expenses are allocated on a defined basis rather than spread uniformly, so that the states and products that generate them carry them.

The value of the decomposition is that it makes cost legible and therefore manageable. When a state's result diverges from plan, a decomposed forecast shows whether the cause is medical utilization, a pharmacy trend, an administrative overrun, or an allocation artefact, and it points the response to the right function. Regional cost structure enters the decomposition as well, since cost varies materially by region; one published total-cost-of-care comparison finds differences ranging from 4.6 percent in the Northeast to 10.1 percent in the West (Milliman, 2025), a spread large enough that a single national cost assumption would misprice most states. The pharmacy line illustrates why the components cannot be forecast as one. Pharmacy cost is driven by drug mix, by the shift toward high-cost specialty agents, and by rebates that arrive well after the prescription is filled. A combined medical-and-pharmacy projection would blur the rebate lag into the medical trend and misstate both the gross cost and its timing. Forecasting pharmacy on its own logic, gross of rebate with the rebate recognised on its actual schedule, keeps the net cost and its timing correct, which matters because the MLR that regulators test is computed on a net basis. The administrative line makes the same point from the other direction: because fixed administrative cost dilutes as membership grows while variable cost scales with it, a plan expanding in one state and contracting in another will see its administrative ratio move in opposite directions in the two states, a movement that only a fixed-versus-variable decomposition can forecast. Decomposing expense by line item and by state is what lets the forecast be both accurate in aggregate and correct where it matters. The same comparison shows that cost also differs by service category rather than uniformly: Figure 6.2 draws on it to show inpatient cost running above the benchmark while outpatient, professional, and other medical categories run below it (Milliman, 2025). This is the empirical case for decomposition in one picture: a plan that projected a single blended trend would mask offsetting movements of this size and lose the ability to see which category is driving a variance. Decomposing expense by line item and by state is what lets the forecast be both accurate in aggregate and correct where it matters.

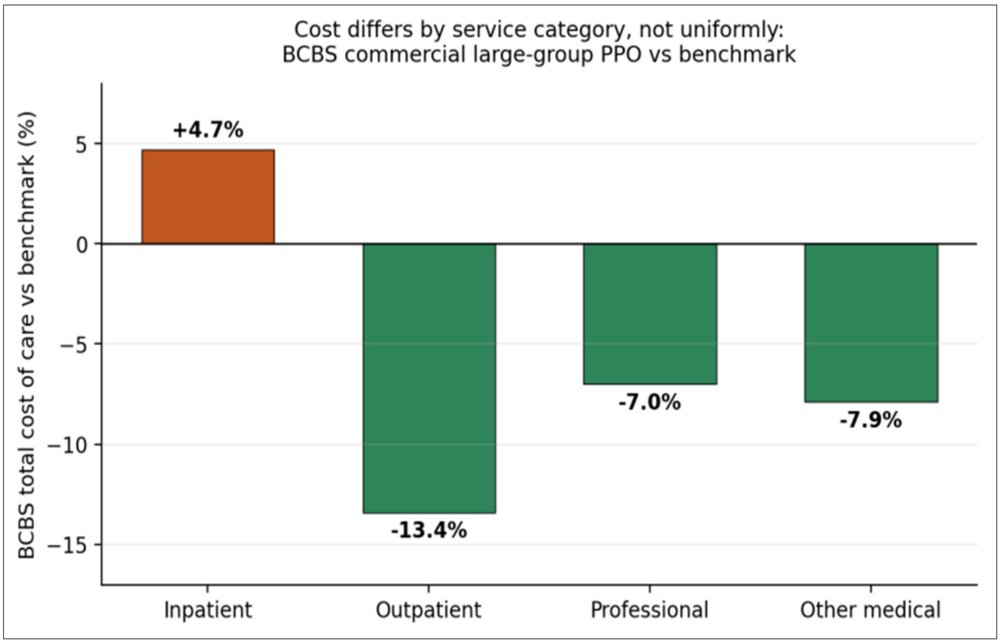

Figure 6.2. Cost differs by service category, not uniformly: total cost of care versus benchmark by service category for commercial large-group PPO networks, showing inpatient above and outpatient, professional, and other medical below the benchmark (Milliman, 2025). The pattern is the empirical case for line-item decomposition.

Scenario-Based Modeling of Key Metrics

Because the Exchange environment is uncertain, Layer 3 models each state under a set of scenarios rather than a single projection. The scenario architecture spans a baseline, an adverse case, a favorable case, and a stress case, so that leadership sees not one number but a range of financial outcomes and the conditions that produce each. Within each scenario the model projects the three ratios that govern Exchange economics: the Medical Loss Ratio (MLR), the Administrative Expense Ratio (AER), and the Per Member Per Month (PMPM) cost. Table 6.1 sets out the scenario matrix.

Table 6.1. Scenario matrix for state-level Exchange forecasting: each key ratio is projected under four scenarios (author's design; the framework fixes the structure, each plan sets the scenario assumptions).

Metric	Baseline	Adverse	Favorable	Stress
MLR	Expected cost/premium	Higher utilization	Lower trend	Regulatory + cost shock
AER	Planned admin load	Enrollment shortfall	Scale efficiency	Fixed-cost dilution loss
PMPM	Central cost trend	Elevated medical/pharmacy	Favorable mix	Compound adverse trend

The scenarios are not one-off exercises but a rolling capability. Multi-year planning horizons require the model to project several years forward, and a rolling forecast cycle re-runs the scenarios as new data arrives, so the forecast is a living instrument rather than an annual artifact. The MLR scenario is watched with particular care, because the ratio is a binding regulatory constraint and its interpretation interacts with plan size and cost structure in ways that make it consequential rather than mechanical (Karaca-Mandic et al., 2013). The construction of the scenarios is disciplined rather than arbitrary. Each scenario is defined by a coherent set of assumptions about the drivers, utilization, trend, enrollment, and regulation, so that the adverse case is not simply the baseline scaled down but a specific combination of conditions the plan can name and monitor. This lets the scenario range function as an early-warning instrument in its own right: when realised conditions begin to track the adverse assumptions, the plan knows which scenario it is entering before the results confirm it. The rolling cycle keeps the scenarios honest, because re-running them against new data each period exposes assumptions that reality has overtaken. A static annual forecast, by contrast, cannot be re-run cheaply and so tends to be defended past its useful life. Modeling MLR, AER, and PMPM by state and by scenario is what lets a plan see, before a coverage year begins, which states are robust across scenarios and which are fragile under the adverse or stress case. Figure 6.3 illustrates the scenario fan: a single metric projected forward along four paths, so that leadership reads not one number but the spread between the favorable and stress cases and the conditions that separate them.

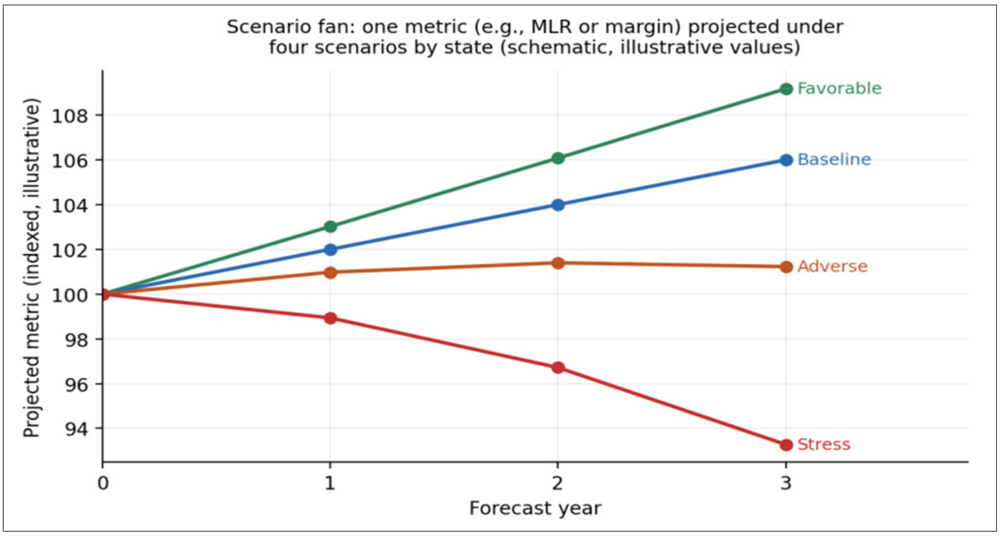

Figure 6.3. Scenario fan (schematic, illustrative values): one metric, such as MLR or margin, projected by state under baseline, adverse, favorable, and stress scenarios; the width between the paths is the financial risk the forecast makes visible (author's construction).

Embedding Regulatory and External Drivers into Forecasting Logic

The defining feature of Layer 3, and its clearest departure from the practice of Chapter 2, is that regulatory and external drivers enter the forecast as structural inputs rather than ex-post adjustments. Figure 6.2 shows the engine: member-month projections, expense decomposition, and regulatory and external drivers all feed a scenario engine that produces the ratios by state and scenario.

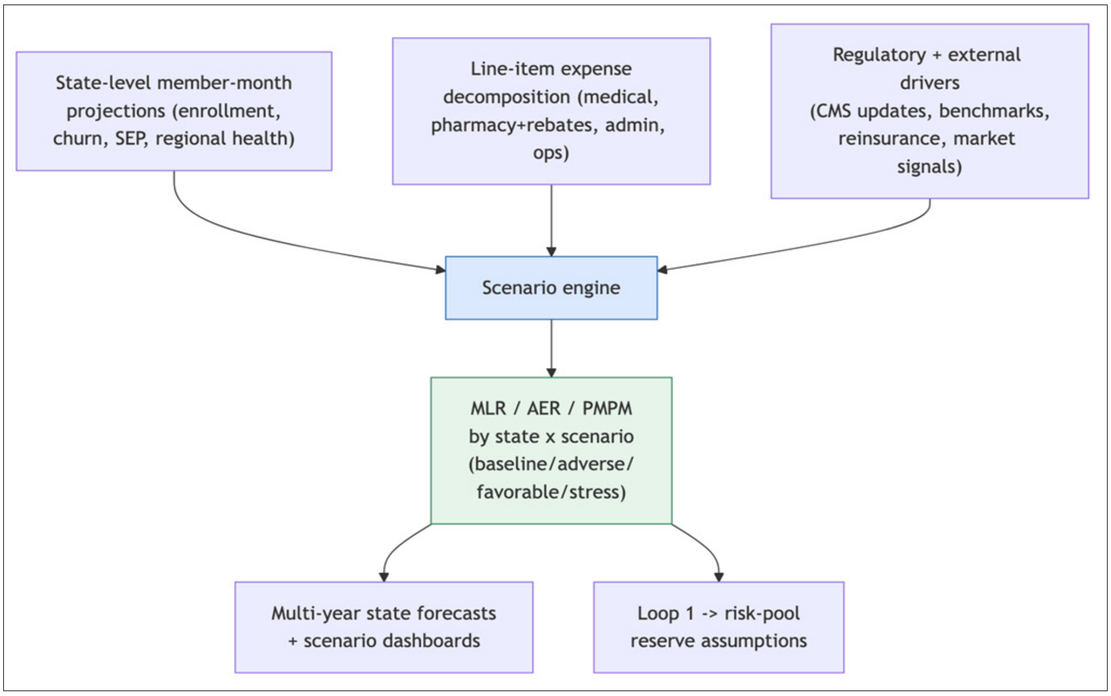

Figure 6.2. State-level forecasting engine: member-month projections, line-item expense decomposition, and embedded regulatory and external drivers feed a scenario engine producing MLR, AER, and PMPM by state and scenario, with output feeding Loop 1 (author's construction).

Embedding means that a regulatory change is expressed in the model's structure, so that it propagates the way it propagates through the business. Reinsurance is the clearest example. Where a state operates a reinsurance program, it compresses premiums measurably, and the effect varies by metal level: in the first implementation year (2018), reinsurance was associated with premium reductions of roughly 10 to 13 percent for bronze plans, 5 to 19 percent for silver, and 11 to 17 percent for gold, relative to 2017 (Oyeka & Wehby, 2023). Figure 6.3 shows the first-year point estimates by metal level for a benchmark enrollee. A forecast that treated reinsurance as a post-hoc adjustment would miss how it changes the premium base and therefore the MLR; a forecast that embeds it projects the effect directly, and by metal level rather than as a single average.

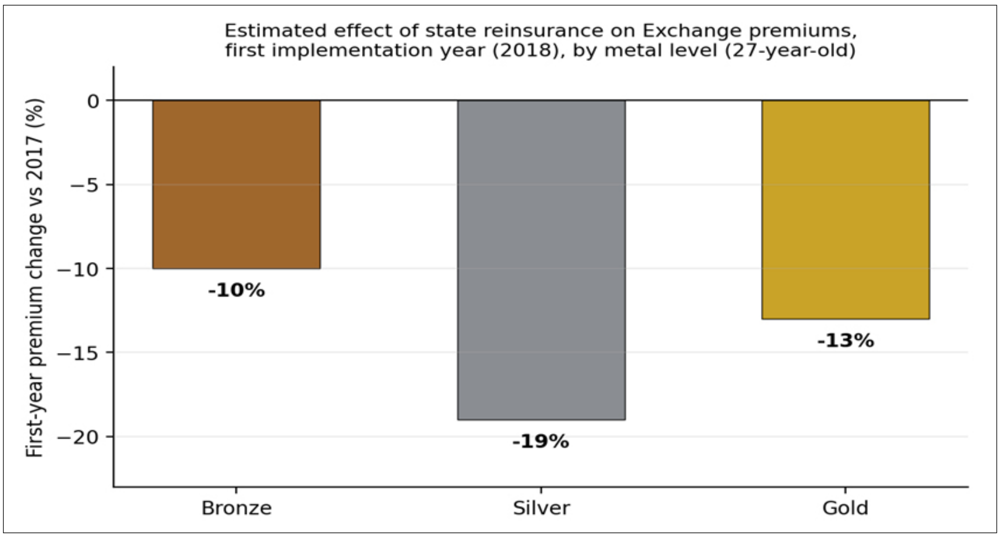

Figure 6.3. Estimated effect of state reinsurance programs on Exchange premiums in the first implementation year (2018), by metal level, for a 27-year-old enrollee relative to 2017 (Oyeka & Wehby, 2023): bronze 10%, silver 19%, gold 13% (point estimates; first-year ranges across specifications are bronze 10–13%, silver 5–19%, gold 11–17%). Reinsurance is embedded as a structural input to the forecast rather than an ex-post adjustment.

Competition and market structure are embedded on the same principle. Insurer entry and exit move premiums materially; the second-lowest-price silver premium that anchors subsidies would have been an estimated 5.4 percent lower in one year had a single large insurer participated fully (Dafny et al., 2015), and market segmentation between the individual and small-group markets shifts both cost and premium (Dickstein et al., 2021). Contract-design and selection studies show how off-cycle enrollment and plan choice move the risk pool a forecast must anticipate (Geruso et al., 2019; Cabral et al., 2018). The rate-setting and risk-transfer rules are embedded on the same principle. CMS rate rules and the annual benefit and payment parameters set the terms under which premiums are filed and subsidies calculated, and the medical-loss-ratio rule imposes the rebate threshold that caps retained margin (Centers for Medicare & Medicaid Services, 2026). Risk adjustment transfers funds between insurers in a state market according to the relative risk they enroll, so a plan's net revenue depends on its competitors' risk profiles as well as its own, a structure the payment literature describes as regulated competition (Layton et al., 2017). A forecast that embeds these rules projects a premium net of expected transfer and tests it against the rebate threshold directly; a forecast that adjusts for them afterward treats as an exception what is in fact the market's operating system. Regional health indicators and market signals enter as structural conditioners of cost and enrollment. Treating these drivers as inputs rather than adjustments is what makes the forecast responsive to the environment that actually determines its accuracy.

Outputs and Operational Use

Layer 3 produces three outputs. The multi-year state-level forecast gives pricing and planning a state-specific, multi-year view. The scenario dashboards present the baseline-to-stress range for each state, so leadership can see robustness and fragility at a glance. The variance and trend reports connect realised results back to the forecast, closing the loop between projection and outcome and feeding the decomposition that explains any divergence.

These outputs are also the layer's connections to the framework. Through Loop 1, the forecast's state-level trend information flows into the risk-pool reserve assumptions of Layer 1, so that a cost trend detected in the Exchange line informs the Medicare Advantage pools exposed to the same providers and regions. Through Loop 2, the forecast receives the reserve's development signals and re-runs its scenarios against them. In the author's operational experience within a large U.S. managed care administrator, building the forecast at state level with embedded drivers and scenario ranges gave executive planning a view of performance by region and product that a national forecast could not, though the framework here is presented as a replicable design rather than a measured result. That design, a state-level, decomposed, scenario-based forecast with regulatory drivers embedded and outputs wired into the risk-pool and reserve layers, is the documented methodology of Layer 3.

Chapter outcome: a fully specified Layer 3 methodology, from state-level member-month projection through expense decomposition, scenario modeling, and embedded drivers, connected to Layers 1 and 2 through the first and second feedback loops. With all three layers now specified, Chapter 7 turns to implementation, validation, and managerial implications.

CHAPTER 7. IMPLEMENTATION, VALIDATION, AND MANAGERIAL IMPLICATIONS

The three layers and their feedback loops describe a system. This chapter shows the system in operation, sets out how a plan would build it, and states honestly what it can and cannot claim. It opens with a worked example that traces a single regulatory event through all three layers, then presents an implementation roadmap, a qualitative account of validation within the author's organization, the risks and limitations of the approach, and the managerial recommendations and future directions that follow. Throughout, the framework is presented as a replicable design; where the worked example needs figures, they are marked as illustrative, and the validation is described in qualitative terms rather than as measured performance.

Integration in Operation: A Worked Example

The clearest way to see the framework work is to follow one event through it. Consider a CMS rate adjustment of the kind the program produces every year: for CY2025, payments to Medicare Advantage plans were set to rise by an average of 3.70 percent, and CMS applied a statutory minimum coding-pattern adjustment of 5.90 percent (Centers for Medicare & Medicaid Services, 2024a). In a fragmented organization, this single announcement lands in three places at three times, and the three responses are reconciled late if at all. In the integrated framework it enters once and propagates through the loops. Figure 7.1 traces the path.

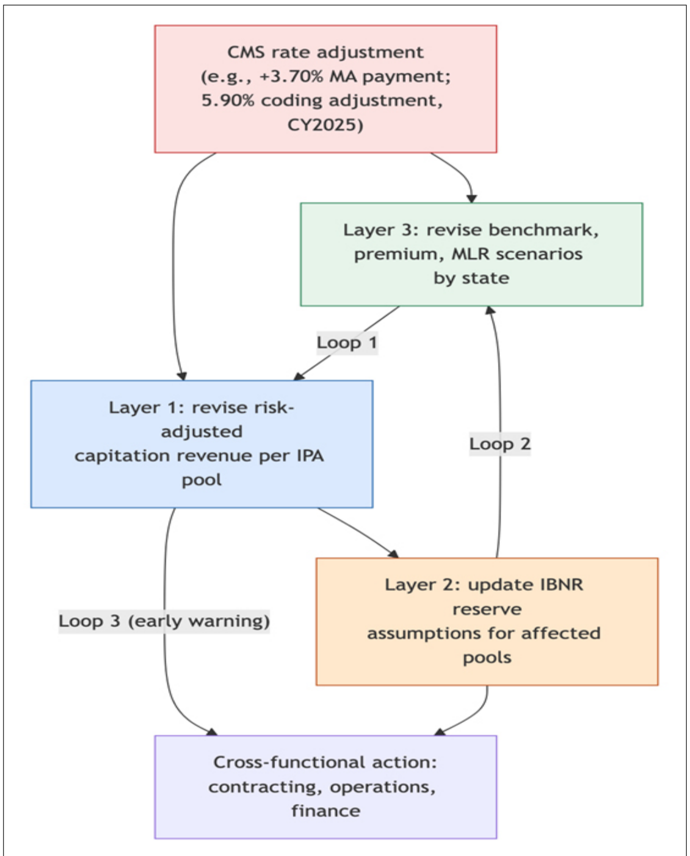

Figure 7.1. Propagation of a CMS rate adjustment through the three layers and the feedback loops: the event revises revenue and forecast, updates reserve assumptions through Loop 1, feeds forecast sensitivity through Loop 2, and triggers cross-functional action through Loop 3 (author's construction).

The propagation runs as follows. Layer 3 revises the state-level forecast first, because the rate change moves the benchmark, the premium, and the MLR scenarios by state. Layer 1 revises the risk-adjusted capitation revenue attributed to each IPA pool, since the coding adjustment and the payment change alter the revenue side of every pool. Through Loop 1, the forecast's revised cost and enrollment trends inform the reserve assumptions applied to the affected pools, so Layer 2 updates the IBNR reserve for those pools on a consistent basis rather than on a stale one. Through Loop 2, the reserve's revised development feeds back into the forecast's scenarios. Through Loop 3, any pool that the revised revenue pushes toward a loss-sharing threshold generates an early-warning signal that routes, via the interpretive reporting layer, to contracting, operations, and finance. If, to illustrate, the coding adjustment reduced a given pool's net position enough to breach its concurrent-indicator threshold, the framework would surface that pool for action within the same cycle in which the rate change was announced, not at the following settlement. The operational protocol that makes this work is a defined cross-layer reconciliation: when an event of this kind occurs, each layer owner updates their layer, the loop hand-offs execute on their defined cadence, and finance confirms that the risk-pool, reserve, and forecast views have moved together. Each functional team has a role in the sequence, and the roles are specified in advance so that the response is a coordinated update rather than three separate reactions.

The contrast with fragmented practice is the point of the example. In a siloed organization the same announcement would reach the actuarial team, the planning team, and the contracting team on their own schedules; each would update its own model against its own data; and the differences between those updates would surface only when a settlement or a close forced a reconciliation, by which time decisions had already been made on inconsistent numbers. The scale of the revenue at stake makes the difference material, since Medicare already pays managed care organizations an estimated 22 percent more per enrollee than fee-for-service (Medicare Payment Advisory Commission, 2024), so a coding or payment change moves a large base. The integrated framework does not make the arithmetic of the response easier; it makes the response coherent, because the single reconciled view and the defined loop hand-offs guarantee that the risk-pool, reserve, and forecast updates are versions of one adjustment rather than three. Coherence achieved within the cycle, rather than reconciliation forced after it, is the operational value the worked example demonstrates.

Implementation Roadmap

A plan does not build all three layers at once. The roadmap sequences the work so that each stage produces usable output and de-risks the next. Table 7.1 sets out the phases, their prerequisites, their owners, and their principal risks.

Table 7.1. Implementation roadmap for the integrated framework (author's design; sequencing may be adapted to a plan's starting point).

Phase	Prerequisite	Primary owner	Principal risk
Pilot (one layer, subset)	Reconciled data for a sample of pools/states	Finance + one partner function	Data quality gaps surface
Calibration	Validated assumptions and benchmarks	Finance + actuarial	Assumptions overfit to sample
Full deployment	Governance and loop hand-offs defined	Finance leadership	Organizational adoption
Steady-state operation	Recalibration cadence in place	Layer owners	Drift if maintenance lapses

The prerequisites are as important as the phases. The framework depends on a reconciled data model, so the first investment is in the data integration described in Chapter 3, not in analytics; a plan that begins by building dashboards on unreconciled data will produce exactly the inconsistent numbers the framework exists to eliminate. Organizational change management is the second prerequisite, because the framework asks functions that have operated independently to accept shared signals and defined hand-offs. Roles, responsibilities, training, and governance must be established, since the feedback loops are processes with owners rather than features of a tool. The pilot is deliberately narrow so that data-quality gaps surface on a manageable scale, and calibration is deliberately separate so that assumptions are validated before the framework is trusted at full scale.

Sequencing the layers within the rollout follows the same logic. A plan can begin with whichever single layer addresses its most acute pain, a fragile risk-pool book argues for starting with Layer 1, a volatile reserve for Layer 2, an unpredictable Exchange line for Layer 3, and can operate that layer usefully on its own while the reconciled data model matures. The feedback loops are added as the second layer comes online, because a loop requires two layers to connect. This staged path means the framework delivers value from the first layer rather than requiring the whole system before any benefit appears, which is what makes the investment defensible to a chief financial officer weighing it against competing priorities. The risk to manage in this staging is that a layer built well in isolation can harden into a new silo if the loops are treated as optional; the roadmap therefore treats the loop hand-offs as deliverables of the deployment phase, not as enhancements to be added later.

Validation and Observed Outcomes

The framework has been implemented and used within the author's organization, a large U.S. managed care administrator operating across Medicare Advantage and Exchange lines of business. The account here is qualitative by design: the plan's operational performance figures are not disclosed, and the value of the framework is described in terms of the changes it produced in how financial decisions were made rather than as a measured improvement.

Three qualitative outcomes are reported. First, reserve transparency improved, because presenting the reserve through the interpretive layer, rather than as an actuarial deliverable, made its movements legible to leadership and connected them to the decisions they bore on. Second, IPA-level risks were detected earlier, because the early-warning design surfaced pools trending toward loss-sharing positions while action was still possible rather than at settlement. Third, cross-functional alignment strengthened, because the feedback loops gave finance, actuarial, contracting, and operations a shared reading of the same quantities instead of separate ones. The observed effect on the plan's financial decision-making cycles was a shift from reacting to settled results toward anticipating them, which is the outcome the framework is engineered to produce. The mechanism behind each observation is the framework's structure rather than any single tool. Reserve transparency followed from placing the reserve in the interpretive layer, so its movements arrived with their operational meaning attached. Earlier risk detection followed from the leading-indicator design of Layer 1, which watches the signals that move before a loss settles. Stronger alignment followed from the loops, which replaced three private readings of the same quantities with one shared reading. None of these outcomes required a new estimator inside a layer; each followed from the connections between layers, which is consistent with the framework's central claim that the contribution is architectural. Reporting these as directional, design-consistent observations rather than as measured effects is the honest position given that the validation environment did not support a controlled comparison, and it is the position the framework's originality does not depend on, since the originality rests on the architecture itself rather than on a performance claim (Born et al., 2021). These are design-consistent observations, not controlled measurements, and they are offered as such.

Risk Management and Limitations

The framework carries operational risks and methodological limitations, and stating them plainly is part of presenting it honestly. The principal operational risk is data quality: the framework's outputs are only as reliable as the reconciled data model beneath them, and a plan with poor data integration will find that the framework surfaces its data problems before it delivers its benefits. The mitigation is to treat data reconciliation as the first deliverable, not an afterthought. A second operational risk is organizational adoption: feedback loops that are not governed will exist on the diagram and not in the operation, so the mitigation is explicit ownership and cadence for each loop. A third is regulatory change: the framework embeds current rules, and a major regulatory shift requires the embedded logic to be updated, which is a maintenance obligation rather than a one-time build.

The methodological limitations are equally real. The framework's scope is Medicare Advantage and the ACA Exchange; it is not presented as a universal model for all lines of business. Its outputs depend on the quality of internal data and on the reasonableness of assumptions, chiefly the completion factors and RAF-driven revenue that Chapter 4 and Chapter 5 identified as the main sources of sensitivity. And because the validation is qualitative, the framework's benefits are described as design intent and observed direction rather than as quantified effect. The mitigation for each limitation is the same: explicit assumptions, documented sensitivity, and governance that keeps the model honest as conditions change (Actuarial Standards Board, 2017; Born et al., 2021).

Managerial Recommendations and Future Directions

For finance leadership considering the framework, the central recommendation is to begin with the data model and the feedback discipline rather than with analytics. A chief financial officer will gain more from a reconciled view that lets one quantity be estimated once than from a more sophisticated model built on inconsistent inputs. The framework's value is architectural, so its adoption should be architectural: build the connections first, and let each layer's sophistication grow within a structure that keeps the layers consistent. For actuarial, risk-management, contracting, and operations teams, the recommendation is to treat the loop hand-offs as part of the job rather than as an imposition, because the early-warning and interpretive mechanisms depend on signals being received and acted on across functional lines. A framework of this kind does not succeed on the strength of its models alone; it succeeds when the functions agree to be governed by shared signals, which is a change in operating culture as much as in analytics. The practical form of that agreement is modest and specific: each function accepts a defined signal on a defined cadence and commits to a defined response, and finance is accountable for the interpretive translation that makes the signal actionable. Where the framework aligns payment and reporting with the decisions they are meant to drive, it does for a plan's internal functions what value-based payment does for its provider relationships, replacing volume-of-activity incentives with outcome-aligned ones (Cattel et al., 2018).

The framework also points to its own extensions. Real-time analytics would tighten the loops, moving the framework from a monthly and forecast-cycle cadence toward continuous signal flow. Predictive and machine-learning methods, already emerging in risk-adjusted payment and in reserving, offer a path to sharper early-warning indicators and reserve estimates; recent work on machine-learning payment formulas, on interpretable pricing models, and on neural approaches to loss reserving indicates where those extensions would attach to the layers (Andriola et al., 2024; Kshirsagar et al., 2021; Schneider & Schwab, 2024). And the architecture is not confined to the two lines of business it was built for: the same three-layer structure with feedback loops could be applied to Medicaid managed care and to dual-eligible plans, which share the capitated, risk-adjusted, reserve-bearing structure that the framework is designed around. These are directions for development, not claims of completed work, and they define a research and practice agenda that the concluding chapter draws together.

Chapter outcome: a complete implementation perspective, from a worked propagation example through a phased roadmap, a qualitative validation, an honest statement of risks and limitations, and a set of managerial recommendations and forward directions.

CONCLUSION

Summary of Findings

This monograph began from a problem visible to any finance team working across Medicare Advantage and the ACA Exchange: the analytics used to manage risk pools, reserves, and forecasts are fragmented, and the fragmentation itself is a source of financial risk. Chapter 2 documented the gaps concretely. Risk-pool estimation is aggregated and retrospective, reserve modeling is static and organizationally isolated, Exchange forecasting is national in grain and treats regulation as an afterthought, and cross-functional communication is one-directional and untailored. These are not four independent problems but four expressions of one structural condition, in which each function optimises its own output and the connections between them are left unmanaged.

The framework developed in Chapters 3 through 6 responds to that condition with an architecture rather than a tool. Three operational layers, IPA-level risk-pool estimation, dynamic IBNR reserving with cross-functional reporting, and state-level Exchange forecasting, are connected by three explicit feedback loops: forecasting trends inform risk-pool reserve assumptions, IBNR development feeds forecast sensitivity, and risk-pool early-warning signals drive cross-functional action. The organizing discipline is to estimate each shared quantity once and let it flow between layers, so that the plan holds one reconciled view rather than several inconsistent ones. The research hypothesis, that a unified architecture with explicit cross-layer feedback loops can reduce fragmentation and improve financial outcomes across all three functional areas at once, is supported by the design and by the qualitative implementation experience reported in Chapter 7, while the limits of that evidence are stated plainly: the validation is directional and qualitative, not a controlled measurement.

Theoretical Contribution

The theoretical contribution is the extension of integrated financial analytics from a set of siloed health-plan functions to a single operational architecture with cross-layer feedback. The literature treats risk adjustment, loss reserving, and marketplace forecasting as separate disciplines, each with its own methods and its own body of work. This monograph does not add a new estimator to any of them. It specifies the architecture that makes the three one system, the data discipline that lets a shared quantity be computed once, and the feedback loops that convert three parallel processes into a self-informing whole. An architecture can be original even when its components are familiar, and the contribution here is precisely that architecture and the connections it formalises.

Practical Contribution

The practical contribution is a replicable blueprint for a working health plan. The framework provides, in concrete form, the specific tools a plan can adopt: a per-IPA financial liability exposure calculation with a leading, concurrent, and lagging early-warning hierarchy; a dynamic IBNR model with an explicit assumption framework and an interpretive reporting layer that carries audience-tailored templates, including a MAPD commission framework; and a state-level, decomposed, scenario-based Exchange forecast with regulatory drivers embedded rather than adjusted for. It provides the data-integration rules that make these outputs consistent and the feedback loops that keep them connected. Each element is presented as an engineered design that a plan can build on the data and platforms it already has, and the implementation roadmap sequences the work so that value appears from the first layer rather than only after the whole system is complete.

Final Recommendations

For organizations and finance leaders considering the framework or its individual layers, three recommendations follow from the work. The first is to begin with the data model and the feedback discipline rather than with analytics, because the framework's value is architectural and a sophisticated model built on unreconciled data reproduces the very fragmentation the framework exists to remove. The second is to treat the feedback loops as governed processes with defined owners, cadences, and hand-offs, since a loop that is not governed exists on the diagram and not in the operation. The third is to adopt the framework incrementally, starting from whichever layer addresses the most acute need and adding the loops as the layers connect, so that the investment is justified by results at each stage. Managed care finance will continue to grow in complexity as its two principal lines of business expand under evolving regulation; a plan that manages that complexity through a connected architecture, rather than through disconnected tools, is better positioned to keep its reserves adequate, its forecasts accurate, and its risk-sharing transparent across the lines of business it serves.

Glossary

Administrative Expense Ratio (AER). The share of premium revenue consumed by a health plan's non-medical operating costs.

Capitation. A fixed per-member-per-month payment to a provider organization in exchange for accepting responsibility for a defined scope of care, transferring utilization risk from payer to provider.

CMS-HCC Model. The CMS Hierarchical Condition Category model, which maps a member's documented diagnoses into condition categories to produce a risk score used in Medicare Advantage payment.

Completion Factor. The estimated fraction of an incurred period's ultimate claim cost that has been paid to date; the basis of chain-ladder and Bornhuetter-Ferguson reserve estimation.

Cross-Layer Feedback Loop. In this framework, an explicit, governed connection through which a signal produced in one operational layer becomes an input to another.

Early-Warning Indicator. A leading, concurrent, or lagging signal that flags an IPA risk pool trending toward a loss-sharing position early enough to act.

Incurred But Not Reported (IBNR). The estimated liability for care that members have received but that has not yet been reported to or fully paid by the plan.

Independent Physician Association (IPA). A legal entity that contracts with a health plan on behalf of a network of physicians, frequently under a capitated risk-sharing arrangement.

MAPD Commission Framework. The structure linking a plan's financial results to the commissions paid to external agents and agencies for Medicare Advantage Prescription Drug plans.

Medical Loss Ratio (MLR). The share of premium a health plan spends on medical care and quality improvement; regulated at a minimum of 80 or 85 percent under the Affordable Care Act.

Member-Month. One member enrolled for one month; the unit in which both revenue and cost accrue in health-plan finance.

Per Member Per Month (PMPM). A cost or revenue figure expressed per enrolled member per month, used to compare across states, products, and time.

Risk Adjustment Factor (RAF). A per-member coefficient derived from the CMS-HCC model that scales risk-adjusted capitation revenue.

Risk Pool. The pooled financial result of a risk-sharing arrangement, in which capitation revenue is set against incurred claims and any surplus or deficit is shared per contract terms.

Special Enrollment Period (SEP). A period outside open enrollment during which individuals may enroll in Exchange coverage following a qualifying event.

Statutory Accounting Principles (SAP). The conservative, solvency-oriented accounting framework U.S. health insurers use for regulatory financial reporting.

REFERENCES

1. Actuarial Standards Board. (2017). *Actuarial Standard of Practice No. 5: Incurred health and disability claims*. URL: http://www.actuarialstandardsboard.org/asops/incurred-health-and-disability-claims_186
2. Actuarial Standards Board. (2021). *Actuarial Standard of Practice No. 28: Statements of actuarial opinion regarding health insurance assets and liabilities.* URL: http://www.actuarialstandardsboard.org/asops/statements-of-actuarial-opinion-regarding-health-insurance-assets-and-liabilities/
3. Agbeko, T., Hiabu, M., Miranda, M. D. M., Nielsen, J. P., & Verrall, R. (2014). Validating the Double Chain Ladder Stochastic Claims Reserving Model. *Variance*. URL: https://doi.org/10.66573/001c.125918
4. Alai, D. H., Merz, M., & Wüthrich, M. V. (2009). Mean Square Error of Prediction in the Bornhuetter–Ferguson Claims Reserving Method. *Annals of Actuarial Science*. URL: https://doi.org/10.1017/s1748499500000580
5. Andriola, C., Ellis, R. P., Siracuse, J. J., Hoagland, A., Kuo, T., Hsu, H., Walkey, A. J., Lasser, K. E., & Ash, A. S. (2024). A Novel Machine Learning Algorithm for Creating Risk-Adjusted Payment Formulas. *JAMA Health Forum*. URL: https://doi.org/10.1001/jamahealthforum.2024.0625
6. Born, P. H., Eastman, E., & Sirmans, E. T. (2021). Managed care or carefully managed? Management of underwriting profitability by health insurers. *The Geneva Papers on Risk and Insurance Issues and Practice.* URL: https://doi.org/10.1057/s41288-021-00239-1
7. Cabral, M., Geruso, M., & Mahoney, N. (2018). Do Larger Health Insurance Subsidies Benefit Patients or Producers? Evidence from Medicare Advantage. *American Economic Review*. URL: https://doi.org/10.1257/aer.20151362
8. Carvalho, B. D. R. d., & Carvalho, J. V. d. F. (2019). A stochastic approach for measuring the uncertainty of claims reserves. *Revista Contabilidade & Finanças*. URL: https://doi.org/10.1590/1808-057x201907860
9. Cattel, D., Eijkenaar, F., & Schut, E. (2018). Value-based provider payment: towards a theoretically preferred design. *Health Economics Policy and Law*. URL: https://doi.org/10.1017/s1744133118000397
10. Centene Corporation. (2025). *Form 10-Q (Q1 2025).* U.S. Securities and Exchange Commission. URL: https://www.sec.gov/Archives/edgar/data/1071739/000107173925000086/cnc-20250331.htm
11. Centers for Medicare & Medicaid Services. (2024a). *2025 Medicare Advantage and Part D rate announcement.* URL: https://www.cms.gov/newsroom/fact-sheets/2025-medicare-advantage-and-part-d-rate-announcement
12. Centers for Medicare & Medicaid Services. (2024b). *Health insurance exchanges 2024 open enrollment report.* URL: https://www.cms.gov/files/document/health-insurance-exchanges-2024-open-enrollment-report-final.pdf
13. Centers for Medicare & Medicaid Services. (2024c). *Risk adjustment.* URL: https://www.cms.gov/medicare/payment/medicare-advantage-rates-statistics/risk-adjustment
14. Centers for Medicare & Medicaid Services. (2026). *Medical loss ratio.* URL: https://www.cms.gov/marketplace/private-health-insurance/medical-loss-ratio
15. Chukhrova, N., & Johannssen, A. (2021). Stochastic Claims Reserving Methods with State Space Representations: A Review. *Risks*. URL: https://doi.org/10.3390/risks9110198
16. Curto, V., Politzer, E., Anderson, T. S., Ayanian, J. Z., Souza, J., Zaslavsky, A. M., & Landon, B. E. (2025). Coding intensity variation in Medicare Advantage. *Health Affairs Scholar*. URL: https://doi.org/10.1093/haschl/qxae176
17. Dafny, L., Duggan, M., & Ramanarayanan, S. (2012). Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry. *American Economic Review*. URL: https://doi.org/10.1257/aer.102.2.1161
18. Dafny, L., Gruber, J., & Ody, C. (2015). More insurers lower premiums: Evidence from initial pricing in the health insurance marketplaces. *American Journal of Health Economics*. https://doi.org/10.1162/ajhe_a_00003
19. Delong, Ł., Lindholm, M., & Wüthrich, M. V. (2021). Collective reserving using individual claims data. *Scandinavian Actuarial Journal*. URL: https://doi.org/10.1080/03461238.2021.1921836
20. Dickstein, M., Ho, K., & Mark, N. (2021). *Market Segmentation and Competition in Health Insurance* (NBER Working Paper). National Bureau of Economic Research. URL: https://doi.org/10.3386/w29406

21. Dodini, S. (2022). Insurance Subsidies, the Affordable Care Act, and Financial Stability. *Journal of Policy Analysis and Management*. URL: https://doi.org/10.1002/pam.22418

22. Douven, R., Kauer, L., Demme, S., Paolucci, F., Ven, W. v. d., Wasem, J., & Zhao, X. (2022). Should administrative costs in health insurance be included in the risk-equalization? An analysis of five countries. *The European Journal of Health Economics*. URL: https://doi.org/10.1007/s10198-022-01436-y

23. Dreksler, S., Allen, C. W., Akoh-Arrey, A., Courchene, J. A., Junaid, B., Kirk, J., Lowe, W. B., O'Dea, S., Piper, J., SHAH, M. U. D., Shaw, G., Storman, D., Thaper, S., Thomas, L., Wheatley, M., & Wilson, M. A. (2015). SOLVENCY II TECHNICAL PROVISIONS FOR GENERAL INSURERS. *British Actuarial Journal*. URL: https://doi.org/10.1017/s1357321714000099

24. Duggan, M., Gupta, A., & Jackson, E. (2022). The impact of the Affordable Care Act: Evidence from California's hospital sector. *American Economic Journal: Economic Policy, 14*(1). https://doi.org/10.1257/pol.20190279

25. England, P. D., Verrall, R., & Wüthrich, M. V. (2012). Bayesian over-dispersed Poisson model and the Bornhuetter & Ferguson claims reserving method. *Annals of Actuarial Science*. URL: https://doi.org/10.1017/s1748499512000012

26. De Felice, M., & Moriconi, F. (2019). Claim Watching and Individual Claims Reserving Using Classification and Regression Trees. *Risks*. URL: https://doi.org/10.3390/risks7040102

27. Fung, V., Peitzman, C. G. K., Shi, J., Liang, C. Y., Dow, W. H., Zaslavsky, A. M., Fireman, B., Derose, S. F., Chernew, M. E., Newhouse, J. P., & Hsu, J. (2018). Adverse Selection into and within the Individual Health Insurance Market in California in 2014. *Health Services Research*. URL: https://doi.org/10.1111/1475-6773.12868

28. George, A. (2016). *The Effects of the ACA on Actuarial Pricing: How the Work of Health Actuaries Has Changed Post-ACA* [Undergraduate honors thesis, Portland State University]. URL: https://doi.org/10.15760/honors.215

29. Geruso, M., & Layton, T. (2017). Selection in Health Insurance Markets and Its Policy Remedies. *The Journal of Economic Perspectives*. https://doi.org/10.1257/jep.31.4.23

30. Geruso, M., & Layton, T. (2020). Upcoding: Evidence from Medicare on squishy risk adjustment. *Journal of Political Economy, 128*(3). https://doi.org/10.1086/704756

31. Geruso, M., Layton, T., & Prinz, D. (2019). Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. *American Economic Journal Economic Policy*. URL: https://doi.org/10.1257/pol.20170014

32. Handel, B., & Kolstad, J. (2022). The Affordable Care Act After a Decade: Industrial Organization of the Insurance Exchanges. *Annual Review of Economics*. URL: https://doi.org/10.1146/annurev-economics-051420-114714

33. Healthcare Financial Management Association. (1997). *P&P Board Statement 11: Accounting and reporting for risk contracts.* URL: https://www.hfma.org/accounting-and-financial-reporting/pp-board-statement-11-accounting-and-reporting-for-risk-contracts

34. Ho, K., Hogan, J., & Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the Medicare Part D program. *The RAND Journal of Economics, 48*(3). https://doi.org/10.1111/1756-2171.12207

35. Humana Inc. (2026). *Form 10-K (FY2025).* U.S. Securities and Exchange Commission. URL: https://www.sec.gov/Archives/edgar/data/49071/000004907126000009/hum-20251231.htm

36. Jacobs, P., & Kronick, R. (2018). Getting What We Pay For: How Do Risk-Based Payments to Medicare Advantage Plans Compare with Alternative Measures of Beneficiary Health Risk?. *Health Services Research*. URL: https://doi.org/10.1111/1475-6773.12977

37. Kan, H., Kharrazi, H., Chang, H., Bodycombe, D., Lemke, K., & Weiner, J. P. (2019). Exploring the use of machine learning for risk adjustment: A comparison of standard and penalized linear regression models in predicting health care costs in older adults. *PLoS ONE*. URL: https://doi.org/10.1371/journal.pone.0213258

38. Karaca-Mandic, P., Abraham, J., & Simon, K. (2013). IS THE MEDICAL LOSS RATIO A GOOD TARGET MEASURE FOR REGULATION IN THE INDIVIDUAL MARKET FOR HEALTH INSURANCE?. *Health Economics*. URL: https://doi.org/10.1002/hec.3002

39. Kshirsagar, R., Hsu, L., Greenberg, C. H., McClelland, M., Mohan, A., Shende, W., Tilmans, N. P., Guo, M., Chheda, A., Trotter, M. V., Ray, S., & Molina, M. A. (2021). Accurate and Interpretable Machine Learning for Transparent Pricing of Health Insurance Plans. *Proceedings of the AAAI Conference on Artificial Intelligence*. URL: https://doi.org/10.1609/aaai.v35i17.17776

40. Layton, T., Ellis, R. P., & McGuire, T. (2015). *Assessing Incentives for Adverse Selection in Health Plan Payment Systems* (NBER Working Paper). National Bureau of Economic Research. URL: https://doi.org/10.3386/w21531

41. Layton, T., Montz, E., & Shepard, M. (2017). *Health Plan Payment in U.S. Marketplaces: Regulated Competition with a Weak Mandate* (NBER Working Paper). National

Bureau of Economic Research. URL: https://doi.org/10.3386/w23444

42. Mack, T. M. (1993). Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *Astin Bulletin*. URL: https://doi.org/10.2143/ast.23.2.2005092
43. Mack, T. M. (2008). The Prediction Error of Bornhuetter/Ferguson. *Astin Bulletin*. URL: https://doi.org/10.2143/ast.38.1.2030404
44. Maisog, J. M., Li, W., Xu, Y., Hurley, B., Shah, H., Lemberg, R., Borden, T., Bandeian, S., Schline, M., Cross, R., Spiro, A., Michael, R., & Gutfraind, A. (2019). *Using massive health insurance claims data to predict very high-cost claimants: a machine learning approach* [Preprint]. arXiv. URL: https://doi.org/10.48550/arxiv.1912.13032
45. McGuire, T. G., Schillo, S., & van Kleef, R. C. (2020). Reinsurance, repayments, and risk adjustment in individual health insurance: Germany, the Netherlands, and the U.S. marketplaces. *American Journal of Health Economics, 6*(1). https://doi.org/10.1086/706796
46. McWilliams, J. M., Weinreb, G., Landrum, M. B., & Chernew, M. E. (2025). Use Of Patient Health Survey Data For Risk Adjustment To Limit Distortionary Coding Incentives In Medicare. *Health Affairs*. URL: https://doi.org/10.1377/hlthaff.2023.01351
47. Medicare Payment Advisory Commission. (2024). *The Medicare Advantage program: Status report* (Report to the Congress: Medicare Payment Policy, Ch. 12). URL: https://www.medpac.gov/document/chapter-12-the-medicare-advantage-program-status-report-march-2024-report/
48. Milliman. (2025). *Total cost of care* [Whitepaper]. URL: https://www.bcbs.com/media/pdf/Milliman-Total-Cost-of-Care-Whitepaper.pdf
49. Milliman. (n.d.). *Health Cost Guidelines Suite.* URL: https://www.milliman.com/en/products/health-cost-guidelines-suite
50. Newhouse, J. P., & McGuire, T. G. (2014). How Successful Is Medicare Advantage?. Milbank Quarterly. URL: https://doi.org/10.1111/1468-0009.12061
51. Obama, B. (2016). United States Health Care Reform. *JAMA*. URL: https://doi.org/10.1001/jama.2016.9797
52. Oyeka, O., & Wehby, G. L. (2023). Effects of state reinsurance programs on health insurance exchange premiums and insurer participation. *Health Services Research*. URL: https://doi.org/10.1111/1475-6773.14205
53. Reschovsky, J. D., Hadley, J., Saiontz-Martinez, C., & Boukus, E. R. (2011). Following the Money: Factors Associated with the Cost of Treating High-Cost Medicare Beneficiaries. *Health Services Research*. URL: https://doi.org/10.1111/j.1475-6773.2011.01242.x
54. Rice, N., & Smith, P. (2001). Capitation and Risk Adjustment in Health Care Financing: An International Progress Report. *Milbank Quarterly*. URL: https://doi.org/10.1111/1468-0009.00197
55. Robinson, J. C. (1999). The Future Of Managed Care Organization. *Health Affairs*. URL: https://doi.org/10.1377/hlthaff.18.2.7
56. Robinson, J. C. (2001). Theory and Practice in the Design of Physician Payment Incentives. *Milbank Quarterly*. URL: https://doi.org/10.1111/1468-0009.00202
57. Robinson, J. C., & Casalino, L. P. (1996). Vertical Integration and Organizational Networks in Health Care. *Health Affairs*. URL: https://doi.org/10.1377/hlthaff.15.1.7
58. Schneider, J. C., & Schwab, B. (2024). Advancing loss reserving: A hybrid neural network approach for individual claim development prediction. *Journal of Risk & Insurance*. URL: https://doi.org/10.1111/jori.12501
59. Sriram, K., & Shi, P. (2020). Stochastic loss reserving: A new perspective from a Dirichlet model. *Journal of Risk & Insurance*. URL: https://doi.org/10.1111/jori.12311
60. UnitedHealth Group. (2026). *Form 10-K (FY2025).* U.S. Securities and Exchange Commission. URL: https://www.sec.gov/Archives/edgar/data/731766/000073176626000062/unh-20251231.htm
61. Veprauskaitė, E., & Adams, M. (2017). Leverage and reinsurance effects on loss reserves in the United Kingdom's property–casualty insurance industry. *Accounting and Business Research*. URL: https://doi.org/10.1080/00014788.2017.1404440